ASSESSMENT OF IMPACT OF COVID-19 ON MENTAL HEALTH USING MACHINE LEARNING TECHNIQUES

A Thesis Submitted
in Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

Submitted by

Priyanka Arora

2K22/PHDCS/01

Under the Supervision of

Dr. Sonika Dahiya

Assistant Professor
Department of Software Engineering



To the

Department of Software Engineering

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India.

October, 2026

Acknowledgement

Reflecting on my experiences in pursuit of my Ph.D., I have realized that my work is not just my achievement but an outcome of collective inputs and encouragement received from various individuals. Due to constraints on space, it becomes impossible for me to acknowledge all of them. Nonetheless, it would give me immense pleasure to thank those individuals who have contributed immensely to the successful completion of my work.

I would like to express my sincere gratitude to my supervisor Dr. Sonika Dahiya, Assistant Professor, Department of Software Engineering, for her motivation, counselling, constant support, and encouragement, which helped me in formulating an urge for this research thesis. Her exemplary hard work and expert guidance have always assisted me in tackling difficulties.

I am grateful to Prof. Ruchika Malhotra, Head, Department of Software Engineering, for extending constant encouragement, administrative support, and providing all the necessary facilities needed to accomplish this research work. I also express my gratitude to the members of the Doctoral Research Committee and the esteemed faculty members for their invaluable feedback, suggestions, and words of encouragement during the periodic progress review meetings, which played a significant role in shaping the quality of this work.

I owe it all to my husband, Ankit Arya, who stood by me through every high and low of this journey. He is a person who helped me become a better individual every day. His constant support, sacrifices, and belief in my potential gave me the courage to persevere and complete this work.

I owe a heartfelt apology and my deepest gratitude to my children, Avyaan, Avika, and Ahaana, for my absence during many moments over these years. Their presence remained a constant source of joy, strength, and inspiration during the most demanding phases of this work.

I express my deepest gratitude to my parents, Mrs. Indu Arora and Mr. Suresh Arora, for their unwavering love, constant encouragement, and the strong values they

instilled in me. Their belief in my abilities has been a continuous source of strength and motivation. I am also grateful to my mother-in-law, Mrs. Rita, whose wisdom and life lesson—to learn by observing—has inspired me and guided me throughout this journey.

I would also like to acknowledge the support of my colleagues and fellow research scholars for their cooperation, discussions, and assistance in various forms. I am thankful to the technical and administrative staff of the department for their timely help and support throughout the course of my study.

Finally, I express my sincere thanks to everyone who directly or indirectly contributed to the successful completion of this thesis.

Priyanka Arora
2K22/PHDCS/01

"This thesis is dedicated to my loving husband, Ankit Arya, and to my three little musketeers—Avyaan, Avika, and Ahaana, whose love, patience, and unwavering support have profoundly shaped my life and this journey. Their faith in me, quiet sacrifices, and constant presence inspired me to persevere through challenges and pursue this work with commitment, resilience, and integrity."



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I, Priyanka Arora (2K22/PHDCS/01), hereby certify that the work which is being presented in the thesis entitled **"Assessment of Impact of COVID-19 on Mental Health Using Machine Learning Techniques"** in fulfillment of the requirements for the award of the Degree of Doctor of Philosophy, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out under the supervision of Dr. Sonika Dahiya, Assistant Professor, Department of Software Engineering, DTU.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

Priyanka Arora
2K22/PHDCS/01
PhD Research Scholar

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

This to certify that **Priyanka Arora** (2K22/PHDCS/01) has carried out her research work presented in this thesis entitled **"Assessment of Impact of COVID-19 on Mental Health Using Machine Learning Techniques"** for the award of Doctor of Philosophy in the discipline of Computer Science from Department of Software Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student herself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Dr. Sonika Dahiya
Assistant Professor
Department of Software Engineering
Delhi Technological University, Delhi-110042
India

Date:

ABSTRACT

Major mental health disorders, like depression and anxiety, constitute a growing global public health challenge grossly exacerbated by large-scale crises, such as that presented by the COVID-19 pandemic. The unprecedented disruption brought on by prolonged lockdowns, social isolation, economic uncertainty, and health-related stressors has increased psychological distress throughout diverse populations, leading to an increase in undiagnosed and untreated mental health disorders. The early recognition and correct classification of depression and anxiety are thus very important to facilitate timely intervention and the development of appropriate mental health policies. Against this background, Machine Learning (ML) and Deep Learning (DL) approaches have emerged to have potential for use in the automated assessment of mental health.

The proposed thesis seeks to move the state-of-the-art in automated detection of depression and anxiety by thoroughly examining classical ML, DL, and hybrid intelligent techniques based on psychological assessment data obtained during and post the COVID-19 era. Contrary to several prior approaches that have a narrow focus and binary classification boundaries, this study will prioritize scalability and multiclass prediction tasks.

The thesis starts by performing a systematic literature review on ML,DL based mental health evaluation techniques, particularly in terms of COVID-19-related depression and anxiety. The literature survey analyzes different techniques, datasets, instruments, feature designs, and metrics that are currently used. Critical gaps in the literature are identified, including the absence of long-term studies, multi-class studies, the consideration of the size of datasets in analyses, rigorous statistical validation of the models, and the use of learning framework combinations. The thesis also examines post-pandemic mental health outcomes, highlighting emerging trends and research

directions. These areas provide a foundation to the methodological contributions made in this thesis.

In the subsequent phase, the thesis examines the effect of the size of the dataset on the predictive capabilities of the traditional classifiers based on ML, such as the K-Nearest Neighbor (KNN) classifier, the Decision Trees (DT) classifier, the Random Forest (RF) classifier, and the Support Vector Machine (SVM) classifier. The experiment will make use of mental health survey data obtained during the COVID-19 pandemic. The results show that the model performs better with an increase in the size of the dataset, and it helps in understanding optimal levels of dataset collection needed to make a mental health model.

The work is further continued with a comparative assessment of DL models to classify depression and anxiety. Different DL models like Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long Short-Term Memory Network (LSTM), Convolutional Neural Network (CNN), and TabNet are used on three different mental health datasets. The data shows that the performance of DL models, especially LSTM and CNN, is significantly better than traditional ML models in modeling complex interactions in features and multiclass classification. This outcome points to the effectiveness of applying DL models to model psychological patterns related to depression and anxiety.

However, to circumvent some limitations in single models, some models in this research use a combination of several models with complementary properties. This combination will be in the form of hybrid classifiers; some examples include DT+KNN classification and RF+NN classification. The empirical outcomes have clearly indicated the ability of the proposed hybrid models to provide significantly higher levels of predictive accuracy and robustness than the stand-alone classifiers, performing up to 99-100% accuracy on some datasets. This confirms the effectiveness of the hybrid learning approach for precise mental health prediction.

The originalities in the thesis lie in the formulation of two predictive models for the detection of depression and anxiety. Firstly, the development of the multiclass Depression Anxiety-TabSVM (DA-TabSVM) hybrid model has taken place in order to tackle the complex, nonlinear nature of the data in the psychological tables. Secondly, the formulation of the Adaptive Sigma KNN (ASKNN) model has taken place. The

model has the unique ability to alter the weight given to the surrounding samples through the adaptation of the sigma parameter. In traditional KNN models, this has not been the case. Thorough experimental studies, including cross-validation, analysis of Area Under the Curve – Receiver Operating Characteristic (AUC-ROC) curves, and statistical significance tests with the Friedman and Wilcoxon signed-rank tests, support that the new models clearly outperform existing baselines in ML. The effect-size analysis via Cohen's d and Cliff's Delta values strengthens the relevance of the obtained improvements. Furthermore, this thesis proposes a Feature-Weighted Adaptive Kernel KNN (FW-AK-KNN) model that incorporates feature importance and adaptive kernel-based neighbor weighting to enhance both accuracy and interpretability in mental health prediction.

On the whole, this thesis makes a valuable contribution to a complete, scalable, and statistically valid framework for the automatic diagnosis of depression and anxiety using ML, DL, and hybrid models. With the development of trustworthy and efficient methods of identifying mental health disorders at an earlier stage, the presented solutions contribute to global initiatives with regard to mental health issues and are in line with the United Nations (UN)'s Sustainable Development Goal (SDG) 3 and the importance of AI solutions in managing this burden worldwide.

List of Publications

Papers Published in International Journals

1. **Priyanka Arora** and Sonika Dahiya. **"Unveiling the Impact of Dataset Size on Machine Learning Models for Anxiety and Depression Prediction amid the COVID-19 Pandemic: Determining Optimal Data Collection Thresholds"**, *Current Psychology*, vol. 44, no. 9, pp. 8106–8119, 2025. (SSCI, Impact Factor: 2.7, Publisher: Springer).
   DOI: 10.1007/s12144-025-07432-8. (Published)

2. **Priyanka Arora** and Sonika Dahiya. **"COVID-19-Induced Depression and Anxiety Detection: A Systematic Literature Review of Machine Learning and Deep Learning Techniques"**, *New Generation Computing*. (SCIE, Impact Factor: 2.1, Publisher: Springer).
   DOI: 10.1007/s00354-025-00305-5. (Published)

3. **Priyanka Arora** and Sonika Dahiya. **"A Novel Adaptive Sigma KNN Model for Depression and Anxiety Detection Following the COVID-19 Pandemic"**, *Scientific Reports*. (SCIE, Impact Factor: 4.9, Publisher: Nature Portfolio).
   DOI: https://doi.org/10.1038/s41598-025-24252-9(Published)

Papers Published in International Conferences

1. **Priyanka Arora** and Sonika Dahiya. **"Exploring Deep Learning Architectures for Mental Health Assessment: A Comparative Analysis"**, in *Proceedings of the International Conference on Artificial Intelligence and Emerging Technology (AISUMMIT2024)*, IEEE, 09 April 2025.
   DOI: 10.1109/GlobalAISummit62156.2024.10947873. (Published)
   Awarded Best Paper Award

2. **Priyanka Arora** and Sonika Dahiya. **"Enhancing Mental Health Diagnosis with DA-TabSVM: A Multi-Class Hybrid Approach for Detecting Depression and Anxiety"**, in *Proceedings of the International Conference on Machine Learning and Data Engineering (ICMLDE 2024)*, Elsevier, 2025. *Procedia Computer Science*, vol. 258, pp. 1348–1364.
   DOI: 10.1016/j.procs.2025.04.368. (Published)

3. **Priyanka Arora** and Sonika Dahiya. **"From Virus to Vulnerability: A Systematic Literature Review of Psychological Aftershocks Following the COVID-19 Pandemic"**, in *Proceedings of the 6th International Conference on Data Analytics & Management (ICDAM-2025)*, Springer, June 2025, vol. 7, pp. 267—284.
   DOI: https://doi.org/10.1007/978-3-032-03751-0_22. (Published)

Papers Under Review in Journals/Conferences

1. **Priyanka Arora** and Sonika Dahiya. **"Feature-Weighted Adaptive Kernel KNN: A Novel Approach for Post-COVID-19 Depression and Anxiety Detection"**. Revision Submitted. (Reviewers Assigned) (SCIE, Impact Factor: 1.2, Publisher: Springer).

2. **Priyanka Arora** and Sonika Dahiya. **"Psychometric Evaluation of the Beck Depression and Anxiety Inventories (BDI and BAI) among Post-COVID Indian College Students: Evidence from Exploratory and Confirmatory Factor Analyses"**. (Communicated).

Contents

ABSTRACT vi

List of Publications ix

List of Tables xix

List of Figures xxiii

Chapter 1: Introduction 1

1.1 Motivation 4

1.2 Research Gaps 5

1.2.1 Limited Literature Reviews 5

1.2.2 Lack of Analysis of Psychological Assessment Tools 5

1.2.3 Lack of Comparative Examination of ML/DL Model Performance 5

1.2.4 Absence of Post-COVID Mental Health Dataset 5

1.2.5 Need for Improved Model Development 5

1.2.6 Need to Analyze ML Performance on Varying Dataset Sizes . 6

1.3 Problem Statement 6

1.4 Overview of the Proposed Methodology 6

1.5 Research Objectives and Contribution of the Thesis 6

1.6 Significance of the Thesis 8

1.7 Thesis Overview 8

Chapter 2: Literature Review 10

2.1 Introduction 10

2.2 Types of Depression and Anxiety 11

2.2.1 Types of Depression 11

2.2.1.1 Major Depressive Disorder (MDD) 11
2.2.1.2 Persistent Depressive Disorder (Dysthymia) 12
2.2.1.3 Postpartum Depression 12
2.2.1.4 Seasonal Affective Disorder (SAD) 12
2.2.1.5 Other Specified Depressive Disorders 12
2.2.2 Types of Anxiety 13
2.2.2.1 Generalized Anxiety Disorder (GAD) 13
2.2.2.2 Panic Disorder 13
2.2.2.3 Social Anxiety Disorder (Social Phobia) 13
2.2.2.4 Specific Phobias 13
2.2.2.5 Other Anxiety Disorders 13
2.3 Risk Factors Associated with Depression and Anxiety 14
2.4 Screening Techniques 14
2.5 Psychological Tools for Identifying Anxiety, Depression, and Other Mental Health Issues 18
2.5.1 Beck depression inventory (BDI) 19
2.5.2 Beck anxiety inventory (BAI) 19
2.5.3 Generalized anxiety disorder-7 (GAD-7) 19
2.5.4 Patient health questionnaire-9 (PHQ-9) 19
2.5.5 Symptom checklist-90 (SCL-90) 19
2.6 Survey Methodology 31
2.6.1 Review Methodology 31
2.6.2 Statistical Insights 33
2.6.3 Research Questions 36
2.7 Open Datasets 38
2.8 Applications of ML and DL in Mental Health Diagnosis 42
2.8.1 ML Models Used for Detecting Depression and Anxiety . . . 42
2.8.1.1 Pre-Processing Algorithms 43
2.8.1.2 Feature Engineering Methods 44
2.8.1.3 Supervised Learning Classifiers 46
2.8.1.4 Supervised Learning Regressors 52
2.8.1.5 Unsupervised Learning 53

2.8.2 DL Models Used for Detecting Depression and Anxiety 80
2.8.3 Hybrid ML and DL Models for Depression and Anxiety Detection 87
2.9 Impact of Dataset Size on Machine Learning Model Performance . . . 88
2.10 Evaluation Metrics 90
2.11 Post-Pandemic Mental Health Crisis 94

Chapter 3: Unveiling the Impact of Dataset Size on Machine Learning Models for Anxiety and Depression Prediction amid the COVID-19 Pandemic: Determining Optimal Data Collection Thresholds 98

3.1 Introduction 98
3.2 Proposed Methodology 99
3.2.1 Datasets Used 99
3.2.2 Mental Health Assessment 101
3.2.3 Data Slicing and Data Augmentation 101
3.2.4 Data Preprocessing and Predictive Modelling 102
3.2.4.1 Machine Learning Algorithms Used 103
3.2.5 Evaluation Metrics 104
3.3 Statistical Analysis of Model Performance 104
3.4 Results 105
3.4.1 Performance of Dataset 1 for GAD-7 and PHQ-9 105
3.4.2 Performance of Dataset 2 for GAD-7 and PHQ-9 113
3.4.3 Performance of Dataset 3 for GAD-7 and PHQ-9: 119
3.4.4 Results of Statistical Analysis on Model Performance 126
3.5 Discussion 129
3.5.1 GAD-7 Results 129
3.5.2 PHQ-9 Results 130
3.5.3 Statistical Results 130
3.5.4 Cross-Validation analysis 131
3.5.5 Limitations 137
3.6 Summary 137

Chapter 4: Exploring Deep Learning Architectures for Mental Health Assessment: A Comparative Analysis 139

4.1 Introduction . . . 139
4.2 Proposed Methodology . . . 140
4.2.1 Datasets Used . . . 140
4.2.2 Mental Health Assessment . . . 141
4.2.3 DL Models . . . 141
4.2.4 Evaluation metrics . . . 142
4.3 Results . . . 143
4.3.1 Analysis of DL models on Dataset 1 . . . 143
4.3.2 DL model performance using Dataset 2 . . . 145
4.3.3 Performance analysis of DL models on Dataset 3 . . . 146
4.4 Discussion . . . 148
4.5 Summary . . . 149

Chapter 5: Enhancing Mental Health Diagnosis with DA-TabSVM: A Multi-Class Hybrid Approach for Detecting Depression and Anxiety 151

5.1 Introduction . . . 151
5.2 Proposed Methodology . . . 153
5.2.1 Dataset Used . . . 153
5.2.2 Mental Health Assessment . . . 154
5.2.3 ML Algorithms . . . 154
5.2.4 DL Models . . . 155
5.2.5 Hybrid Models . . . 155
5.2.5.1 DT+KNN . . . 155
5.2.5.2 RF+NN . . . 156
5.2.5.3 DA-TabSVM . . . 157
5.2.6 Evaluation metrics . . . 160
5.3 Results . . . 161
5.3.1 Performance analysis of ML and DL models on Depression . . . 161
5.3.2 Evaluation of ML and DL models' performance in relation to anxiety . . . 165
5.4 Discussion . . . 169
5.5 Summary . . . 172

Chapter 6: A Novel Adaptive Sigma KNN Model for Depression and Anxiety Detection Following the COVID-19 Pandemic **173**

6.1 Introduction . . . 173
6.2 Proposed Methodology . . . 174
6.2.1 Datasets Used . . . 174
6.2.1.1 Proposed Dataset . . . 175
6.2.1.2 Benchmark Datasets used . . . 183
6.2.2 Mental health assessment . . . 189
6.2.3 Psychometric Validation . . . 189
6.2.4 Predictive Modeling using KNN and its variants . . . 190
6.2.4.1 KNN . . . 190
6.2.4.2 Adaptive KNN (AKNN) . . . 190
6.2.4.3 Fuzzy KNN (FKNN) . . . 191
6.2.4.4 K-means Clustering-Based KNN (KMKNN) . . . 191
6.2.4.5 Weight-Adjusted KNN (WKNN) . . . 192
6.2.4.6 Generalized Mean Distance KNN (GMDKNN) . . . 192
6.2.4.7 Mutual KNN (MKNN) . . . 192
6.2.4.8 Ensemble Approach KNN (EKNN) . . . 193
6.2.4.9 ASKNN . . . 193
6.2.4.10 Advantages of Adaptive Sigma KNN . . . 196
6.2.5 Data Preprocessing and Predictive Modelling . . . 199
6.2.6 Evaluation Metrics . . . 199
6.2.7 Statistical Analysis of Performance of ASKNN . . . 199
6.3 Results . . . 199
6.3.1 Experimental Findings of the Dataset . . . 200
6.3.1.1 Descriptive and Statistical Analysis of the Dataset . . . 200
6.3.1.2 Correlation and Reliability Analysis . . . 202
6.3.1.3 Factor Analysis (EFA and CFA) . . . 204
6.3.2 Model Performance Evaluation Across All Datasets . . . 205
6.3.2.1 Accuracy . . . 205
6.3.2.2 Precision . . . 207
6.3.2.3 Recall . . . 209

6.3.2.4 F1-score . . . 211
6.3.2.5 5-Fold Cross Validation . . . 213
6.3.3 Performance of the ASKNN model on Primary Dataset . . . 216
6.3.3.1 Depression . . . 216
6.3.3.2 Anxiety . . . 218
6.3.4 Statistical Analysis and Significance Testing . . . 221
6.3.4.1 Friedman Test Analysis . . . 221
6.3.4.2 Wilcoxon Signed-Rank Test Results . . . 221
6.3.4.3 Effect Size Analysis . . . 223
6.4 Discussion . . . 224
6.5 Summary . . . 226

Chapter 7: Feature-Weighted Adaptive Kernel KNN: A Novel Approach for Post-COVID-19 Depression and Anxiety Detection 227

7.1 Introduction . . . 227
7.2 Proposed Methodology . . . 228
7.2.1 Datasets used . . . 229
7.2.2 Mental health assessment . . . 229
7.2.3 ML Algorithms . . . 230
7.2.4 Overview of the Proposed Model . . . 230
7.2.4.1 Feature Weighting Using RF . . . 230
7.2.4.2 Locally Adaptive Kernel Function . . . 230
7.2.4.3 Kernel-Weighted Voting Mechanism . . . 231
7.2.5 Evaluation metrics . . . 231
7.3 Results . . . 232
7.3.1 Feature Contribution Analysis for Depression Detection using FW-AK-KNN . . . 234
7.3.2 Feature Contribution Analysis for Anxiety Detection using FW-AK-KNN . . . 235
7.4 Discussion . . . 237
7.5 Summary . . . 239

Chapter 8: Conclusion, Future Work & Social Impact 241

8.1 Conclusion 241
8.1.1 Lessons Learnt 243
8.1.2 Ethical Implications and Responsible Use 244
8.1.2.1 Privacy Concerns 244
8.1.2.2 Potential Biases 244
8.1.2.3 Responsible Use 244
8.1.2.4 Limitations 244
8.2 Future Work 246
8.3 Social Impact and Applications 246

References **249**

Biography – Supervisor **279**

Biography – Student **280**

List of Tables

1.1 Anxiety and depression prevalence worldwide (in millions) between 1990 and 2020 2

1.2 Global Prevalence of Anxiety and Major Depressive Disorders in 2020 2

2.1 Techniques used to detect mental health problems 16

2.2 Instruments Used to Detect Depression 19

2.3 Instruments Used to Detect Anxiety 23

2.4 Instruments Used to Detect Stress 24

2.5 Instruments Used to Detect Other Conditions 24

2.6 Research questions related to depression and anxiety detection 37

2.7 Analyzing survey papers related to the detection of depression and anxiety through a comparative lens, with a focus on the research questions posed in these papers. 37

2.8 Various open datasets available for detecting depression and anxiety . 39

2.9 Year-wise previous studies that used ML-based techniques for mental health detection 55

2.10 Year-wise previous studies that used ML-based mental health detection along with feature selection 68

2.11 Comparative analysis of prior DL-based studies for depression and anxiety detection 84

2.12 Overview of existing approaches in mental health assessment using hybrid models 88

2.13 Comparative analysis of related work and their limitations 90

3.1 Description of datasets used 100

3.2 Anonymized Illustrative Snapshot of the Open-Source Dataset D1 . . 100

3.3 Anonymized Illustrative Snapshot of the Open-Source Dataset D2 . . 100
3.4 Anonymized Illustrative Snapshot of the Open-Source Dataset D3 . . 100
3.5 Hyperparameters for each classifier 103
3.6 Performance of Dataset–1 for GAD-7 and PHQ-9 using different machine learning models. 108
3.7 Misclassification rates of machine learning models across different dataset sizes for Dataset–1. 111
3.8 Performance of Dataset–2 for GAD-7 and PHQ-9 using different machine learning models. 115
3.9 Misclassification rates of machine learning models across different dataset sizes for Dataset–2 117
3.10 Performance of Dataset–3 for GAD-7 and PHQ-9 using different machine learning models 122
3.11 Misclassification rates of machine learning models across different dataset sizes for Dataset–3 124
3.12 Significant p-values (at the 5% level) and F-statistic from ANOVAs of effects of dataset size on performance metrics 126
3.13 Friedman Test Results for PHQ-9 and GAD-7 127
3.14 Summary of Spearman Rank Correlation Between Dataset Size and Model Performance After Holm Adjustment 128
3.15 CV scores of all datasets for GAD-7 and PHQ-9 using different machine learning models 133
3.16 Performance of wine dataset for different machine learning models . . 136

4.1 An overview of the datasets utilised in earlier research. 140
4.2 Advantages and disadvantages of the models used to forecast anxiety and sadness 142
4.3 Configuration and hyperparameters of the deep learning models . . . 142
4.4 Analysis of DL models on dataset 1 for anxiety 143
4.5 Analysis of DL models on dataset 1 for depression 144
4.6 Analysis of DL models on dataset 2 for anxiety 145
4.7 Analysis of DL models on dataset 2 for depression 145

4.8 Analysis of DL models on dataset 3 for anxiety 147
4.9 Analysis of DL models on dataset 3 for depression 147

5.1 Hybrid Model Comparison with Conventional ML and DL Models . . 152
5.2 Description of datasets used 154
5.3 Anonymized Illustrative Snapshot of the Dataset 3 154
5.4 Hyperparameters for DT+kNN 156
5.5 Hyperparameters for RF+NN 157
5.6 Hyperparameters for DA-TabSVM 158
5.7 Algorithm of working of Hybrid Models 160
5.8 Performance analysis of all the models to detect depression. 164
5.9 Performance analysis of all the models to detect anxiety. 168
5.10 Accuracy Comparison of Hybrid Models DT+kNN and RF+NN . . . 171
5.11 AUC-ROC Comparison of Hybrid Models for Depression 171
5.12 AUC-ROC Comparison of Hybrid Models for Anxiety 171

6.1 A provisional diagnosis of BDI 176
6.2 A provisional diagnosis of BAI 176
6.3 Results of questions based on the BDI 178
6.4 Results of questions based on the BAI 181
6.5 Overview of datasets employed in this work, comprising primary depression and anxiety datasets, open-source benchmark datasets, and other medical datasets for comparative work 185
6.6 Anonymized Illustrative Snapshot of the Primary Depression Dataset D1a 189
6.7 Anonymized Illustrative Snapshot of the Primary Anxiety Dataset D1b 189
6.8 Summary of KNN Variants and Their Key Characteristics 197
6.9 Descriptive Statistics and Factor Analytic Results for BDI Items . . . 201
6.10 Descriptive Statistics and Factor Analytic Results for BAI Items 202
6.11 Cronbach's Alpha for Total and Subscales of BDI and BAI 204
6.12 EFA and CFA Summary for BDI and BAI 204
6.13 Accuracy of various KNN implementations on multiple datasets, such as depression, anxiety, and general medical datasets. 206

6.14 Precision scores for different KNN variants over different datasets. . . 208
6.15 Recall scores for different KNN variants over different datasets, establishing the enhanced effectiveness of ASKNN at accurately identifying positive cases. 210
6.16 F1 scores of different KNN-based models on different datasets. 212
6.17 5-fold cross-validation mean accuracy scores and standard deviation for various KNN-based models. 214
6.18 Friedman Test Results for Model Comparison 221
6.19 Wilcoxon Signed-Rank Test (p-values) Comparing ASKNN with Other Models 222
6.20 Effect Size Analysis 223
6.21 Comparative theoretical overview of computational details for different KNN-based models 226

7.1 Summary of datasets used in the evaluation of FW-AK-KNN 229
7.2 Performance comparison of the proposed FW-AK-KNN model 232
7.3 Feature Importance Scores for BDI-based Depression Detection using FW-AK-KNN 235
7.4 Feature Importance Scores for BAI-based Anxiety Detection using FW-AK-KNN 236
7.5 Ablation study results on the BDI and BAI datasets using various kernel functions and gamma values. The highest accuracy and F1-score were obtained using the Epanechnikov kernel with a gamma value of 0.1 on both datasets. 238

List of Figures

1.1 Over time, anxiety and depression illnesses have become more common worldwide. . . . 3
1.2 In 2020, anxiety and depression illnesses will be very prevalent worldwide. 3

2.1 Types of depression. . . . 12
2.2 Types of anxiety. . . . 14
2.3 Anxiety and depression are diagnosed using a variety of methods. . . 15
2.4 A chord diagram demonstrating the author's relationship to the psychological instrument(s). . . . 31
2.5 PRISMA flow diagram for study selection. . . . 33
2.6 Distribution of Articles According to Indexing. . . . 34
2.7 Year-wise distribution of studies. . . . 34
2.8 Countries under study. . . . 35
2.9 ML and DL models were used in earlier research to identify mental health issues. . . . 35
2.10 Depiction of the most used ML models. . . . 80
2.11 A chord diagram demonstrating the author's relationship to DL and ML approaches. . . . 86
2.12 The application of ML and DL for depression and anxiety detection. . 87
2.13 Frequency of different mental health assessment tools utilized in chosen studies for the assessment of post-COVID-19 psychological symptoms. 95

3.1 a Visual representation of proposed research work, b Provisional diagnosis and severity of GAD-7 & Figure 1c Provisional diagnosis and severity of PHQ-9 . . . 99

3.2 Comparison between the performance metrics of GAD-7 and PHQ-9 on varied dataset sizes for Dataset-1 106
3.3 Comparison between the performance metrics of GAD-7 and PHQ-9 on varied dataset sizes for Dataset-2 114
3.4 Comparing the GAD-7 and PHQ-9 performance measures on various sizes of Dataset-3 119
3.5 Comparative analysis of misclassification rates for different machine learning models for GAD-7 120
3.6 Comparative analysis of misclassification rates for different machine learning models for PHQ-9 121
3.7 Comparative analysis of cross-validation scores for different machine learning models for GAD-7 131
3.8 Comparative analysis of cross-validation scores for different machine learning models for PHQ-9 132

4.1 Score-Based Distribution of Anxiety and Depression Severity Levels . 141
4.2 Performance of DL models for anxiety and depression on dataset 1. . . 144
4.3 Performance of DL models for depression on dataset 1 using the Confusion Matrix. 144
4.4 Performance of DL models for depression on dataset 1 using the Confusion Matrix. 144
4.5 Performance of DL models for anxiety and depression on dataset 2. . . 146
4.6 Performance of DL models for depression on dataset 2 using the Confusion Matrix. 146
4.7 Performance of DL models for depression on dataset 2 using the Confusion Matrix. 146
4.8 Performance of DL models for anxiety and depression on dataset 3. . . 147
4.9 Performance of DL models for depression on dataset 3 using Confusion Matrix. 147
4.10 Performance of DL models for depression on dataset 3 using Confusion Matrix. 148

4.11 a. Evaluation of the accuracy of many DL models for anxiety in comparison. b. Evaluation of the accuracy of many DL models for depression in comparison. 148
4.12 Evaluating the ability to predict depression using ROC Curves for Anxiety across all datasets using Radar Plot 149
4.13 Evaluating the ability to predict depression using ROC Curves for Anxiety across all datasets using Radar Plot 149

5.1 Proposed DA-TabSVM architecture. 158
5.2 ML model's performance in detecting depression (a) Accuracy; (b) Precision; (c) Recall; (d) F1-score 162
5.3 Comparison of the performance of hybrid models in identifying depression. 163
5.4 Performance of DA-TabSVM for depression using Confusion Matrix. . 163
5.5 Comparison of ML models' performance to detect anxiety (a) Accuracy; (b) Precision; (c) Recall; (d) F1-score. 166
5.6 Comparison of hybrid models' performance to detect anxiety. 167
5.7 Performance of DA-TabSVM for anxiety using Confusion Matrix. . . 167
5.8 Comparison of accuracy of hybrid models' performance to detect depression (a) Dataset 1; (b) Dataset 2; (c) Dataset 3 169
5.9 Comparison of accuracy of hybrid models' performance to detect anxiety (a) Dataset 1; (b) Dataset 2; (c) Dataset 3 170
5.10 Comparison of accuracy across DT+KNN and RF+NN. 171
5.11 Comparison of hybrid models' AUC-ROC scores for depression detection. 171
5.12 Comparison of hybrid models' AUC-ROC scores for anxiety detection. 172

6.1 BDI score ranges and depression severity levels. 175
6.2 BAI score ranges and anxiety severity levels. 176
6.3 BDI Response Analysis Using Circos Plot 180
6.4 BAI Response Analysis Using Circos Plot 183
6.5 Adaptive neighborhood in ASKNN. 196

6.6 a Prevalence and severity of depression in students after the COVID-19 pandemic b Prevalence and severity of anxiety in students after the COVID-19 pandemic 200
6.7 Significant Depression Factors 203
6.8 Significant Anxiety Factors 203
6.9 CFA path diagram for BDI that displays factor loadings and correlations between latent variables and observable items. 204
6.10 CFA path diagram for BAI that displays factor loadings and correlations between latent variables and observable items. 204
6.11 Accuracy comparison of different KNN variants across multiple datasets.206
6.12 Precision comparison of different KNN variants across multiple datasets.208
6.13 Recall comparison of different KNN variants across multiple datasets. 211
6.14 F1-Score comparison of different KNN variants across multiple datasets.212
6.15 CV score comparison of different KNN variants across multiple datasets.216
6.16 Confusion matrix for depression detection employing ASKNN 217
6.17 AUC-ROC curve for depression detection employing ASKNN 218
6.18 Confusion matrix for anxiety detection employing ASKNN 219
6.19 AUC-ROC curve for anxiety detection employing ASKNN 220
6.20 Wilcoxon Signed-Rank Test (p-values) Comparing ASKNN with Other Models 223
6.21 An effect size analysis depicting Cohen's d for KNN variants with ASKNN, and b An effect size analysis depicting Cliff's delta for KNN variants with ASKNN 224
6.22 Comparison between the performance of KNN and ASKNN for various datasets a Accuracy b Recall c Precision d F1 score 225

7.1 Working of KNN with FW-AK-KNN a KNN assigns equal votes b FW-AK-KNN weighs neighbors by distance 231
7.2 Working of KNN with FW-AK-KNN a KNN assigns equal votes b FW-AK-KNN weighs neighbors by distance 233
7.3 Feature Importance Analysis for Depression Detection using FW-AK-KNN 236

7.4 Feature Importance Analysis for Anxiety Detection using FW-AK-KNN 237
7.5 Significant Depression and Anxiety Features 239

List of Abbreviations

ML Machine Learning vi
DL Deep Learning vi
KNN K-Nearest Neighbor vii
DT Decision Trees vii
RF Random Forest vii
SVM Support Vector Machine vii
ANN Artificial Neural Network vii
RNN Recurrent Neural Network vii
LSTM Long Short-Term Memory Network vii
CNN Convolutional Neural Network vii
ASKNN Adaptive Sigma KNN vii
DA-TabSVM Depression Anxiety-TabSVM vii
FW-AK-KNN Feature-Weighted Adaptive Kernel K-Nearest Neighbors viii
UN United Nations viii
SDG Sustainable Development Goal viii
AUC-ROC Area Under the Curve – Receiver Operating Characteristic viii
OCD obsessive-compulsive disorder 2
PTSD Post-Traumatic Stress Disorder 2
GAD Generalized Anxiety Disorder 2
SARS Severe Acute Respiratory Syndrome 94
HIV Human Immunodeficiency Virus 94
AI artificial intelligence 4

EEG electroencephalogram 10
MRI Magnetic Resonance Imaging 10
MDD Major Depressive Disorder 11
SAD Seasonal Affective Disorder 12
GAD Generalized Anxiety Disorder 2
HRV Heart Rate Variability 14
DSM-5 Diagnostic and Statistical Manual of Mental Disorders 15
BDI Beck depression inventory 6
BAI Beck anxiety inventory 6
PHQ-9 Patient health questionnaire-9 19
SCL-90 Symptom checklist-90 19
DASS-21 Depression, Anxiety, and Stress Scale 20
CES-D 7 Center for Epidemiological Studies Depression Scale 20
PHQ-4 Patient health questionnaire-4 20
PHQ-8 Patient health questionnaire-8 20
PHQ-2 Patient health questionnaire-2 21
PHQ-15 Patient health questionnaire-15 21
DASS-9 Depression Anxiety Stress Scale-9 21
EPDS Edinburgh Postnatal Depression Scale 21
STAI State-Trait Anxiety Inventory 22
ZSDS Zung Self-Rating Depression Scale 22
BSI-18 Brief Symptom Inventory-18 22
BDC Burns Depression Checklist 22
GAD-7 Generalized Anxiety Disorder-7 23
GAD-2 Generalized Anxiety Disorder-2 23
SAI State Anxiety Inventory 23
SAS Self-rating Anxiety Scale 23
PSS Perceived Stress Scale 24
PSS-10 Perceived Stress Scale-10 24

COPE questionnaires Coronavirus Perinatal Experiences 24
SCL-5 Symptom Check List 24
SBQ-R Suicidal Behaviors Questionnaire-Revised 24
ISI Insomnia Severity Index 24
CD-RISC Connor–Davidson resilience scale 25
SSQ Support Questionnaire 25
DCPR Diagnostic Criteria for Psychosomatic Research 25
SRMH Self-rated mental health 25
SPANE-8 Scale of Positive and Negative Experience 8 26
SWLS Satisfaction with Life Scale 26
IES-R Impact of Events Scale-Revised 26
Y-BOCS Yale-Brown Obsessive Compulsive Scale 27
OCD Obsessive-Compulsive Disorder 2
BABS Brown Assessment of Belief Scale 27
FSS Fatigue Severity Scale 27
GSES General Self-Efficacy Scale 27
SCS Self-reported Coping Style 28
COPE-NVI-25 Coping Orientations to Problems Experienced 28
BSCS Brief Self-Control Scale 28
LOC Locus of Control 28
BFI-10 Big Five Inventory 28
DMRS-SR-30 Defense Mechanisms Rating Scale - Self Report 28
PFI Professional Fulfillment Index 28
MHI-5 Mental Health Inventory 29
MBI Maslach Burnout Inventory 29
IES Impact of Event Scale 29
PSQI Pittsburgh Sleep Quality Index 29
DCSQ Demand-Control-Support Questionnaire 29
GHQ-12 General Health Questionnaire 29

K-6 Kessler 6 Scale 30
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses 32
SLR Systematic Literature Reviews 33
RFE Recursive Feature Elimination 45
SHAP Shapley Additive Explanation 45
NB Naïve Bayes 47
CART Classification and Regression Trees 49
XGBoost Extreme Gradient Boosting 51
AdaBoost Adaptive boosting 51
LR Logistic Regression 51
SVR Support Vector Regression 52
DT Regression Decision Tree Regression 53
GMM Gaussian Mixture Model 54
BIRCH Balanced Iterative Reducing and Clustering using Hierarchies 54
DCNN Deep Convolutional Neural Network 82
DNN Deep Neural Network 82
MSE Mean Square Error 91
RMSE Root Mean Square Error 92
RMSEA Root Mean Square Error of Approximation 204
FP false-positive 92
FN false-negative 92
TP true-positive 92
TN true-negative 92
PPV Positive predictive value 92
MCC Matthew's Correlation Coefficient 93
AKNN Adaptive KNN xvi
FKNN Fuzzy KNN xvi
KMKNN K-means Clustering-Based KNN xvi
WKNN Weight-Adjusted KNN xvi

GMDKNN Generalized Mean Distance KNN xvi
MKNN Mutual KNN xvi
MHC–SF Mental Health Continuum–Short Form 26
EKNN Ensemble Approach KNN xvi
EFA Exploratory Factor Analysis 190
CFA Confirmatory Factor Analysis 190
NFI Normed Fit Index 190
CFI Comparative Fit Index 204
GFI Goodness-of-Fit Index 190
TLI Tucker-Lewis Index 204
AGFI Adjusted Goodness-of-Fit Index 190
KMO Kaiser Meyer Olkin 190
FW-AK-KNN Feature-Weighted Adaptive Kernel KNN viii
BMI Body Mass Index 47
ECG Electrocardiogram 47

Chapter 1

Introduction

The issue of mental health is one of the crucial components of overall health and wellness, and it has been recognized as such in the framework of the UN SDG, as part of the aim for "Good Health and Well-Being". The goal is aiming to reduce premature mortality from non-communicable diseases, including mental health disorders, by 2030 [1]. Mental illnesses like anxiety and depression may have a huge impact on the personal, social, educational, and professional lives of the patients. But stigma, ignorance, and lack of access to professional mental health care may cause delays in diagnoses and treatment, especially in developing nations like India.

The pandemic which started in the year 2020 has led to some unusual health care steps like lockdowns, traveling restrictions, and disruption in education and work [88, 165, 216]. These events, along with social isolation, financial instability, and the loss of family members, led to greater stress levels and heightened fears of a corresponding mental health crisis [133, 135, 154, 181, 189].

The COVID-19 pandemic has left no lives unaffected [133, 221]. Unknown to many, people were discreetly battling growing mental health issues, masking the actual scope of the catastrophe. Social isolation and loneliness are inherently associated with worse physical and mental health results impacting children and youth [20, 135, 139, 154, 163, 182, 232].This research focuses on the less visible mental health consequences of the COVID-19 pandemic, particularly the increase in anxiety and depression. Depressive disorders are characterized by persistent depressive symptoms, diminished interest or pleasure, fatigue, changes in sleep or appetite, and difficulty

concentrating, while anxiety disorders are marked by sensations of apprehension and dread [77]. Anxiety disorders include a wide range of conditions, such as panic disorder, Generalized Anxiety Disorder (GAD), social anxiety disorder, Obsessive-Compulsive Disorder (OCD), and Post-Traumatic Stress Disorder (PTSD). Seeking help for mental health problems poses a serious stigma in India, and individuals often avoid seeking professional help, resulting in delayed or missed diagnosis and affecting their personal, social, and economic well-being. Depressive disorders are the greatest cause of non-fatal health losses globally [260].

The prevalence of anxiety and depression disorders worldwide from 1990 to 2020 is shown in Table 1.1. Following natural disasters, the incidence of psychiatric diseases and psychological discomfort rises [16]. Fig. 1.1 shows how the prevalence of anxiety and depression disorders changed globally between 1990 and 2020. Anxiety disorders affected 194.9 million persons in 1990, while depressive illnesses affected 170.8 million [50]. These figures rose to 267.2 million for anxiety, and 311.1 million for severe depressive disorder by 2015 [251]. According to Table 1.1, 301.4 million individuals suffered from anxiety disorders and 279.6 million from depressive disorders in 2019. Both conditions significantly increased as a result of the pandemic, as Table 1.2. According to pre-pandemic forecasts, there will be 193 million cases of severe depressive disorder and 298 million cases of anxiety disorders in 2020. However, post-pandemic data showed that 246 million people had severe depression and 374 million had anxiety disorders [206], which has been depicted in Fig. 1.2.

Table 1.1: Anxiety and depression prevalence worldwide (in millions) between 1990 and 2020

Year	**Anxiety**	**Depression**
1990	194.9	170.8
2015	267.2	311.2
2019	301.4	279.6
2020	374.0	246.0

Table 1.2: Global Prevalence of Anxiety and Major Depressive Disorders in 2020

Year / Disorder	**Cases in 2020 before COVID-19**	**Additional cases in 2020 due to COVID-19**	**Total cases in 2020**
Anxiety	298	76.2	374
Depression	193	53.2	246

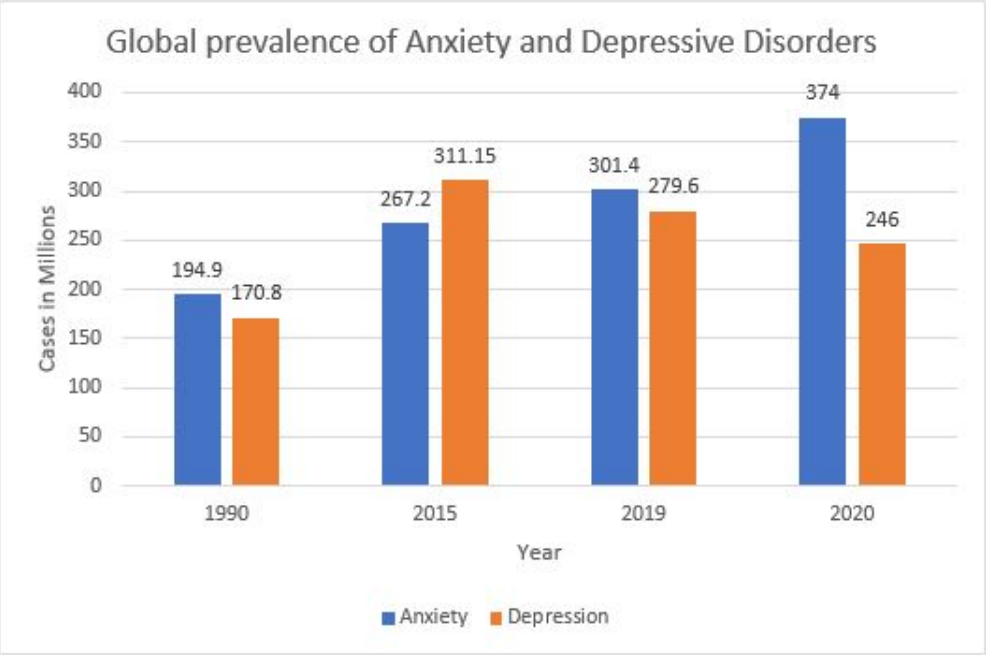


Figure 1.1: Over time, anxiety and depression illnesses have become more common worldwide.

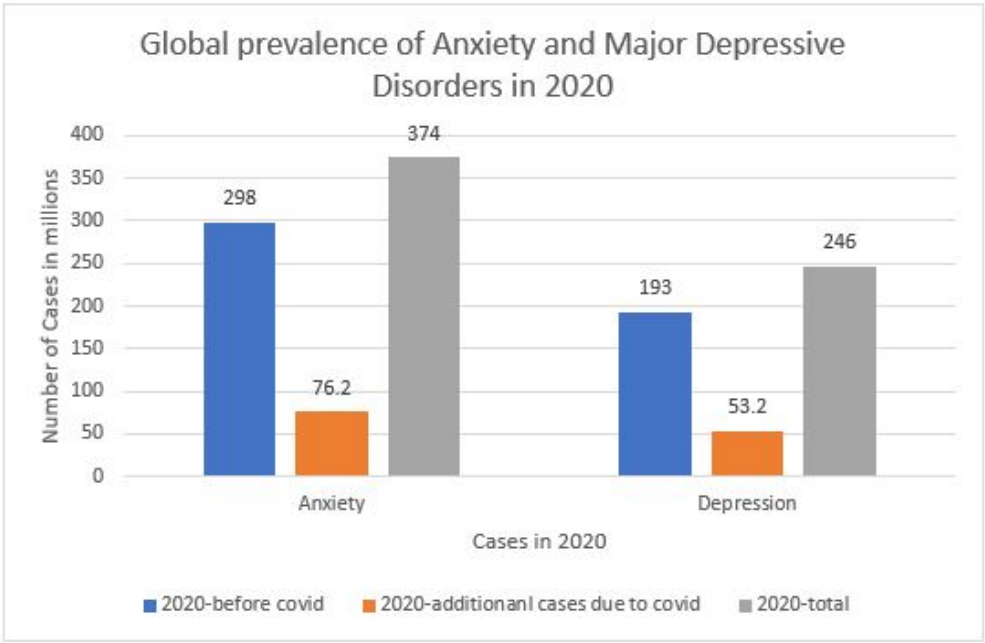


Figure 1.2: In 2020, anxiety and depression illnesses will be very prevalent worldwide.

1.1 Motivation

Past pandemics have demonstrated that serious infectious diseases can have a profound psychological impact, often leading to ”moral panic” and a sense of stigma that spreads globally [76]. COVID-19 was accompanied by an ”infodemic”: a huge amount of frightening information and disinformation that increased public apprehension. False information continues to drive societal concern since individuals now rely on social media for information. Behavioral responses, such as panic buying, emerge due to the fear of shortages of supplies. Such compulsive buying is according to heightened arousal and feelings of scarcity in the context of the pandemic [101]. Effective usage of the government's social media outlets will mitigate the effects of this infodemic by reducing public panic and increasing proper compliance with new regulations [233]. That calls for the understanding and tackling of broader psychological impacts of the pandemic. It also refers to issues such as lockdowns, information overload, and social isolation.

While governmental communication strategies and official social media interventions have been shown to mitigate public panic to some extent, these measures primarily address information dissemination rather than the underlying psychological distress experienced by individuals. Prolonged lockdowns, social isolation, and continuous exposure to distressing content have compounded mental health challenges, highlighting the limitations of conventional policy-driven and awareness-based interventions. These situations highlight the necessity of systematic, scalable, and objective methods for early detection and monitoring of mental health disorders like anxiety and depression.

In this regard, data-driven computational methods present a viable way to deal with the wider psychological effects of pandemics. By leveraging ML and artificial intelligence (AI) methodologies, it becomes possible to analyze large-scale mental health data, uncover latent patterns associated with psychological distress, and support timely screening beyond traditional clinical settings. This necessity motivates the present research to explore advanced ML, DL, and hybrid intelligence-based frameworks capable of robustly detecting depression and anxiety in post-pandemic scenarios.

1.2 Research Gaps

There is still a dearth of recent, thorough, and methodically organised literature reviews that critically examine and contrast ML and DL approaches for the detection and prediction of anxiety and depression, despite the expanding body of research on mental health assessment during the COVID-19 pandemic. The following research gaps have been discovered and are described in this section based on a critical study of the body of existing literature and current research trends.

1.2.1 Limited Literature Reviews

The lack of fresh, comprehensive literature reviews on ML and DL models used to detect and forecast anxiety and sadness brought on by the COVID-19 pandemic is the research gap.

1.2.2 Lack of Analysis of Psychological Assessment Tools

There is an absence of detailed analysis of the commonly used psychological assessment tools in this context, indicating the need for a more specialized investigation.

1.2.3 Lack of Comparative Examination of ML/DL Model Performance

A thorough assessment and comparison of ML and DL models for the detection of anxiety and depression is lacking in current research.

1.2.4 Absence of Post-COVID Mental Health Dataset

There is no post-COVID mental health dataset available, making it difficult to assess the actual long-term impact of COVID-19 on mental health.

1.2.5 Need for Improved Model Development

Developing an improved model presents an opportunity for more accurate detection of depression and anxiety.

1.2.6 Need to Analyze ML Performance on Varying Dataset Sizes

There is a need to analyze ML model performance across varying dataset sizes in the context of anxiety and depression detection.

1.3 Problem Statement

The key problem addressed in this research is the lack of sufficiently robust and reliable ML and DL approaches for anxiety and depression detection that adequately address challenges related to model performance, interpretability, dataset size, and the use of sensitive mental health data. This research therefore investigates and develops ML and DL based approaches to improve the effectiveness and reliability of automated anxiety and depression detection while addressing the identified research gaps.

1.4 Overview of the Proposed Methodology

The methodology proposed in this research involves the use of ML, DL, and hybrid methods for the classification of depression and anxiety severity. Data for the study were collected using the Beck depression inventory (BDI) and Beck anxiety inventory (BAI) instruments, and the resulting questionnaire responses were used as input to various ML and DL models for severity classification. As part of this research, three models, namely DA-TabSVM, ASKNN, and FW-AK-KNN, were proposed for the classification of depression and anxiety severity.

1.5 Research Objectives and Contribution of the Thesis

This thesis's main goal is to create and assess ML and DL based models for the precise identification and forecasting of anxiety and depression, utilising psychological screening tools that have been validated. The thesis achieves the following objectives:

- **Objective 1: Literature Review** - To conduct a systematic literature review and analysis of existing research on anxiety and depression detection employing machine learning and deep learning models.

  - **Contribution:** In order to assess the advantages and disadvantages of different ML and DL techniques for identifying anxiety and depression, the suggested thesis does a thorough, methodical literature analysis. (Completed by Journal 2 and Conference 3)

- **Objective 2: Comparison of Existing ML/ DL Techniques** - To study, implement, and analyze the performance of Deep Learning/Machine Learning classification models: Decision Tree, Random Forest, Support Vector Machine, KNN employed for anxiety and depression detection induced during COVID-19.

  - **Contribution:** Four ML algorithms for mental health assessment—RF, DT, SVM, and KNN—are compared in this work. The study emphasises the importance of dataset size in improving ML models' predictive skills for reliable and accurate evaluations. This work explores the application of DL techniques in the analysis of depression and anxiety, aiming to enhance understanding, diagnosis, and treatment strategies. Five distinct DL models—namely, NN, CNN, RNN, LSTM, and TabNet are evaluated and compared within the context of depression and anxiety analysis on three different datasets in conference paper 1. (Completed by Journal 1 and Conference 1).

- **Objective 3: Improved Algorithm / Novel Approach** - To study and analyze the impact of COVID-19 on ongoing college students employing improved Deep Learning/Machine Learning model for depression and anxiety detection.

  - **Contribution:** The proposed work implements existing algorithms DT+kNN and RF+NN to efficiently predict mental health conditions like depression and anxiety using data from psychological instruments. It also introduces DA-TabSVM, a novel hybrid model that combines TabNet with SVM for the detection of depression and anxiety disorders. The suggested hybrid strategy consistently achieved the maximum precision, recall, F1-score, and accuracy in the performance study of the DA-TabSVM model in identifying depression and anxiety using three datasets, outperforming or matching traditional ML and other hybrid models. KNN and extensions

have been extensively used in disease detection. In this work, ASKNN is proposed, an improved version of KNN that adapts neighbor influence dynamically with an adaptive sigma parameter, enhancing predictive accuracy and stability in mental health classification. This work introduces a FW-AK-KNN model that improves upon the traditional KNN approach by integrating feature importance into the distance calculation and applying adaptive kernel-based weighting to neighbors, thereby enhancing both predictive accuracy and interpretability in mental health prediction. (Completed by Conference 2, Journal 3, Journal 4, and Journal 5).

1.6 Significance of the Thesis

This thesis is important because it can use cutting-edge ML and DL techniques to improve the detection and prediction of anxiety and depression. This work seeks to contribute to both theoretical understanding and practical applications by filling in the existing research gaps, such as the lack of post-COVID mental health datasets and the restricted comparative studies of ML/DL models.

The outcomes of this thesis can aid mental health professionals in early detection and intervention, inform policy decisions, and provide a framework for integrating federated learning approaches in privacy-preserving mental health systems. Additionally, it improves comprehension of model performance across different dataset sizes, guaranteeing improved generalisation and practicality.

1.7 Thesis Overview

This thesis is organised to give readers a thorough grasp of the study problem, methods, and results. There are eight chapters in this thesis. The impact of the COVID-19 pandemic on mental health and the rationale for this study are covered in Chapter 1, *Introduction*. The literature review of earlier studies on the identification of anxiety and depression is included in Chapter 2, *Literature Review*. It also addresses gathering the underlying information and insights needed to identify anxiety and despair after the COVID-19 epidemic. It describes the psychological instruments, types

of depression and anxiety, and various techniques used to diagnose depression and anxiety. Furthermore, it presents the ML and DL models used to automate depression and anxiety detection, including different data processing techniques for psychological instruments and the various ML and DL based models. It explains the publicly available datasets for the use of psychological tests to identify anxiety and depression. Chapter 3, *Unveiling the Impact of Dataset Size on Machine Learning Models for Anxiety and Depression Prediction amid the COVID-19 Pandemic: Determining Optimal Data Collection Thresholds*, addresses the problem of determining the optimal dataset size required for effective use of ML and DL models.

Chapter 4, *Exploring Deep Learning Architectures for Mental Health Assessment: A Comparative Analysis*, compares existing DL architectures used for depression and anxiety detection. Chapter 5, *Enhancing Mental Health Diagnosis with DA-TabSVM: A Multi-Class Hybrid Approach for Detecting Depression and Anxiety*, proposes an improved hybrid model for depression and anxiety detection and introduces a novel model, DA-TabSVM.

Chapter 6, *A Novel Adaptive Sigma KNN Model for Depression and Anxiety Detection Following the COVID-19 Pandemic*, presents a novel model, ASKNN, for depression and anxiety detection. Chapter 7, *Feature-Weighted Adaptive Kernel KNN: A Novel Approach for Post-COVID-19 Depression and Anxiety Detection*, proposes a novel model, FW-AK-KNN, for depression and anxiety detection. This model uniquely identifies the important features that contribute to depression and anxiety detection using RF for feature ranking, which is further fed to the proposed model for depression and anxiety detection. . Finally, Chapter 8, *Conclusion, Future Work & Social Impact*, summarizes the conclusions of this research, highlights potential future work, and discusses the social applications of the findings.

Chapter 2

Literature Review

2.1 Introduction

Among the most common psychiatric problems in the world are mental health disorders, especially anxiety and depression. Their prevalence has increased recently, and the COVID-19 epidemic has made psychological anguish worse because of social isolation, health-related anxieties, and unstable finances [21,179]. In addition to having a negative impact on an individual's well-being, these diseases have a substantial negative impact on society, the economy, and healthcare, underscoring the importance of early detection and treatment [91].

A comprehensive understanding of depression and anxiety, including their types, clinical manifestations, and associated risk factors, is essential for developing reliable detection mechanisms. In addition to traditional clinical assessments, various screening and diagnostic approaches have been explored, ranging from self-report psychological instruments to physiological and behavioral measures such as electroencephalogram (EEG), social media-based sentiment analysis, and Magnetic Resonance Imaging (MRI). While these methods provide valuable insights, they often suffer from limitations such as subjectivity, reporting bias, and limited accessibility and scalability.

Recently, considerable efforts have been made to employ ML and DL paradigms for automatic mental health evaluation. Many researchers have explored the potential of conventional classifiers, hybrid systems, and complex learning architectures to enhance predictive performance and reliability. This chapter aims to conduct a systematic review

of these methods, specifically focusing on their implementation during and after the outbreak of the COVID-19 pandemic.

This chapter analyzes the methodologies, data sources, feature extraction methods, and evaluation procedures adopted by existing works. The objective is to highlight current research trends and challenges faced by existing methods, as well as open problems. Despite notable advancements, existing approaches often face limitations related to handling tabular data, model generalization, interpretability, and performance consistency across diverse datasets.

Before proceeding further, it is important to understand depression and anxiety and their types in detail, which are discussed in the following section.

2.2 Types of Depression and Anxiety

Anxiety and depression are two of the most common mental health conditions that many individuals worldwide are currently dealing with. Both of these issues occur in different forms, and learning about these can help a person identify whether he or she is dealing with depression and/or anxiety. Fig. 2.1 provides a schematic overview of the classification of depression.

2.2.1 Types of Depression

Feelings of melancholy, a lack of interest in activities, and other symptoms are indicative of depression, which is classified as a mood illness [21]. Depression can also be described by the types below:

2.2.1.1 Major Depressive Disorder (MDD)

Major Depressive Disorder (MDD) is the most frequent and severe type of depression. This condition can be defined as a set of both emotional, cognitive, and physical symptoms, such as feeling sad, lacking energy, feeling worthless, having impaired sleep, and finding difficulty concentrating. These symptoms persist for at least two weeks. Additionally, this condition is recurrent [179].

2.2.1.2 Persistent Depressive Disorder (Dysthymia)

This type of depression is a less severe kind of depression that persists for at least two years. Sufferers of this type of depression have low mood, low energy levels, low self-esteem, and problems carrying out daily activities. Although the symptoms of Persistent Depressive Disorder are not severe, the chronic form of the condition affects the quality of life of the patient.

2.2.1.3 Postpartum Depression

Postpartum Depression is a condition that affects women after childbirth. Symptoms of postpartum depression include extreme sadness, anxiety, irritability, and changes in sleep and appetite.

2.2.1.4 Seasonal Affective Disorder (SAD)

Seasonal Affective Disorder (SAD) is also a kind of depression that is linked to the seasons. It mainly occurs when people do not get enough sunlight, which is usually when winter is around. There are also symptoms of sadness, lethargy, and a craving for carbs.

2.2.1.5 Other Specified Depressive Disorders

This category would include depression types that are distressing to the patient but do not fit into any of the aforementioned categories. This category would include depression varieties that cause the patient anguish but do not fit into any of the aforementioned categories.

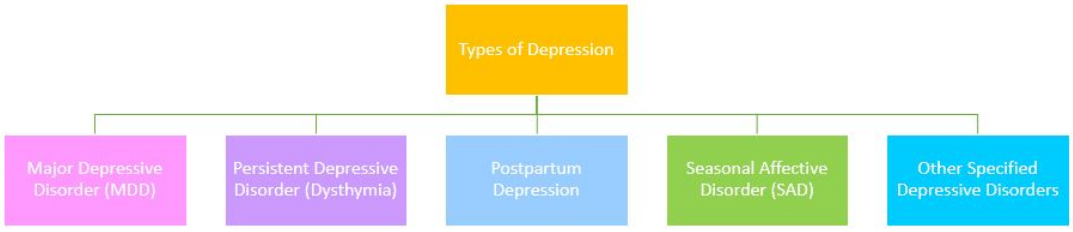


Figure 2.1: Types of depression.

2.2.2 Types of Anxiety

Anxiety disorders refer to excessive fear, worry, or nervousness that can substantially interfere with everyday functioning. A diagrammatic overview of depression classification is shown in Fig. 2.2. Typical kinds include:

2.2.2.1 Generalized Anxiety Disorder (GAD)

GAD involves excessive worry about various things in life, such as job, health, or relationships. The symptoms of GAD involve restlessness, muscle tension, fatigue, irritability, and problems with concentration.

2.2.2.2 Panic Disorder

In panic disorder, there are sudden episodes of attacks of panic, and these are intense moments of fear, along with palpitations, sweating, trembling, difficulty breathing, and a sense of doom. These occurrences are typically unanticipated and may have an impact on the person's quality of life.

2.2.2.3 Social Anxiety Disorder (Social Phobia)

Excessive fear and avoidance of social interactions brought on by a fear of mockery or condemnation are characteristics of social anxiety disorder. It manifests with signs such as blushing, sweating, palpitations, and avoidance of social situations. It can manifest in two different forms: generalized social phobia and specific phobias.

2.2.2.4 Specific Phobias

An excessive and illogical fear of particular things and circumstances is known as a specific phobia. Spiders, flying, and heights are a few examples of specialised phobias.

2.2.2.5 Other Anxiety Disorders

Anxiety disorders that do not fit neatly into the above categories but still cause significant effects on functioning include separation anxiety, selective mutism, and anxiety due to medical conditions.



Figure 2.2: Types of anxiety.

2.3 Risk Factors Associated with Depression and Anxiety

There are several famous risk factors for unhappiness and anxiety in the field of psychopathology, but their etiology is complex. Genetics is a significant factor that brings the hereditary character of many diseases into focus. People are especially susceptible if there is a family history of depression or anxiety. The actual chemical imbalance in the brain may have a lot to do with these diseases, particularly defects in the neurotransmitters related to serotonin, dopamine, and norepinephrine that control mood regulation. Trauma is another ascertained risk factor, which can be the result of a traumatic accident, emotional or physical abuse, or any other distressing event. The psychological impact on the individual owing to the trauma associated with such an incident has deep-rooted involvement with the development of disorders like PTSD. Finally, chronic stress resulting from life stresses like work strain or interpersonal conflicts has been found to increase the risk factor. Long-term stress can alter the morphological and functional structure of the brain. These aspects relate to each other to show how complex the etiology of depression and anxiety can be and why treatment should be multi-dimensional and preventative in nature.

2.4 Screening Techniques

There are varying tools to screen for depression. The psychological instruments can be used in screening for depression as well as MRI, EEG, visual review based on facial expression, Heart Rate Variability (HRV), biological markers, and social media through

sentiment analysis, as shown in Table 2.1. EEG is a signal meant for the detection of activity in the brain. This activity is a diagnostic indicator for depression. EEG can detect depression because it senses emotional fluctuation and abnormal brain activity, two common features of depression [36].

Fig. 2.3 shows various techniques used to diagnose depression and anxiety. This review focuses on studies that used psychological instruments for detecting depression and anxiety.

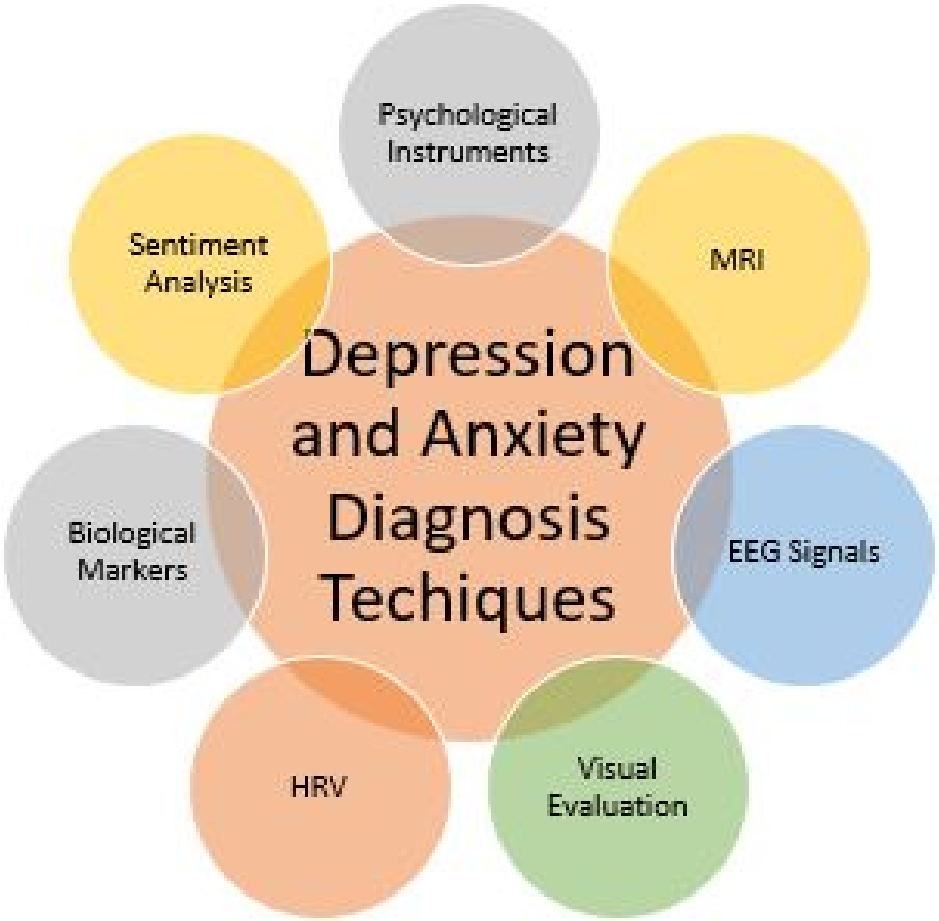


Figure 2.3: Anxiety and depression are diagnosed using a variety of methods.

Clinical examination is usually used in clinical settings to establish anxiety diagnoses, including interviews with experts, self-report questionnaires, observation of symptoms and behaviors, and adherence to the diagnostic criteria described in manuals, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). MRI and EEG can offer further insight into the neural correlates and brain areas associated with anxiety, and therefore, the neurobiology that underlies this condition can be understood

in detail. Because the abnormalities in structure and function on an MRI could be manifestations of other biological processes or behaviors, it is impossible for doctors and psychiatrists to diagnose any patient with MDD using that particular scan. Thus, MRI scans are ineffective for diagnosing most mental disorders [270].

Table 2.1: Techniques used to detect mental health problems

Method	**Description**	**Strength**	**Weakness**
Psychological Instruments	These include self-report questionnaires and structured interviews designed to assess various aspects of mental health, such as depression and anxiety.	Widely validated and accepted. Can be easily administered. Cost-effective. Useful for large-scale screenings.	Subject to self-report bias. May not capture all aspects of mental health. Can be influenced by the individual's current state or willingness to be truthful.
MRI	An imaging technique used to visualize detailed internal structures, including the brain, helps in identifying structural and functional abnormalities associated with mental health disorders.	Provides detailed images of brain structure. Helps identify structural and functional abnormalities. Can aid in understanding the neurobiology of disorders.	Expensive. Not effective for diagnosing most mental disorders alone. Structural abnormalities may be due to other factors.

Continued on next page

Method	Description	Strength	Weakness
EEG	A technique for capturing brain electrical activity that can be used to identify aberrant brain activity and emotional swings that could point to mental health problems.	Measures brain electrical activity. Can detect emotional fluctuations and abnormal brain activity. Useful for identifying neural correlates of disorders.	Requires specialized equipment and expertise. May not pinpoint specific disorders without additional context.
Facial Expression Evaluation	A technique that uses visual analysis of facial expressions to infer emotional states, often enhanced with AI and ML algorithms.	Non-invasive. Can be automated with AI. Provides real-time data.	May be influenced by cultural differences. Limited by the individual's ability to express emotions visibly.
HRV	Measurement of the variation in time between heartbeats, which can indicate stress and emotional regulation.	Non-invasive. Provides real-time physiological data. Can indicate stress and emotional states.	Influenced by physical activity and other physiological factors. May not be specific to mental health disorders.
Biological Markers	Biological indicators (e.g., hormones, neurotransmitters) that may be associated with mental health conditions, provide a potentially objective measure of mental health.	Potentially objective indicators of mental health. Can be used in conjunction with other methods.	Still under research for many mental health conditions. May require invasive procedures (e.g., blood tests).

Continued on next page

Method	Description	Strength	Weakness
Sentiment Analysis (social media)	Analyzing text data from social media to infer emotional states and mental health trends based on language use.	Can analyze large amounts of data. Provides insights into real-time mental states. Useful for detecting population-level trends.	Privacy concerns. Data may not be representative of clinical conditions. Subject to variability in language use and context.
Clinical Assessments	Comprehensive evaluations are conducted by mental health professionals, including interviews, observations, and adherence to diagnostic criteria like those in the DSM-5.	In-depth and personalized. Can include expert interviews, observations, and standardized criteria DSM-5. Holistic approach.	Time-consuming. Requires trained professionals. Subject to human bias and variability in assessment quality.

Among these approaches, psychological instruments are selected as the primary data source due to their standardized structure, clinical validation, and suitability for large-scale, non-invasive mental health assessment.

2.5 Psychological Tools for Identifying Anxiety, Depression, and Other Mental Health Issues

Tables 2.2, 2.3, 2.4, and 2.5 provide an overview of several psychological tools. Fig. 2.4 shows a chord diagram demonstrating the author's relationship to the psychological instrument(s).

2.5.1 Beck depression inventory (BDI)

A test to gauge how severe depression symptoms are, which comprises 21 items. Every item receives a value between 0 and 3; higher scores indicate more severe depression. [17].

2.5.2 Beck anxiety inventory (BAI)

A 21-item test is used to gauge anxiety levels. Respondents score the degree to which each symptom afflicted them throughout the previous week. Higher scores indicate more anxiety, and they range from 0 to 63. [83].

2.5.3 Generalized anxiety disorder-7 (GAD-7)

A seven-question test for determining anxiety is called GAD-7. The cutoff values for mild, moderate, and severe anxiety vary, and the severity of symptoms is rated from 0 to 21 [226].

2.5.4 Patient health questionnaire-9 (PHQ-9)

The DSM-IV criteria are the basis for the 9-item depression questionnaire known as the Patient health questionnaire-9 (PHQ-9). It gauges how severe depression symptoms are, and results help clinicians make decisions about diagnosis and therapy [123].

2.5.5 Symptom checklist-90 (SCL-90)

The Symptom checklist-90 (SCL-90) is a 90-item broad-scale psychological measure assessing nine dimensions of symptoms, ranging from depression, anxiety, and somatization, to obsessive-compulsive symptoms. It is found in both research and clinical situations to measure overall psychological distress [27].

Table 2.2: Instruments Used to Detect Depression

S No	**Psychological Instrument**	**Description**	**Used by**

1	BDI	The BDI is a widely used self-report questionnaire designed to assess an individual's agreement with a set of statements on depression throughout the past two weeks in order to determine the intensity of their depressive symptoms [17].	[61, 143]
2	Depression, Anxiety, and Stress Scale (DASS-21)	The DASS-21 is a popular self-report questionnaire that evaluates a person's emotional health by gauging the severity of their feelings of stress, anxiety, and depression using a 21-item scale [145].	[58, 64, 128, 149, 244]
3	PHQ-9	The PHQ-9, which consists of nine questions about depression symptoms, is a popular self-report instrument for determining the severity of depression in individuals [123].	[37, 38, 46, 48, 63, 100, 140, 170, 185, 187, 188, 192, 235, 238, 242, 267, 268]
4	Center for Epidemiological Studies Depression Scale (CES-D 7)	A short questionnaire called CES-D 7 counts the frequency with which respondents reported having seven different depressed symptoms throughout the preceding week [204].	[22, 43, 55, 97]
5	Patient health questionnaire-4 (PHQ-4)	It is a very effective ultra-brief screening test, and it does so for both depression and anxiety disorders. It is comprised of a four-item composite measure that combines those two-item scales: the PHQ–2 and the GAD–2.	[26, 29, 183]
6	8-item Patient health questionnaire-8 (PHQ-8) scale	It is a self-report instrument intended to evaluate and quantify a person's level of depression [125].	[5]

7	Patient health questionnaire-2 (PHQ-2)	The PHQ-2 is a quick, two-question screening instrument used to determine whether a person has depression symptoms [122].	[93, 219]
8	Patient health questionnaire-15 (PHQ-15)	It alludes to the self-report questionnaire used to gauge how severe somatic symptoms or physical complaints are in various individuals. It is commonly used in assessments for somatic symptom disorders and related diseases. The 15 questions focus on somatic-related physical symptoms [124].	[187]
9	Depression Anxiety Stress Scale-9 (DASS-9)	It is a short questionnaire, which measures the amount of agreement of a person with statements that are connected to his/her degrees of depression, anxiety, and stress [264].	[131]
10	Hamilton Rating Scale	A questionnaire that is used to gauge depression levels in clinical and healthcare settings [131].	[121, 150]
11	Edinburgh Postnatal Depression Scale (EPDS)	To identify potential cases and quantify the strength of postpartum depression symptoms, the EPDS is used to screen the symptoms of postnatal depression in women after a delivery [52].	[166]

12	Zung Self-Rating Depression Scale (ZSDS)	Following the first few weeks of COVID-19, Fong et al. employed the State-Trait Anxiety Inventory (STAI) to measure anxiety and depression in American adults. ZSDS, also called the Zung Depression Inventory, is a self-administered survey used to measure an individual's level of depression. It was created by Dr. William W.K. Zung and is frequently utilised in research and therapeutic contexts. The scale aids in determining the presence and intensity of depression symptoms [276].	[73]
13	Brief Symptom Inventory-18 (BSI-18)	It is a survey for self-report that gauges an individual's emotional and psychological well-being by measuring their degree of suffering across a variety of symptoms, including somatization, anxiety, and depression [73].	[166]
14	Burns Depression Checklist (BDC)	BDC to predict depression. Among the 15 typical symptoms of depression that BDC asks respondents about are sorrow, low self-esteem, and hopelessness [34].	[274]

Table 2.3: Instruments Used to Detect Anxiety

S No	Psychological Instrument	Description	Used by
1	Spanish version of the STAI	By measuring how often a person agrees with statements about their symptoms, the STAI is a widely used self-report questionnaire that distinguishes between state anxiety (temporary or situational anxiety) and trait anxiety (general or permanent anxiety). This provides information regarding the person's anxiety levels both now and in the future [224].	[61, 73, 143, 219]
2	BAI	A self-report tool called the BAI is used to gauge how severe anxiety is in psychiatric populations [83].	[63]
3	Generalized Anxiety Disorder-7 (GAD-7)	A brief self-report questionnaire that assesses a person's agreement with statements regarding worry and anxiety in order to detect and assess the severity of anxiety symptoms. [226].	[8, 38, 48, 97, 100, 166, 170, 187, 188, 219, 238, 242]
4	Generalized Anxiety Disorder-2 (GAD-2) scale	A brief self-report questionnaire that is used to assess for GAD. It offers two key questions about excessive worry and restlessness during the past two weeks, and the responses are used to assess the respondent's level of anxiety symptoms [8].	[5]
5	State Anxiety Inventory (SAI)	A psychological assessment tool called the SAI is used to gauge someone's momentary, current level of anxiety [225].	[43]
6	Zung's Self-rating Anxiety Scale (SAS)	Dr. William Zung developed the Zung's SAS, a psychological assessment instrument that measures an individual's degree of anxiety symptoms using a self-report questionnaire [275].	[140, 192, 252]

Table 2.4: Instruments Used to Detect Stress

S No	Psychological instrument	Description	Used by
1	Perceived Stress Scale (PSS)	A self-report questionnaire known as the PSS measures to what extent the person agreeing with statements about their perceived level of stress and their capacity to handle life's obstacles will determine how much they think stress exists in their lives [49].	[63, 137]
2	10-item Perceived Stress Scale-10 (PSS-10)	It is a psychological measurement tool that assesses the degree to which a person perceives that they have experienced stress over the past month; that is, how unexpected, uncontrollable, and overwhelming a person considers their situation in life to be [164].	[71]
3	Coronavirus Perinatal Experiences (COPE questionnaires)	COPE surveys have been created to evaluate how the COVID-19 pandemic has affected perinatal experiences, including mental health, anxiety, and stress throughout pregnancy and the postpartum period [237].	[166]
4	Symptom Check List (SCL-5)	For predicting psychological distress [27].	[14]

Table 2.5: Instruments Used to Detect Other Conditions

S No	Psychological instrument	Description	Used by
1	Suicidal Behaviors Questionnaire-Revised (SBQ-R)	A self-report instrument called the SBQ-R is used to evaluate suicidal thoughts and actions. It was created to assess suicide risk in research and therapeutic contexts [178].	[149]
2	Insomnia Severity Index (ISI) scale	A popular self-report tool called ISI is used to evaluate the impact and severity of insomnia [15].	[48, 149]

Continued on next page

S No	Psychological instrument	Description	Used by
3	Connor–Davidson resilience scale (CD-RISC)	CD-RISC is a psychological test. It measures resilience as the ability to deal with and adapt to adverse circumstances, stress, and personal challenges [51].	[238]
4	Support Questionnaire (SSQ)	SSQ is a tool used to measure the perceived social support of individuals [207].	[238]
5	Diagnostic Criteria for Psychosomatic Research (DCPR)	It is a diagnostic system designed to detect psychosomatic syndromes and symptoms not yet fully addressed by the DSM or International Classification of Diseases [184].	[238]
6	Self-rated mental health (SRMH)	It is an assessment tool used to investigate a person's view about their entire mental health and general mental well-being through self-evaluation and rating [136].	[55, 129]
7	Resilience Scale	The Resilience Scale evaluates a person's capacity for adjustment and recovery from stress or adversity, offering information on how resilient they are [107].	[173]
8	Adult Hope Scale	The Adult Hope Scale measures the ability of individuals to set and work toward goals as well as their confidence in their ability to achieve the same to measure their level of hope [117].	[173]
9	Spiritual Wellbeing Scale	A person's feeling of purpose, his or her relation with a higher power, and the level of total spiritual satisfaction were measured through the Spiritual Wellbeing Scale developed by Paloutzian and Ellison [3].	[173]

Continued on next page

S No	Psychological instrument	Description	Used by
10	Scale of Positive and Negative Experience 8 (SPANE-8)	SPANE-8 gauges an individual's favourable and unfavourable experiences over a certain time frame created by [132].	[131]
11	The Satisfaction with Life Scale (SWLS)	SWLS is a self-report questionnaire that gauges an individual's subjective well-being and overall life satisfaction by determining how much he agrees with statements about his life and happiness [57].	[131]
12	Mental Health Continuum–Short Form (MHC–SF)	MHC-SF is an acronym for the brief self-report questionnaire designed to assess mental health and wellness as well as people's emotional, social, and psychological functioning on a spectrum from sluggish to flourishing [113].	[131]
13	DSM-5	The DSM-5 criteria are employed in the design of the self-report questionnaire commonly known as the PCL-5. Such a self-report questionnaire thus becomes an important tool in establishing, in addition to monitoring, a diagnosis for PTSD [28].	[37, 166]
14	Impact of Events Scale-Revised (IES-R)	The IES-R is a self-reporting tool used to assess the degree of discomfort associated with symptoms following a specific traumatic event, such as intrusion, avoidance, and hyperarousal [256].	[58]

Continued on next page

S No	Psychological instrument	Description	Used by
15	COVID-19 Impact Survey	The COVID-19 Impact Survey is a tool for gathering information on various facets of everyday life and well-being while investigating the effects of the COVID-19 pandemic on people and communities [53].	[105, 199]
16	Yale-Brown Obsessive Compulsive Scale (Y-BOCS)	It is a standardized clinical evaluation tool used to assess the severity of OCD symptomatology. It assesses the frequency and severity of obsessions and compulsions and provides an organized approach to the assessment and follow-up of symptomatology in OCD [79].	[61, 245]
17	Brown Assessment of Belief Scale (BABS)	It is a clinical rating scale that helps doctors assess the seriousness and level of impairment of delusional beliefs upon general mental health [62].	[61]
18	Fatigue Severity Scale (FSS)	The FSS is a self-report questionnaire that is often used to measure levels of physical and mental fatigue by assessing how much people agree with fatigue-related statements [213].	[159]
19	PTSD (Posttraumatic Diagnostic Scale-Chinese) (PDS-C)	A psychological diagnostic tool used to identify and rate the symptoms of PTSD [217].	[43, 140]
20	General Self-Efficacy Scale (GSES)	GSES measures a person's general self-efficacy—belief in their ability to deal with diverse complex situations [210].	[140]

Continued on next page

S No	Psychological instrument	Description	Used by
21	Self-reported Coping Style (SCS) scale	A psychological diagnostic tool used to gauge how well someone is managing stress, problems, and challenging circumstances [261].	[140]
22	Coping Orientations to Problems Experienced (COPE-NVI-25)	COPE-NVI-25 measures the coping mechanisms one uses to deal with challenges and problems in life [72].	[71]
23	Brief Self-Control Scale (BSCS)	BSCS assesses an individual's capacity for self-regulation regarding emotions, impulses, and behaviors [234].	[71]
24	Locus of Control (LOC) Scale	LOC Scale measures whether an individual believes life events are controlled internally or by external forces [172].	[71]
25	10-item Big Five Inventory (BFI-10)	The BFI-10 assesses the Big Five personality traits using 10 items [84].	[71]
26	Defense Mechanisms Rating Scale - Self Report (DMRS-SR-30)	It is a psychological assessment tool measuring defense mechanisms used to cope with psychological stressors [56].	[187]
27	Stanford Professional Fulfillment Index (PFI)	It is a survey method designed to gauge healthcare workers' resilience, job happiness, and professional fulfilment [243].	[144]

Continued on next page

S No	Psychological instrument	Description	Used by
28	SF-20 Quality of Life scales	The SF-20 is a short 20-item quality-of-life questionnaire measuring physical, emotional, and social well-being [33].	[235]
29	Mental Health Inventory (MHI-5)	Measures emotional distress and mental well-being using 5 items [198].	[235]
30	Maslach Burnout Inventory (MBI)	A structured questionnaire measuring burnout, especially in the workplace [155].	[236]
31	Impact of Event Scale (IES)	A psychological instrument assessing subjective distress from traumatic events [256].	[236]
32	Pittsburgh Sleep Quality Index (PSQI)	A survey assessing sleep disruptions and quality during the preceding month [35].	[242]
33	Demand-Control-Support Questionnaire (DCSQ)	A survey tool that evaluates workplace stress by measuring job demands, control, and social support [110].	[242]
34	General Health Questionnaire (GHQ-12)	A 12-item screening tool for psychological distress and mental disorders [59].	[242]
35	Utrecht Work Engagement Scale	Measures vigor, dedication, and absorption in work tasks [209].	[242]

Continued on next page

S No	Psychological instrument	Description	Used by
36	Kessler 6 Scale (K-6)	A six-item assessment for mental health issues like anxiety and sadness [75].	[263]
37	Self-designed questionnaire	A custom-created survey tool tailored to gather specific information relevant to a study.	[24, 25, 47, 86, 99, 106, 109, 127, 162, 171, 174, 176, 194, 212, 255]

In all, researchers have used a range of depression and anxiety measures and questionnaires to evaluate mental health in various demographics and circumstances. These tools are as fundamental to understanding and solving the psycho-emotional problems in clinical, occupational, or pandemic-related situations and offer important information for the provision of support and solutions. These techniques are still being used by researchers to provide light on the different ways that the COVID-19 epidemic affects people's mental health worldwide.

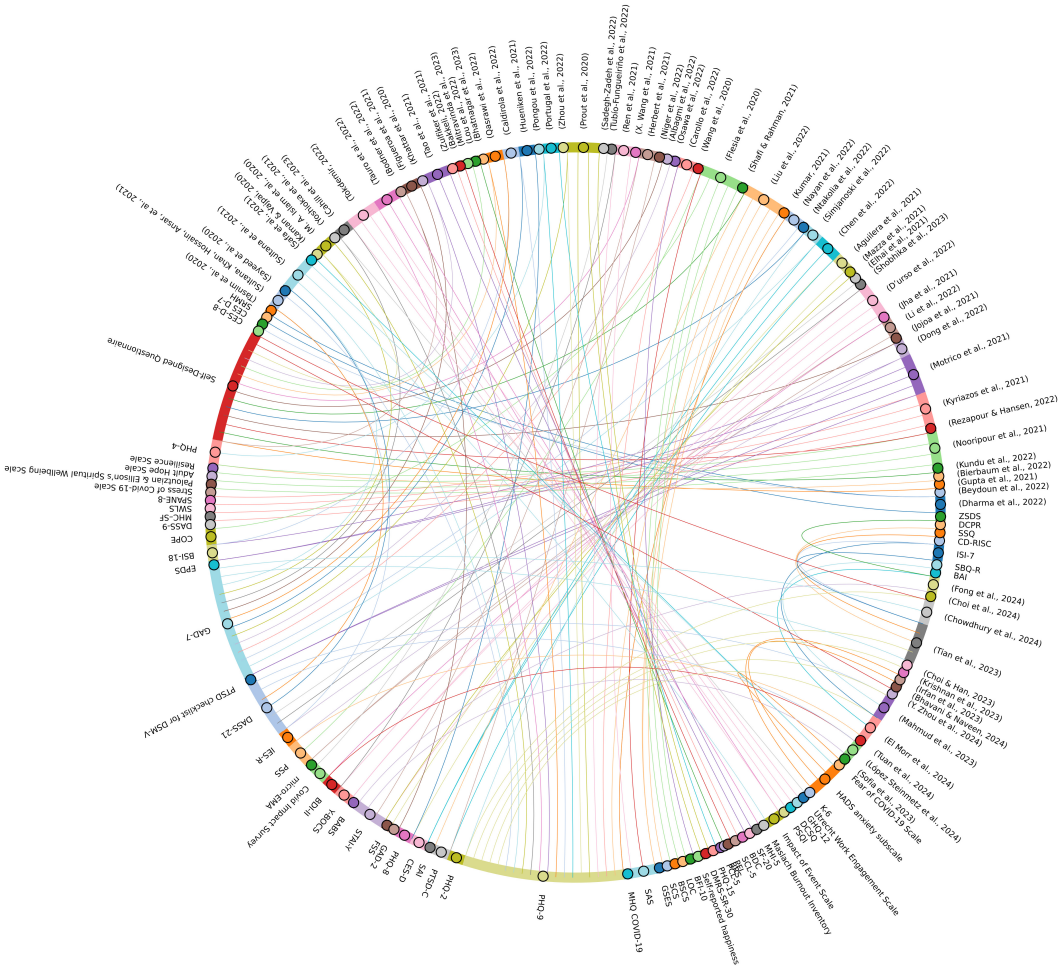


Figure 2.4: A chord diagram demonstrating the author's relationship to the psychological instrument(s).

2.6 Survey Methodology

This section describes in detail the methodological procedures that underlie our selection and exclusion criteria for sources involved in this work. This section also presents our well-articulated research questions that guide this investigation.

2.6.1 Review Methodology

The purpose of this section is to thoroughly discuss how COVID-19 affects mental health, as well as the use of ML and DL techniques in the diagnosis of depression and anxiety. The authors outline the methodical approach to the literature review in

this section. A systematic review employing Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria is being used to screen the studies for comprehensiveness. Two components of the PRISMA criteria are used to include studies in this review. Research studies that demonstrate the prevalence of mental health issues caused by the COVID-19 pandemic make up the first set of requirements. The research that employ ML and DL for depression and anxiety fall under the second criterion. With access to the following databases of journals, a broad survey was carried out: Science Direct, IEEE Xplore Digital Library, Springer Link, Google Scholar, and ACM Digital Library.

Fig. 2.5 displays the PRISMA type of representation used for the research paper selection. This work carefully examined and narrowed down 66 research publications that particularly address the detection of sadness and anxiety from an initial pool of over 700 research papers spread over the chosen databases.

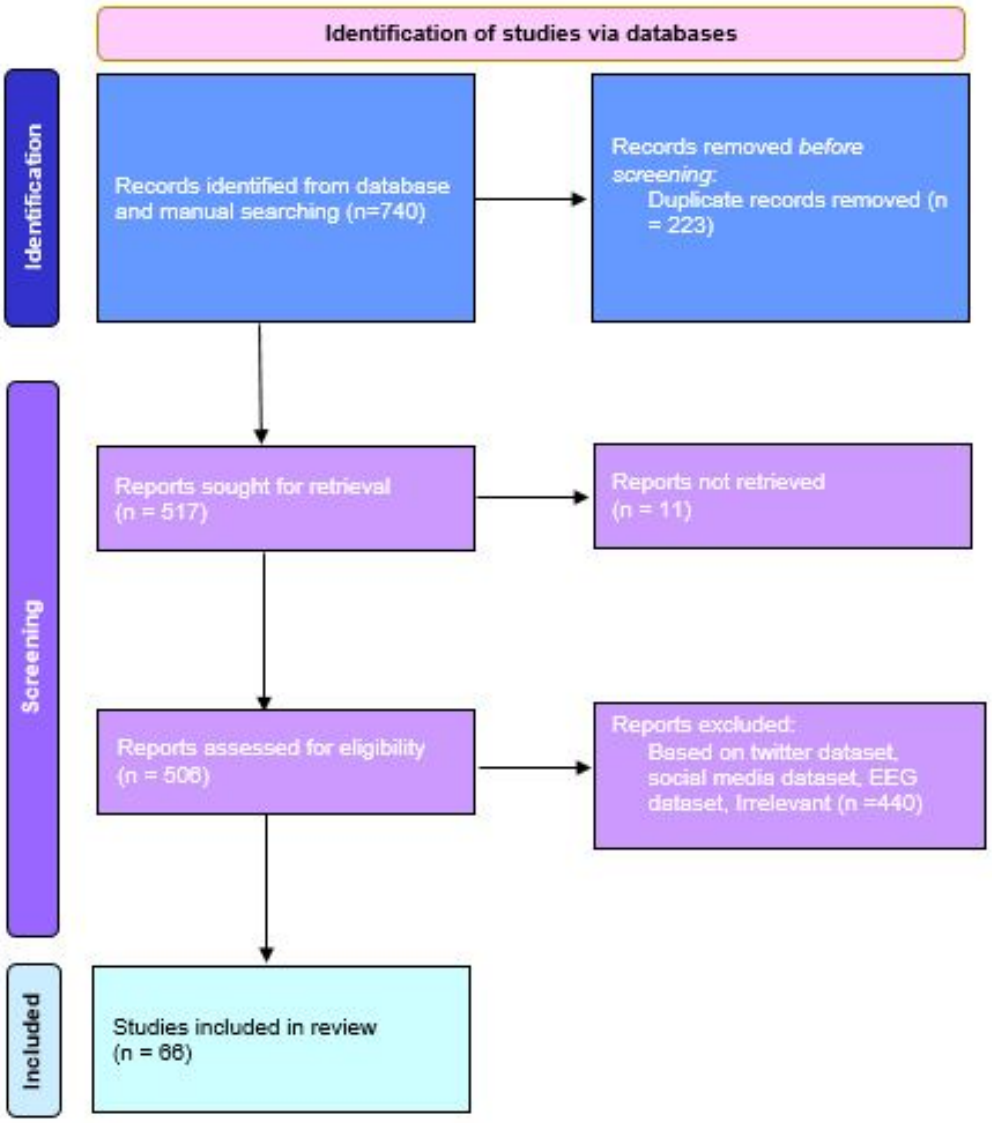


Figure 2.5: PRISMA flow diagram for study selection.

2.6.2 Statistical Insights

The keywords or search terms employed in these databases encompass the following: Depression, Anxiety, Covid-19, Machine Learning, and Deep Learning. This research mostly employs advanced ML and DL techniques in conjunction with questionnaire-based evaluations (such as the BDI, BAI, and DASS-21). The breakdown of these 66 studies reveals a classification into 5 Systematic Literature Reviews (SLR), 57 employing ML methods, and 4 focusing on DL approaches. Out of the 66 publications that were ultimately chosen, 62 are journal papers, and 4 are conference papers. The 66 publications that were chosen include 44 that are included in prestigious databases like

SCI, SSCI, and ESCI, 11 that are listed in Scopus, and 4 that are taken from conference proceedings and 7 other articles, as shown in Fig. 2.6. This graphical representation delineates the diverse indexing sources that underpin our reviewed literature.



Figure 2.6: Distribution of Articles According to Indexing.

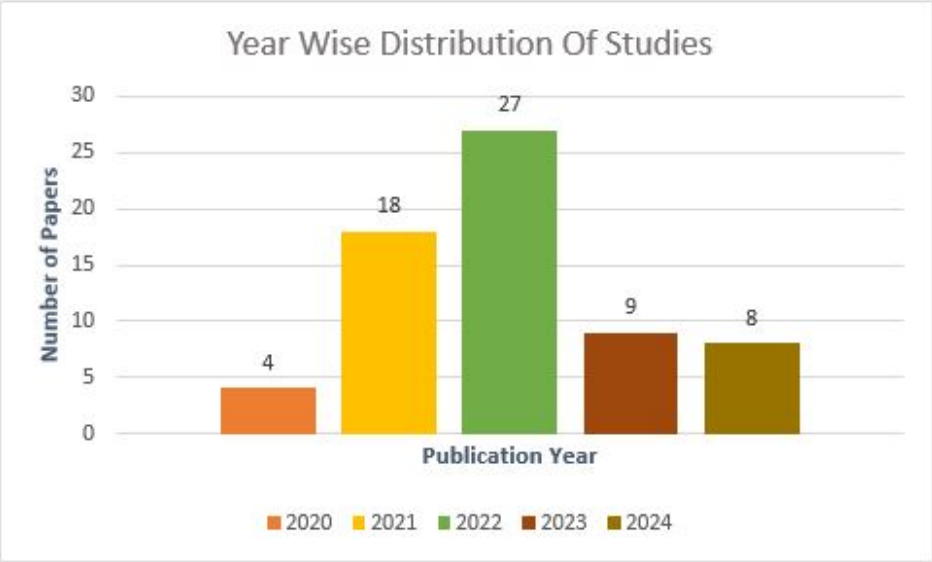


Figure 2.7: Year-wise distribution of studies.

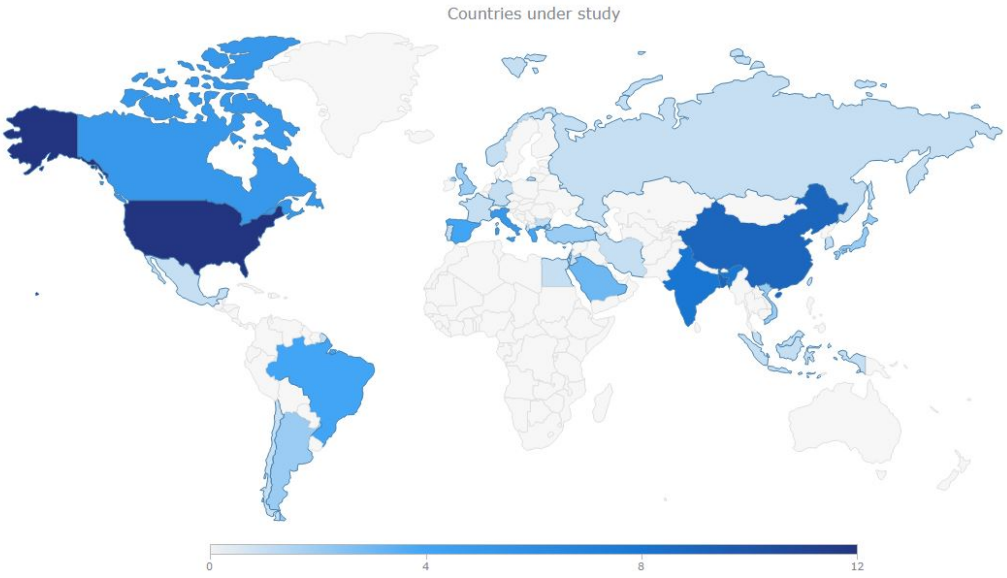


Figure 2.8: Countries under study.

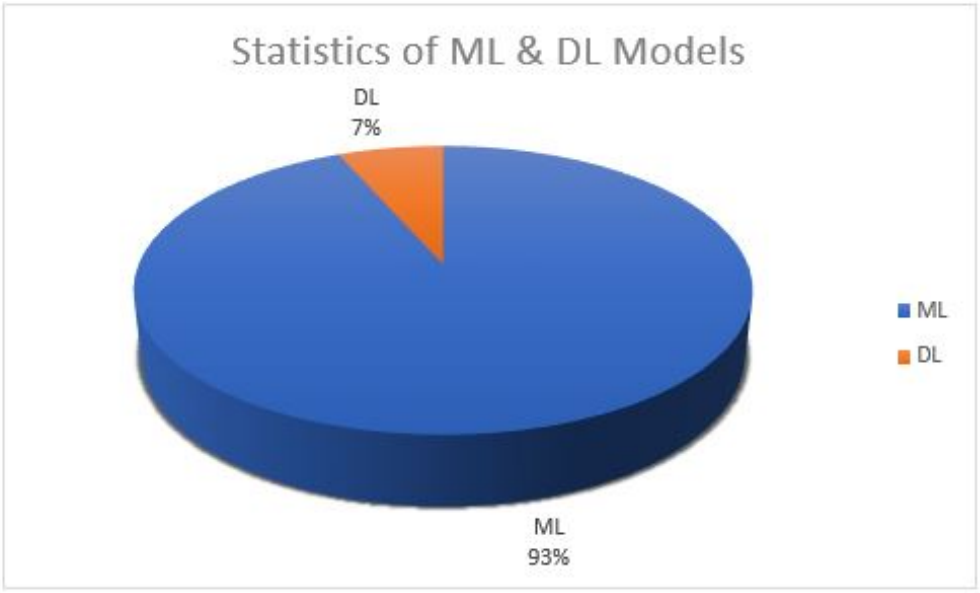


Figure 2.9: ML and DL models were used in earlier research to identify mental health issues.

Fig. 2.7 gives an in-depth depiction of the temporal distribution of research studies, including a comparison of conference and journal papers and the prevalence of each paper type by individual years. In the analysis, it was observed that the year 2024 had the lowest number of publications on the topic, with only eight articles, while the year 2022 saw the highest number of publications, with a total of thirty-two published articles.

The geographical distribution of research study locations is observed in Fig. 2.8. The research studies that were carried out in various nations are depicted on this map. China comes in second with nine studies, followed by Bangladesh and India with eight, and the United States with twelve. Significant contributions are also made by the UK, Canada, Italy, Brazil, Greece, and Spain. Several nations carried out two or fewer investigations, while many others only carried out one. This map highlights international cooperation in knowledge advancement while showcasing international research activities. While few research have employed DL models for the same purpose, ML models are widely used for the detection of mental health issues, as Fig. 2.9.

2.6.3 Research Questions

This work's main objective is to motivate up-and-coming researchers in the field. This work addresses various research questions crucial for the identification of depression and anxiety using ML and DL techniques, with several of these questions elaborated in Table 2.6. Furthermore, Table 2.7 examines survey papers concerning the detection of depression and anxiety through a comparative approach, with an emphasis on the research queries presented in these papers.

Table 2.6: Research questions related to depression and anxiety detection

S. No.	Research Questions
Q1	Is the mental health condition well-defined?
Q2	What is the COVID-19 pandemic?
Q3	Did previous pandemics impact the mental well-being of the population?
Q4	Has the COVID-19 pandemic deteriorated the mental health of an individual?
Q5	What are the methods and techniques used for detecting depression and anxiety?
Q6	What factors contribute to the risk of experiencing depression and anxiety?
Q7	What are the key advantages and challenges linked to the techniques used in the identification of depression and anxiety?
Q8	What are the different types of ML and/or DL techniques in depression and anxiety detection?
Q9	What are the performance evaluation measures for validating ML and/or DL-based depression and anxiety detection techniques?
Q10	What roles do ML and/or DL play in the context of identifying depression and anxiety?
Q11	Which datasets are frequently employed within the corresponding field?
Q12	How is depression and anxiety detection using ML and DL different from the other approaches?
Q13	What opportunities for future research are available in the field of detecting depression and anxiety using ML and/or DL?
Q14	Has the mental health of individuals been affected after the COVID-19 pandemic?

Table 2.7: Analyzing survey papers related to the detection of depression and anxiety through a comparative lens, with a focus on the research questions posed in these papers.

Survey	Review Year	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
Mittal et al., 2022	2022	✓	×	×	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	×
Chahar et al., 2021	2021	×	×	×	✓	✓	×	✓	✓	×	✓	✓	×	✓	×
Yogeswaran & Morr, 2022	2022	×	×	×	✓	×	×	×	×	×	✓	×	×	✓	×
Jiménez-Mijangos et al., 2022	2022	✓	×	×	✓	✓	✓	✓	✓	✓	×	×	×	✓	×
Aleem et al., 2022	2022	✓	×	×	×	✓	×	✓	✓	✓	✓	✓	×	✓	×
Proposed work	2024	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

2.7 Open Datasets

The effective development of computer algorithms for the detection of anxiety and depression requires access to a wide range of well-structured datasets. These datasets serve as the foundation for creating and assessing ML and DL models. They provide researchers with a variety of choices for experimentation because they differ in terms of the number of characteristics and tuples. This work presents a thorough overview of the many open datasets for depression and anxiety identification in Table 2.8. This section answers research question 11. This compilation provides information on the resources that are easily available to the scientific community to promote future research on this important topic.

Table 2.8: Various open datasets available for detecting depression and anxiety

S. NO.	Dataset Name	Dataset Link	Dataset Size (N)	The period of data collection	Psychological Instrument Used
1	Wanting to seek help during the pandemic for mental health and substance use concerns among the LGBTQI2S+ young adults dataset.	https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP3/DXFGAZ	1414	From December 2020 until March 2021	Score for past-week depression, past-year suicidal ideation, comorbidity, disability, stress, and general and mental health assessments
2	COVID Isolation on Sleep and Health in Healthcare Workers	https://www.openicpsr.org/openicpsr/project/127081/version/V1/view?path=/openicpsr/127081/fcr:versions/V1&type=project	842	March 28, 2020, to April 29, 2020	To determine whether healthcare professionals' sleep, emotions, and health-related behaviors are affected, 29 questions are asked.
3	Dataset of the COVID-19 lockdown survey conducted by GIPEyOP in Spain	https://doi.org/10.1016/j.dib.2021.107700	8387	April 28, 2020, to May 14, 2020	NA
4	Dataset of the COVID-19 post-lockdown survey conducted by GIPEyOP in Spain	https://doi.org/10.1016/j.dib.2021.107763	1175	September 23, 2020, to October 14, 2020	NA

Continued on next page

Table 2.8 continued

S. NO.	Dataset Name	Dataset Link	Dataset Size (N)	The period of data collection	Psychological Instrument Used
5	Covid Impact Survey	https://www.covid-impact.org/	17764	April 20, 2020 to April 26, 2020 and May 4, 2020 to May 10, 2020	Covid Impact Survey
6	Self-perceived loneliness and depression during the Covid-19 pandemic: a two-wave replication study	https://doi.org/10.5522/04/20183858	Wave 1 (n = 435), Wave 2 (n = 263)	April 17, 2020 to July 31, 2021	IPAQ-SF, Pittsburgh Sleep Quality Index, CASES, Sleepiness Scale, LQ, CHAOS, GAD-7, PHQ-9, aggression and personality questionnaires
7	mental-health-dataset	https://github.com/Hu-Li/mental-health-dataset	5108	June 1, 2020 to June 7, 2020	SCL-90, SAS, SDS, MMPI, EPQ, 16PF
8	Covid Impact Survey	https://www.covid-impact.org/	8790, 8975, 7506	April–June 2020	COVID-19 Impact Survey
9	Cross-sectional survey	pone.0277238.s002.dta	2756	July 2020 to October 2020	PHQ-4
10	cross-sectional survey	https://discover.schlesingergroup.com/covid19-mental-health.html	6021	May 2020 to December 2020	9 anxiety and depression questions
11	The Mental Health of Canadian Workers During COVID-19 Pandemic	https://osf.io/87vgs/	1576	April 21, 2021 to June 1, 2021	Self-rated mental health

Continued on next page

Table 2.8 continued

S. NO.	Dataset Name	Dataset Link	Dataset Size (N)	The period of data collection	Psychological Instrument Used
12	Pregnancy and Mental Health Data during COVID-19	https://doi.org/10.7910/DVN/FCDGEB	3569	August to November 2020	PHQ-9, GAD-7
13	OSMI Mental Health in Tech Survey	https://bitbucket.org/GowriSrinivasa/mental-health-in-tech/src/main/	1433	2016 to 2021	63 mental health questions
14	Depression-Repository	https://github.com/Sabab31/Depression-Repository	604	April and August 2020	BDC
15	National Health and Nutrition Examination Survey (NHANES)	https://www.cdc.gov/nchs/nhanes/index.htm	5398	NHANES 2013–2014	SF-20, PHQ-9
16	cross-sectional survey	https://doi.org/10.1371/journal.pone.0238162.s001	476	May 6, 2020 to May 12, 2020	PHQ-9, GAD-7

2.8 Applications of ML and DL in Mental Health Diagnosis

Applications of ML and DL offer benefits in the diagnosis of mental health, including automation, early detection, personalisation, and many other aspects that can combine a multitude of sources that have become more comprehensive and better traditional approaches to achieving and improving the sheer depth and substance of knowledge in this area. This section answers research questions 8 and 10.

2.8.1 ML Models Used for Detecting Depression and Anxiety

The goal of computer science is to provide guidelines and practices that allow a computer to study on its own and acquire knowledge [205]. ML transforms data into input for applications and algorithms that give them the ability to carry out activities that humans could complete. Over time, this data makes machines more accurate and helps them work better [201]. The three main recognized categories of ML are reinforcement learning, unsupervised learning, and supervised learning. This section answers research question 12. The typical workflow of any ML model contains Data Processing, Modelling, and Deployment.

For instance, accuracy values of 86% for anxiety and 84% for depression have been obtained in a comparison between DT, RF, and XGBoost by [48]. Irfan et al. [99] have found very high performance of GB, XGBoost, KNN, RF, DT, and LightGBM models; however, it should be noted that their study was performed on Saudi Arabian healthcare workers. DT, NB, RF, and SVM have been compared in the work of Bhatnagar et al. [24], achieving 78.9% accuracy with RF, but limited to a smaller data set. Similar findings on the comparison of several ML classifiers have been obtained in the work of Kumar et al. [128] and Mitravinda et al. [162]

Moreover, there are some constraints found in the reviewed articles. Smaller or homogeneous sample size, limited geography, subjective measures, and variations in measurement approaches could influence the validity and applicability of the models developed. As an example, Kundu et al. [129] used RF to predict mental health outcomes and found some limitations regarding sample size and non-response,

whereas Qasrawi et al. [188] showed potential ML performance in depression and anxiety screening; however, their results were constrained due to a limited dataset and incomplete screenings. This indicates that even high ML classification accuracy does not guarantee robustness of the ML model when detecting mental disorders. The section given below discusses various preprocessing techniques used in previous studies. Tables 2.9 and 2.10 shows an overview of ML models used to identify mental health issues.

2.8.1.1 Pre-Processing Algorithms

Pre-processing methods are essential for improving data quality because they reduce noise, remove redundancy, and convert raw data into a format that is appropriate for training and evaluating machine learning models.

Linear Discriminant Analysis (LDA): It is a statistical method for classifying data and reducing dimensionality in ML and statistics. It determines a linear feature combination that best divides two or more classes or groups in a dataset for tasks like pattern recognition and classification. LDA was proposed by R. Fischer in 1936 [70].

Principal Component Analysis (PCA): Dimensionality reduction is typically achieved in data analysis and ML using the statistical technique known as PCA. A dataset is transformed into a new coordinate system with orthogonal new axes (principal components) that are arranged by how much variance they explain. By identifying the most crucial details, cutting down on background noise, and supporting feature selection or visualization, PCA facilitates the simplification of complex data [153].

Synthetic Minority Oversampling Technique (SMOTE): SMOTE is a data augmentation technique used in ML that pays special attention to class imbalance in classification tasks. By interpolating between real cases, it creates artificial examples from the minority class such that the distribution of classes will finally balance out. This could be a crucial tool for helping machine learning learn from the minority class in order to enhance the classifier's performance on potentially uneven datasets [41,42].

2.8.1.2 Feature Engineering Methods

By converting unprocessed data into discriminative and meaningful features, feature engineering significantly improves the accuracy, efficiency, and generalisation of ML models.

SelectKBest One of the famous feature selection methods in ML is SelectKBest, which scores the value of each feature in a dataset to choose the top "k" features according to a predetermined scoring system. Using SelectKBest during preprocessing allows for more accurate and efficient selection of pertinent features within ML models, thus removing less useful variables to curtail dimensionality [134].

Particle Swarm Optimization (PSO) PSO is an optimisation algorithm that draws inspiration from the way flocks of birds and schools of fish behave naturally. PSO iteratively looks for the optimal solution in a problem space using a population of potential solutions known as "particles." Each particle moves with a velocity based on its experience and the surrounding particles' experiences [120]. Feature selection via PSO allows selection of the most relevant and informative characteristics, improving model performance and reducing dimensionality [255].

Maximum Relevance Minimum Redundancy (mRMR) mRMR focuses on emphasizing the most important and distinctive information while trying not to overlap or have similarity between features selected. By selecting a feature that are more pertinent and nonredundant, it aids in being able to improve model performance and reduce complexity [266].

Boruta Boruta is an approach used in ML to find and rank the most important features from a given dataset. It does so by comparing the importance of each feature in a dataset against a set of "shadow" features that are randomly created. A real feature is deemed important enough to be included in the ultimate subset of the feature if it can outperform its corresponding "shadow" versions significantly. In this method, Boruta emphasizes the most informative aspects of the data and ignores the less important ones, thus improving model accuracy [229].

RELIEFF RELIEFF stands for Relief-F, which is an algorithm used by ML tasks for data classification. It applies the process of how useful each feature is to the class labels of the dataset while predicting the value of the significance of each feature. RELIEF assigns a weight or score to each attribute in order to indicate its importance in discriminating between classes. The task of classification seems more relevant in the case that weights associated with a feature are relatively high compared to when their weights happen to be relatively low [118].

Feature Importance Ranking Measure (FIRM) FIRM approach is, in fact, a feature selection technique mainly applied to ML in order to examine and rank the relative contribution of each characteristic in a dataset. The goal of FIRM is to determine which characteristics have the biggest impact on the performance of a particular task or model [271].

Recursive Feature Elimination (RFE) It trains the model iteratively, until it has removed all the least important features, in order to reach the required number of selected features or a desired level of performance. RFE is particularly very useful when you have to handle high-dimensional datasets or you just need to identify the significance of features automatically [87].

Shapley Additive Explanation (SHAP) SHAP, or Shapley Additive Explanation, is an approach in both ML and explainable AI that attempts to shed light on feature contributions to the prediction given by a model. This is a namesake of the cooperative game theory concept called Shapley values, which determines how much each participant contributes to the success of a coalition. You can now rank the features that are in your model in terms of how much each one impacts predictions after including SHAP values in your model. The more a feature impacts predictions, the more it would be considered to contribute to the decisions made by the model [147].

Markov Blanket A Markov blanket of variables in a probabilistic model is such that while it is conditionally independent of all other variables in the model, it contains all the information sufficiently adequate to predict another specific variable. It simplifies the demanding probabilistic inferences and reasoning processes by keeping intact the

direct influences and the dependencies of the target variable. Additionally, Markov blankets can be utilised in feature selection to identify the most pertinent features for a given target variable [119].

CorrelationAttributeEval It is a feature selection method that ranks and assesses features according to how well they correlate with a dataset's target variable or class labels. This kind of feature selection makes use of filters [78].

Pearson's Correlation During feature selection, features are ranked according to their linear relationship to the target variable using Pearson's correlation coefficient. It then selects further, more robust linear relationships that are incorporated into the model, thus making the calculation smoother and model interpretability easier [211].

FP-Growth Feature selection techniques such as FP-growth are applied to pattern mining that reveals frequent patterns of feature combinations that correlate with the target variable. These patterns are thus considered when features have to be picked to model over, which helps in discovering and retaining meaningful feature sets for ML tasks [89].

2.8.1.3 Supervised Learning Classifiers

Supervised learning classifiers are machine learning algorithms that use labelled training data to anticipate the correlations between input features and class labels. New data instances are predicted into the established classes by the trained classifiers. Some examples of supervised learning classifiers are LR, DT, SVM, KNN, and NN. This section has been critically discussed in section 2.8.1. The section given below discusses various ML techniques used in previous studies. Tables 2.9 and 2.10 shows an overview of ML models used to identify mental health issues.

The application of supervised machine learning classifier techniques to the diagnosis of depression, anxiety, stress, and other mental health disorders has been studied before. Different researchers have looked at individual classifiers as well as their ensembles, and some of the most common types of such classifiers used include RF, SVM, DT, LR, KNN, and boosting. This variation in the claimed performance is due to the difference

in data sets, sample sizes, subjects, features, target states, and test methodologies employed. Thus, it may not be possible to compare the performance measures stated in different studies.

Naïve Bayes (NB) An NB Classifier is a straightforward ML method that divides data into groups or classes using the Bayesian principle. The NB Classifier is predicated on the idea that, given the class value, the attribute values are independent of one another. If the traits are interconnected, their effectiveness can be affected [89].

KNN KNN is a technique that finds new records by merging other historical data with the K most current historical data. To identify the K nearest neighbours, calculate the distance between the new sample and the training sample. Next, assess the category of the new sample and compare it with the categories of each neighbour. The new sample will also fall into the same category if all of the neighbours do [253].

KNN and Adaptive KNN Models K-Nearest Neighbour (KNN) is a widely used supervised learning algorithm that has attracted attention in healthcare because of its simplicity, adaptability, and effectiveness in classification tasks. The exponential growth of ML in healthcare has advanced the detection, diagnosis, and prognosis of diseases [40]. Conventional diagnosis often involves laboratory examinations, imaging technologies, and expert clinicians, which can be time-intensive, costly, and difficult to access in some developing communities [215]. Consequently, ML-based approaches have emerged as useful tools for computer-based and data-driven medical diagnosis [108].

KNN has been applied to the diagnosis of various diseases, including cancer, diabetes, cardiovascular diseases, and infectious diseases [130,146,167,203]. For breast cancer screening, KNN has been used to distinguish between benign and malignant tumors using histopathological images, mammography data, and gene expression patterns [45]. Similarly, KNN has been applied to diabetes prediction using risk indicators such as blood glucose, Body Mass Index (BMI), and insulin resistance. In cardiovascular disease diagnosis, KNN has been employed to classify patients using Electrocardiogram (ECG) signals, blood pressure, and lipid profiles [116, 130, 218].

KNN has also been investigated for infectious disease detection, including COVID-19, tuberculosis, and malaria [146, 190, 195]. During the COVID-19 pandemic, KNN was applied to symptom-based classification, CT scan analysis, and risk assessment. For tuberculosis detection, KNN has been used with chest X-ray images and sputum-related data, while malaria diagnosis has involved classification of infected and non-infected samples from microscopic blood smear images [116, 190]. These applications demonstrate the flexibility of KNN across different healthcare datasets and classification problems.

Compared with DL algorithms, KNN can be applied to relatively small datasets without requiring extensive model training or large computational resources. This characteristic makes KNN relevant to healthcare applications where data availability may be limited because of privacy concerns, data collection difficulties, or other practical constraints. KNN has therefore also been investigated for automated mental health screening and classification [65].

However, conventional KNN relies on a fixed distance metric and is sensitive to the local distribution of data. These limitations can affect its classification performance when datasets contain heterogeneous features or complex relationships. In mental health assessment, psychometric instruments provide quantitative measurements of psychological characteristics, but the resulting data may exhibit variability, nonlinear relationships, and differences among individuals. Consequently, the effectiveness of a fixed-distance KNN model may be limited when applied to such datasets.

Several modifications of KNN have been proposed to address the limitations of the conventional approach. These include Ensemble KNN, Weight-adjusted KNN, and Fuzzy KNN, which attempt to improve classification by incorporating multiple classifiers, feature or instance weighting, or fuzzy membership mechanisms [54, 92, 112]. Although these approaches address some limitations of conventional KNN, challenges related to generalization, sensitivity to data distribution, model complexity, and computational requirements remain.

Overall, the existing literature demonstrates the applicability of KNN in healthcare and its potential for classification using relatively small datasets. However, comparatively limited research has investigated adaptive distance mechanisms for KNN-based depression and anxiety prediction using standardized psychological

assessment instruments. In particular, there is a need for approaches that can dynamically adapt the distance measure according to the local characteristics of the data rather than relying solely on a fixed distance metric. This gap motivates the investigation of an ASKNN approach for depression and anxiety prediction presented in Chapter 6.

SVM SVM is a kind of machine learning method that uses a hypothesis space composed of the high-dimensional feature space of a linear function. These systems are trained using a learning approach based on optimisation theory that considers learning bias. The idea behind SVMs is to find the best hyperplane for classification. [102].

DT In classification tasks, where the objective is to predict categorical labels or class memberships, DTs are frequently utilized. The input space is divided into regions by the DT algorithm, and each region is assigned a unique class label. Particularly helpful are jobs like sentiment analysis, spam identification, and medical diagnosis. With hierarchical, sequential structures that recursively partition the data, DT provides a means to describe the rules that underlie the data [168]. The main objectives of DTCs are: 1) to accurately classify the training sample as much as feasible; 2) to classify unknown data with the highest potential accuracy, which generalizes beyond the training sample; 3) to be simple to update as additional training samples are available; 4) and have as simple a structure as possible [200].

Classification and Regression Trees (CART) It is used for tasks involving both classification (categorizing data) and regression (forecasting numerical values). To produce predictions, data is recursively divided into subsets according to the most important characteristics [240].

C5.0 C5.0 is a sophisticated DT method that improves on the ID3 algorithm in several ways and is frequently used for classification jobs. It can handle categorical and continuous data and chooses attributes based on information gained [180].

CTREE The DT algorithm CTREE was created specifically for categorical data. It can be used in cases where attribute values are not continuous since it builds DTs using

statistical methods appropriate for discontinuous, non-numeric variables [95].

DT Ensemble Learning Within ensemble approaches like RF and Gradient Boosting, DTs are frequently utilized as the basic learners. In this situation, several DTs are joined to build more robust and precise models. RF is an illustration of a set of DTs used for both regression and classification. Bagging and boosting are the two basic categories into which the ensemble approaches can be divided.

Bagging The process of bagging entails producing several predictor iterations and aggregating them to produce an aggregated predictor. The aggregate determines the average of each number when generating a numerical prediction. On the other hand, when predicting a category, it determines the most common choice by a plurality vote. By generating bootstrap replicas of the training set and employing them as fresh training sets, several iterations are produced. Tests conducted on actual and simulated data sets have demonstrated that bagging may lead to substantial enhancements in accuracy. This enhancement has been noted in other situations, such as subset selection in linear regression and the use of regression trees and categorization. One crucial component is the prediction approach's instability. If the act of perturbing the learning set leads to substantial alterations in the generated predictor, then the technique of bagging can enhance the accuracy [31].

RF The subject of categorisation has made extensive use of this ensemble learning approach. Several DTs are built within an RF model; each DT serves as a classifier and generates an output. By reducing variance and increasing bias, the generic bagging method used by the RF classifier improves model performance. The category with the most votes is chosen as the final output by the RF algorithm, which aggregates the results of all classifications [32].

Bagged CART method The bagged CART approach applies bootstrap aggregation (bagging) on DTs (usually CART) in an ensemble setting. It creates a large number of unique DTs, each trained on a different collection of data, and then aggregates their predictions. This technique reduces variation and enhances the model's accuracy and stability by averaging out individual tree biases [31].

Boosting A set of DTs is iteratively constructed using an ensemble learning method known as "boosting with DTs." Every tree in the succession is trained to correct the mistakes of its predecessors, with a focus on misclassified cases. Until the required level of accuracy is reached or a predefined number of trees is reached, this iterative process is continued. Combining the individual predictions of each tree in the sequence results in the final forecast. Boosting is a widely applicable technique that enhances the accuracy of any learning system.

Gradient Boosting Gradient Boosting is a boosting ensemble technique for generating DTs one after the other. Every tree corrects the errors of the one before it, focusing on the incorrectly classified cases [169].

Extreme Gradient Boosting (XGBoost) XGBoost can be described as a scalable ML system for tree boosting. Through the use of significantly fewer resources than current methods, XGBoost scales beyond billions of examples [44].

Adaptive boosting (AdaBoost) One of the ML algorithms is called AdaBoost. To create a more effective learning algorithm, it may be combined with different learning algorithms. It builds a learning method with stronger classifiers by combining it with weak classifiers [74].

LightGBM Another gradient-boosting framework with an eye toward efficiency and the capacity to handle huge datasets is LightGBM [111].

CatBoost A gradient boosting method created specifically to complement categorical features [6].

Logistic Regression (LR) A type of regression known as logistic regression uses a logistic function to forecast the likelihood that an event will occur. Similar to other types of regression analysis, LR uses a set of predictor variables that can either be numeric or categorical in nature [228].

2.8.1.4 Supervised Learning Regressors

In order to predict numerical results based on labelled training data, supervised learning regressors are frequently employed to model the relationship between input features and continuous target variables.

Linear Regression To put it simply, linear regression is the process of creating a statistical model that, given one or more independent variables, features, predicts one dependent variable, the target. A linear connection is used to make a forecast about the target variable, symbolized with the graphical representation of a straight line. Find the line that minimizes the discrepancy between the anticipated values and the actual data points as an ideal fit [157].

Lasso Regression A feature selection and prediction method called Lasso regression places restrictions on the parameters that lead the variable reduction coefficients to approach zero. Finding the subset of characteristics that reduces the prediction error for a certain response variable is the aim of lasso regression [19, 239].

Neural Network Regression An ML technique called neural network regression uses artificial neural networks, which are made up of many interconnected neurons or units, to carry out regression tasks. It is particularly suited for applications like predicting stock prices, home prices, or any other problem where the objective is to estimate a real-valued output since it learns from data patterns to forecast continuous numeric values [223].

Elastic Net Strongly linked predictors likely to be in or out of the model collectively, which is encouraged by the elastic net's grouping effect. When there are many more predictors (p) than observations (n), the elastic net is especially helpful [273].

Support Vector Regression (SVR) SVR is characterized by kernel utilization, sparse solutions, and VC control of the margin and support vector count. SVR is a valuable technique for accurately predicting functions with real-valued outputs. SVR utilizes a symmetrical loss function that imposes equal penalties on both overestimation

and underestimation errors. This characteristic makes SVR a supervised learning technique [13].

Decision Tree Regression (DT Regression) DT can be utilized for regression tasks even though they are typically connected with classification. DT predicts continuous numerical values rather than discrete class labels in regression. For applications like estimating housing values based on characteristics like size, location, and number of bedrooms, DT regression is useful [142].

Ridge Regression It modifies the traditional linear regression cost function by adding a penalty term that discourages the model from assigning excessively large coefficients to predictor variables. Reducing the impacts of multicollinearity and avoiding overfitting makes the model more resilient and broadly applicable [160].

2.8.1.5 Unsupervised Learning

Without using labeled results, an algorithm using unsupervised learning examines and explores data to discover hidden patterns, structures, or groups. To find important information in datasets, techniques like clustering and dimensionality reduction are used.

K-Means Clustering The approach, which iteratively updates the centroids to reduce within-cluster variance, assigns each data point to the cluster with the nearest centroid. For tasks like customer segmentation, image reduction, and data preprocessing, K-Means is frequently utilized [90].

K-Means++ Clustering K-Means++ applies a probabilistic approach to select an initialization for centroids much more likely to yield a good final clustering solution compared to random initialization. With this method, the sensitivity of an initial placement of centroids in traditional K-Means has been reduced [161].

Agglomerative Hierarchical Clustering This is an agglomeration method in which a point is taken as a separate cluster first and then merges progressively with other points based on similarity. The hierarchical form generated using this bottom-up approach,

commonly visualized as a dendrogram, displays how groups are nested into each other. Agglomerative clustering allows you to select the number of clusters at varying levels of detail and is useful for discovering hierarchical associations in data [30].

Gaussian Mixture Model (GMM) The GMM method of clustering presumes that the data points actually belong to a composite of several Gaussian distributions. Every such Gaussian has some mean and covariance matrix, and their combination would represent the distribution of the data points. By this GMM clustering, a single data point may belong to multiple clusters, because it is possible for a data point to be probabilistically distributed amongst several clusters. It is somewhat flexible and can easily identify clusters of all sizes and orientations, and it works better if the data is produced from underlying Gaussian components [193].

Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) BIRCH is an excellent clustering algorithm, especially designed for massive datasets. The data is organised into a hierarchical tree structure, with each node in the tree denoting a cluster. To save memory usage, data summaries are created within each node and also reduce computational complexity. Such a method is indispensable while dealing with high-dimensional data or while applying clustering over streaming data, where data is continuously arising [249].

Spectral Clustering The eigenvalues and eigenvectors of a similarity matrix, which are frequently obtained from pairwise similarities between data points, are used in spectral clustering. Contrary to conventional techniques, spectral clustering is good for locating clusters that may not have a straightforward geometric shape and can capture complicated, non-linear correlations between data points. When working with data where conventional distance-based measures might not be effective, it is especially helpful [250].

Table 2.9: Year-wise previous studies that used ML-based techniques for mental health detection

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
1	[143]	SVM, SVR, RF, LogReg, RR	The best models are SVR and RR (e.g., R2 for SVR and RR: 0.56, 95% CI: 0.45, 0.64 and 0.45, 0.63, respectively) and SVM and LogReg (AUPRC: 0.76, 95% CI: 0.69, 0.81)	1. The study makes use of a longitudinal dataset. 2. The study looked at several machine learning algorithms and emphasised the significance of model interpretability and feature selection.	1. ML models lack generalizability due to demographic focus. 2. Future studies should include diverse demographics.
2	[48]	DT, RF, and XGBoost	XGBoost anxiety accuracy = 86%, depression accuracy = 84%, insomnia accuracy = 81%	1. Shows ML models' value in predicting mental health for stakeholders.	1. Only tree-based ML models used

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
3	[99]	Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), KNN, RF, DT, and Light Gradient Boosting Machine (LightGBM)	DT, GB, XGB accuracy for depression and anxiety=100%	1. Most algorithms achieved over 99% accuracy, demonstrating effectiveness in detecting anxiety and depression.	1. Results are specific to Saudi Arabian healthcare workers. 2. More validation with diverse datasets and real-world testing is needed.
4	[24]	DT, NB, RF, SVM	RF achieved maximum accuracy = 78.9%	1. Potential for accurate identification of students at high risk of serious mental collapse.	1. Small dataset

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
5	[162]	LR, DT, RF, KNN, AdaBoost, XGBoost, GB, K-Means clustering, K-Means++ clustering, Partition around medoids	LR, DT, RF, CNN, AdaBoost, XGBoost, and GB achieved accuracies of 91.414%, 85.353%, 91.414%, 89.898%, 88.383%, 93.434%, and 93.939%, respectively.	1. The findings of the study will assist companies in identifying workplace issues that affect mental health and in putting into practice effective interventions for employee well-being.	1. Small dataset
6	[128]	DT, RF, NB, SVM, and kNN algorithms	SVM depression accuracy = 94.9%, stress accuracy = 97.4%, and anxiety accuracy = 97.8%.	1. Rajyoga meditation has been demonstrated to improve pandemic resilience.	NA

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
7	[55]	RF and LASSO	RMSE = 0.300 for all datasets	1. Clear nonlinear relationships between predictors and outcomes were found through ML.	1. Unexpected social identity interactions were not uncovered. 2. Non-linear correlations in outcomes crucial for comprehending mental health issues.
8	[22]	LASSO, RF, XGBoost, and SVM models	NA	1. Predicted COVID-19 concern in older U.S. adults. 2. Used sociodemographic data and signs of depression as predictors.	1. Small sample size affected analyses and self-reported data validity
9	[129]	RF	RF accuracy = 86%	1. Effectively identifies individual predictors and interactions that are statistically significant.	1. Small sample size. 2. Missing responses. 3. Pandemic effects on risk variables were not discussed.

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
10	[140]	ANN model	ANN accuracy = 90.0%.	1. Interventions in mental health are essential for enhancing well-being. 2. In clinical practice, a potential screening device.	1. Self-report scales may introduce biases. 2. Homogeneous demographics limit generalizability.
11	[38]	SVR and MLR models	MSEs for the SVR performances were 2.04 and 2.29 for the training and test data, respectively. For the MLR, MSEs were 1.97 and 2.39 for the training and test data, respectively.	1. Data-driven approach explores physical and mental aspects. 2. Shows U-shaped trajectories for managing expectations and support.	1. Analysis is preliminary with limited datasets. 2. Cross-sectional vs. longitudinal designs restrict depth. 3. Individual-level variable effects are unclear. 4. UK-centric focus limits generalizability.

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
12	[171]	Svm, knn	SVM accuracy = 80%, whilst the kNN accuracy = 83%.	1. Virtual Cognitive behavioral therapy provides ways to deal with various behavioral disorders.	1. Limited feature selection may affect how complete the model is. 2. Small sample sizes have an impact on the generalizability and overall data representation.
13	[245]	ML models for classification (LDA and SVM) and regression (linear regression and SVR)	A sensitivity of 79% and a specificity of 88% are used to predict suicidal thoughts. SVM and SVR yielded the most optimal outcomes.	1. Clinical and sociodemographic data can predict OCD symptom changes. 2. ML aids in identifying high-risk patients for optimal care.	1. Because of the limited sample size, phone-based examinations, and lack of information on drugs and co-occurring illnesses, care must be taken when extrapolating results.

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
14	[185]	Pattern regression analyses, Linear epsilon-insensitive support vector machine (ε-SVM) regression	PTSD Model: Twofold CV: R=0.35, r2=0.12, NMSE = 0.96, P=0.001. Fivefold CV: r=0.34, r2=0.12, NMSE = 0.90, P=0.001. Depressive Symptoms Model: Two-fold CV: r=0.36, r2=0.13, NMSE = 0.90, P=0.001. Fivefold CV: r=0.38, r2=0.15, NMSE = 0.86, P=0.001.	1. The evaluation of traumatic incidents, especially connected to the COVID-19 pandemic.	1. The sample may not represent all Brazilian healthcare professionals, and confounders like work experience and job roles could affect results. 2. Future validation should use an independent sample.
15	[188]	SVM, KNN, NB, RF, NN, DT, and GB	For depression, the Gradient Boosting accuracy was 83.3% and the RF accuracy was 83.2%; for anxiety, it was 82.9% and 81.3%.	1. To facilitate focused preventive interventions, ML can assist in screening high-risk pregnant/postpartum women for anxiety and depression.	1. Small dataset, incomplete assessments, inconsistent model performance across demographics, and ethical issues.

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
16	[144]	Penalized L, SVM, NN, RF, and GB.	Average AUROC: 0.595 (95% CI: 0.35-0.808), Accuracy: 0.567 (95% CI: 0.39-0.74), MAE: 0.602 (95% CI: 0.412-0.826). Temporal Features: MAE: 0.596 (95% CI: 0.391-0.826), AUROC: 0.581 (95% CI: 0.343-0.790).	1. This study's goal is to use passively acquired data to create a burnout screening tool.	1. The results may not apply to other regions or physician cohorts because the study was carried out in a single university medical center with a limited number of trainee physicians. 2. Response bias
17	[26]	FP-growth algorithm, DT	DT with an anxiety model and DT with a depression model both achieved an accuracy of 68.72% and 68.68%, respectively	1. The association principles of frequent pattern mining provide relevant patterns.	1. The DT findings were not significantly impacted by feature selection via pattern mining.

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
18	[131]	CART, C5.0, RF, CTREE, GBM, Treebag	CART mean accuracy = 0.77	1. Self-report bias is lessened by ML models using simple demographic inputs.	1. Unbalanced sample in terms of demographics from COVID-19 and gender. 2. It was impossible to use many assessment techniques.
19	[64]	lasso and elastic net algorithms, SVM, random forest, extreme gradient boosted regression	Elastic net accuracy = 64.93%	1. Bivariate analyses linked COVID-19 anxiety to perceived mortality threat	1. Early pandemic data from China limit generalizability.
20	[93]	SVR and Gradient Boosting Regression (GBR) and	RMSE of up to 16.6%	1. The literature on COVID-19 benefits from a multimethod approach using scales, analytics, and ML.	1. Low retention and small sample sizes reduce study power. 2. Future validation with larger samples is necessary.

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
21	[159]	elastic net penalized logistic regression	elastic net AUC = 77%	1. Understanding mechanisms and risk factors is necessary for recovery and tailored treatment.	1. Small dataset
22	[127]	DT, SVM, kNN, and LR	NB accuracy = 92.15%	1. The study recognizes various causes of psychological trauma from the pandemic and lockdowns.	1. There was only one week of the study's longitudinal follow-up. 2. Negative impacts may not represent the whole range of psychological experiences.
23	[5]	linear regression model	NA	1. Text messaging aids in mental health strategy dissemination.	1. Attrition bias possible from incomplete final evaluations.

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
24	[212]	RF, DT, SVM, LR, NB, kNN, Perceptron, and GBM models	DT acc = 67.28%, LR acc = 67.47%	1. ML assists in making personalized medical decisions and stress risk identification. 2. Model appropriate for experts in public health and psychiatry. 3. Enables the identification of people who are susceptible to mental disease.	1. Small dataset

S. NO.	Previous Work	Models used	Results	Advantages	Challenges
25	[252]	XGBoost model	XGBoost model accuracy = 80%.	1. The application of the XGBoost model in institutional and pandemic contexts is discussed.	1. Small sample size limits representation. 2. No data on participants' family COVID-19 exposure. 3. Family exposure could independently affect anxiety levels.

Apart from classifier selection, there are other methods that some researchers used such as feature selection and feature engineering methods in order to determine informative features and lower the dimensions of mental health datasets. Methods such as Random Forest based feature importance, RFE, ReliefF, MRM, SelectKBest, SHAP, Markov Blanket, and correlation methods have been used in conjunction with different ML classifiers.

For example, Tuan et al. [244] applied RFE and ET with KNN, MLP, and SVM and reported an accuracy greater than 98.4% using SVM. RezaPour et al. [194] investigated multiple ML classifiers with RFE and reported 96.32% accuracy using RF with SMOTE. Nayan et al. [170] used ReliefF with LR, RF, SVM, LDA, KNN, and NB and reported accuracies of 89.76% for depression and 93.09% for anxiety using RF. Zulfiker et al. [274] combined mRMR, SelectKBest, and SMOTE with different ML models and reported promising classification performance, demonstrating the potential benefit of combining feature selection with data-balancing strategies

However, the reviewed studies use different feature-selection techniques, datasets, feature sets, and validation strategies, making it difficult to determine whether the reported performance improvements are attributable to the feature-selection method, the classifier, or the characteristics of the dataset. In addition, several studies rely on cross-sectional data, small samples, online surveys, or geographically restricted populations [61, 121, 238]. Feature importance and explainability approaches such as RF-based feature ranking and SHAP have also been used to identify influential predictors and interpret model decisions [58, 63, 176]. However, the use of different feature representations and evaluation procedures remains a challenge for meaningful comparison across studies.

Therefore, further investigation is required to understand the contribution of feature selection and feature weighting methods to reliable depression and anxiety detection. Table 2.10 summarises previous research that used ML models and feature selection methods to identify mental health issues.

Table 2.10: Year-wise previous studies that used ML-based mental health detection along with feature selection

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
1	[47]	DT, RF, and XGBoost	MLR and XGBoost models were fed feature selection techniques based on quantum annealing (QA)	XGBoost anxiety accuracy=86% depression accuracy= 84%, insomnia accuracy= 81%	1. Investigated depression variables post-COVID-19. 2. Introduced a quantum annealing approach for feature selection.	1. Need for dataset analysis from other countries to generalize findings.
2	[267]	AdaBoost Classifier, Ridge Classifier, CatBoost Classifier, Light Gradient Boosting Machine, and standard ML classifiers	RFECV algorithm	Adaboost depression accuracy= 0.7917	1. Incorporated sociodemographic, psychological, and epidemic-related factors into the model.	1. For cross-site validation, a large, diverse sample is crucial but hard to achieve globally.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
3	[63]	XGBoost, AdaBoost, NB, KNN, SVM, RF, and LR.	Random forest feature importance ranking	AdaBoost AUC = 72.96%, NB AUC = 76.37%, and RF AUC = 78.27%	1. The MLP-based model exhibited the highest AUC and satisfactory accuracy.	1. Limited generalizability and potential biases from sample demographics and convenience.
4	[244]	KNN, MLP, and SVM	RFE and ET	SVM accuracy ¿ 0.984	1. Integrating indicators aims to improve understanding and tailor mental health therapies.	1. Focus on recovered COVID-19 patients limits generalizability
5	[150]	DT, kNN, and NB	Feature importance applied	KNN has an accuracy of 92.32%	1. The methodology included both negative and positive questions, along with weekly records,	1. Decision Tree offers better latency but less accuracy than KNN; comparing algorithms is complex and resource-intensive.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
6	[149]	XGBoost, AdaBoost, NB, KNN, SVM, RF, and LR.	RFE	SVM accuracy=79% for suicide	1. Initial effort to predict suicidal tendencies in Bangladeshi university students using ML.	1. Certain crucial attributes may have been omitted from the Adaboost forecasting model.
7	[121]	J48 DT, SVM	Random forest feature importance ranking	SVM accuracy =100%	1. Enhancing student mental well-being through targeted policies and interventions.	1. COVID-19 restrictions limited data collection to online methods.
8	[46]	quantum annealing (QA)-based optimization, MLR, and XGBoost	quantum annealing (QA)-based optimization, MLR and XGBoost algorithms	The top 10 features were identified by QA with an accuracy of .750 MLR and XGBoost algorithms	1. Analyzed longitudinal data from 16 South Korean areas. 2. Considered geographical factors in analysis.	1. Key characteristics, like specific diseases, were missing. 2. Generalizability is limited to South Korea.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
9	[238]	SVM, RF, LR, and XGBoost	XGBoost model	XGBoost accuracy for depression and anxiety = 0.89	1. The study used 19 features for efficient and high-performing clinical screening.	1. The cross-sectional design limits causation conclusions. 2. Downsampling reduced accuracy.
10	[194]	DT, Multinomial LR, NB, RF, Gradient Tree Boosting, XGBoost, kNN, SVM, NN, and CatBoost	RFE	SMOTE RF = 96.32	1. Aimed to help healthcare facilities improve staff mental health during COVID-19.	1. For complicated and accurate models, more diverse data from more places is required.
11	[58]	Light Gradient Boosting Machine, GradientBoost, and RF	SHAP	NA	1. Novel use of ML and visualizations to find trends. 2. Covered healthcare workers' roles during the initial pandemic.	1. Limited understanding of long-term outcomes due to no follow-up.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
12	[137]	EN, SVR, GBRT, ensemble classifier	EN	An ensemble classifier achieves up to 85.2% accuracy.	1. Identifies risks a week in advance.	1. Small dataset
13	[61]	RF, SVM	Feature importance	Depression RF acc = 68%, Anxiety RF acc = 75%, Obsessive and Compulsive Symptoms SVM acc = 71%, Belief Symptoms SVM accuracy = 92%.	1. Could help health authorities with public health initiatives for disadvantaged groups.	1. Different baseline data approaches made comparisons difficult. 2. Small sample size and lack of baseline data limit representativeness.
14	[43]	LR with LASSO	Feature ranking	Resilience AUC = 0.83 and chronicity AUC = 0.85	1. Longitudinal data capture heterogeneity and suggest adding post-disaster worry.	1. Small sample size 2. LGMM may misidentify group numbers due to skew and kurtosis.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
15	[219]	gradient tree boosting, RF, and elastic net	The SMILE scale was examined as an independent lifestyle	Spanish sample Elastic Net accuracy = 0.79, Brazilian sample accuracy = 0.76	1. Implications for targeted strategies to reduce impact.	1. Cross-sectional design limits causal inference.
16	[174]	RF, MLP, XGBoost, LR, SVM, KNN, DT	ReliefF	XGBoost model accuracy = 69.47% with 40 features.	1. Highlighted the beneficial effects of family relationships and coping mechanisms.	1. Broad diagnostic categories limited understanding of specific disorders.
17	[170]	LR, RF, SVM, LDA, kNN, and NB	ReliefF	depression RF acc = 89.76%, and anxiety RF acc = 93.09%	1. Highlighted the beneficial effects of family relationships and coping mechanisms.	1. Cross-sectional design prevents causal link identification. The lockdown led to challenging data collection and the use of online surveys.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
18	[176]	multivariable LR model	SHAP	Positive factors: meaning in life, spouse, trust in neighbors, and female gender. Negative factors: poor health, income anxiety, psychiatric diseases, and isolation.	1. Japan's first COVID-19 happiness study with 25,000+ participants.	1. Exploratory design limits causal link establishment. Results may not generalize to other nations or non-pandemic contexts.
19	[8]	SVM and J48 DT	CorrelationAttributeEval	J48 obtained 93.50% accuracy for three classes and 95.66% accuracy for two classes.	1. Researchers advise Saudi Arabian decision-makers to use this model when preparing their mental health care services.	NA

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
20	[14]	GBM, RF, DT, SVM, NN, RG, LS and EN	RF	RF ACC = 0.71	1. Demographic, socioeconomic, behavioral, and epidemiological variables can be used in ML models to quickly screen for and diagnose mental health problems.	1. Only a depression diagnosis was used in the study.
21	[37]	Gradient-boosted decision trees and the SHAP	MRMR	First Wave sensitivity = 76.5% and a Second Wave specificity = 77.8%	1. Longitudinal approach revealed PMDD onset factors during pandemics.	1. Small sample size; self-report evaluation; and data only up to 2020.
22	[199]	NB, LR, RF, and SVM	Markov blanket	RF ACC in April 2020 = .89	1. Interpretable ML revealed complex correlations between predictors of anxiety and depression.	1. The cross-sectional design limited tracking of long-term mental health changes and link direction.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
23	[187]	RF, regression trees	random forest, regression trees	NA	1. The study discovered risk variables for psychological distress by examining how they interact.	1. The cross-sectional design lacked causal clarity, and selection bias, and identifying distress triggers was a challenge.
24	[268]	DT, LR with LASSO, RF, and gradient-boosting tree	LR with LASSO	Gradient-boosting tree model accuracy = 0.829	1. Multifold cross-validation with 2,574 samples and 1,187 events prevents overfitting, while predictive performance is assessed with a separate test dataset.	1. Large sample sizes and diverse ethnicities are crucial for cross-site model validation, but achieving global diversity is challenging.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
25	[255]	GCBA, IGCBA	Global Chaos Bat Back Propagation Neural Network	IGCBA-BPNN-4 accuracy = 92.55%.	1. According to experimental findings, IGCBA-BPNN outperforms current approaches.	1. IGCBA reduces input dimensions but doesn't optimize BPNN parameters.
26	[97]	XGBoost	SHAP, PCA	XGBoost model r2=0.39	1. The study used a representative Canadian sample and revealed complex anxiety-depression correlations with ML.	1. The cross-sectional design limited long-term mental health tracking
27	[105]	RF, NB, and LR	Markov blanket	RF and SVM accuracy = 80%, while NB and LR accuracy = 77%	1. The study enhances AI comprehensibility for population-level forecasting.	1. Results are specific to the USA. 2. The dataset is insufficient for deep analysis.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
28	[173]	ANFIS	Linear Correlation	ANFIS stress accuracy = 80%	1. Priority is given to prediction optimization and feature selection.	1. A detailed discussion on the predictor explaining ability is not given.
29	[192]	LR	Correlation analysis	LR depression accuracy = 73.5% and anxiety accuracy = 81.42%	1. ML techniques like SMOTE, AIC, multivariate LR, and ROC curves boost scientific rigor and reliability.	1. The small sample size limited the investigation of school variation; the Chi-square test showed no urban-rural differences.
30	[235]	Ensemble Binary Classifier, SVM, DT, ANN, KNN	Feature extraction	Ensemble classifier accuracy for PHQ-9 = 95.4%	1. Predictive performance is improved by combining the classification approach with binary ground truth.	1. The reliability and sensitivity of the ensemble system must be tested using additional datasets.

Continued on next page

S. NO.	Reference	Models used	Feature Engineering	Results (Best Model Results)	Advantages	Challenges
31	[274]	KNN, AdaBoost, GB, XGBoost, Bagging, GB, Bagging, and AdaBoost	mRMR, SelectKBest, SMOTE	AdaBoost with SelectKBest accuracy = 92.56%	1. This study looked for the most frequent contributing variables to depression.	1. Using BDC as the gold standard for depression is a research flaw. 2. No biological markers were included.
32	[71]	LR, SVM, NB, RF	CFS	roc area for Logistic=0.782, SVM=.697, Naïve Bayes=.776, RF=.776	1. Pandemic stress levels were higher than in non-emergency situations.	1. Online hiring may miss disadvantaged communities. 2. The sample was mostly educated, young individuals.

Fig. 2.10 shows the most used models in the studies using a pie chart. The following graph illustrates the variety of approaches used in the field by representing several ML models. Along with less well-known models like LASSO, NB, and XGBoost, it includes well-known models like RF, SVM, DT, kNN, LR, and Gradient Boosting. There are also more specialized models, including AdaBoost, CatBoost, and CART, to name a few. This collection of models demonstrates the wide range of methodologies present in the ML field.

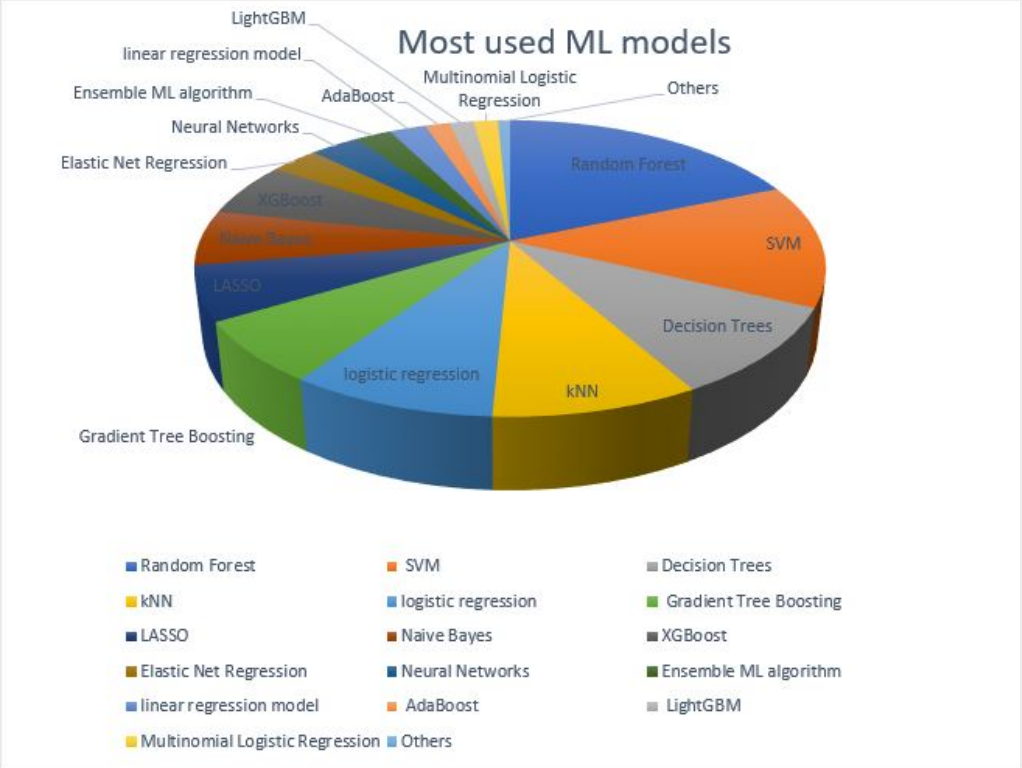


Figure 2.10: Depiction of the most used ML models.

2.8.2 DL Models Used for Detecting Depression and Anxiety

Due to their ability to automatically extract complex patterns from data, DL techniques have drawn a lot of attention in recent years. As a result, they are highly successful at diagnosing anxiety and depression.

Previous studies have explored DL and neural network-based approaches for automated mental health detection using different populations and data sources. For example, [25] compared LSTM with SVM, ADA Boost, RF, K-NN, LR, and MLP and

reported 100% accuracy using LSTM; however, such a high accuracy raises concerns regarding possible overfitting and requires validation using diverse and independent datasets. [73] used MLP and reported an average MSE of 0.5300, although the use of data collected primarily during the COVID-19 pandemic may limit the generalizability of the findings. Similarly, [86] compared ANN with SVM, DT, and RF and reported 98% accuracy using SVM, while the study noted comparatively lower performance of ANN due to variation within the dataset. [109] investigated LR, KNN, DT, RF, Bagging, and DNN, with LR, DT, and Bagging achieving 100% accuracy; however, the study highlighted concerns related to individual variability, demographic and cultural differences, and data privacy and bias. These studies demonstrate the potential of neural network and DL-based approaches for mental health detection, while also indicating the need for sufficiently representative datasets, robust validation, and consideration of generalizability. A summary of DL models used for detecting mental health issues is presented in Table 2.11. Due to their ability to automatically learn complex patterns and representations from data, DL techniques have attracted considerable attention in the field of automated mental health assessment. Previous studies have explored neural network and DL-based approaches for detecting depression and anxiety using diverse populations, data sources, and modelling strategies.

For example, [25] compared LSTM with SVM, AdaBoost, RF, K-NN, LR, and MLP and reported 100% accuracy using LSTM. Although this result demonstrates the potential of LSTM for mental health prediction, such exceptionally high performance may indicate overfitting and therefore requires validation using diverse and independent datasets. Fong et al. [73] used an MLP-based approach and reported an average MSE of 0.5300 for predicting anxiety and depression. However, the dataset was collected primarily during the COVID-19 pandemic, which may limit the generalizability of the findings to other periods and populations. Similarly, Gupta et al. [86] compared ANN with SVM, DT, and RF and reported 98% accuracy using SVM. Their findings showed comparatively lower performance of ANN, indicating that DL performance may be sensitive to dataset variability and characteristics. Kaman and Vajpai [109] investigated LR, KNN, DT, RF, Bagging, and DNN for depression prediction among Indian IT workers, with LR, DT, and Bagging achieving 100% accuracy. However, the authors highlighted challenges related to individual variability, demographic and

cultural differences, data privacy, and potential bias.

Overall, these studies demonstrate the potential of DL and neural network-based approaches for automated depression and anxiety detection. However, the reported results vary considerably across datasets, populations, data sources, and evaluation settings. In particular, some studies report exceptionally high classification performance, while others demonstrate that dataset variability can adversely affect DL models. Furthermore, existing studies have predominantly focused on specific populations or data sources, and comparatively limited work has systematically compared multiple DL architectures using consistent datasets and standardized psychological assessment instruments. These limitations highlight the need for a controlled comparative evaluation of DL models for depression and anxiety assessment, which is addressed in Chapter 4. The major DL techniques relevant to this research are described below.

CNN CNNs are composed of neurons that, like traditional artificial neural networks, undergo optimization through learning. Every individual neuron will persistently receive input and perform a computational task, similar to a scalar [177].

ANN Modelled after biological neural networks, ANNs excel in finding patterns and meaning in complicated data. For issues that are too complicated to be resolved using conventional methods, they are frequently employed in a variety of scientific domains. Neurons from three different types of layers—input, hidden, and output layers—make up ANNs. Because ANNs work well with nonlinear problems, they have been effectively used to identify anxiety and sadness [247].

Deep Neural Network (DNN) In comparison, deeper neural networks have learned high-level characteristics with more complexity and abstraction [231]. Deep Neural Networks have shown outstanding levels of predicting accuracy that may be comparable to that produced by a human's ability.

Deep Convolutional Neural Network (DCNN) Deep architectures often outperform shallow structures when taking on challenging learning issues. The layer-wise stacking of numerous linear and non-linear processing units enables the learning of complicated

representations at various levels of abstraction. Deep CNNs may perform at a level comparable to humans by learning the invariant representations if given enough training data [114].

RNN An RNN is any network whose neurons engage in bidirectional communication. Sequential and time-series data may be used to extract features and long-term dependencies using RNNs. At least one link between every non-linear unit in the stack of non-linear units of the RNNs forms a directed cycle. Any dynamical system may be modelled by an RNN with the right training; nevertheless, long-term dependency learning is a major problem for RNN training [114].

LSTM A particular kind of RNN, which is called LSTMs, was created to deal with the vanishing gradient issue that arises when creating models with long-range dependencies. They preserve memory cells that have the capacity to hold data for extended periods of time. As a result, they are able to record the long-term interdependencies of sequential data, which enhances the accuracy of estimates [10].

Tabnet With the use of attentive mechanisms and gradient-based learning, TabNet is a DL network for tabular data that can collect a variety of high-dimensional data and identify which attributes are crucial for producing predictions that are both accurate and understandable [10].

Table 2.11: Comparative analysis of prior DL-based studies for depression and anxiety detection

S. No.	Reference	Models Used	Best Model / Result	Task Type	Advantages	Challenges
1	[25]	LSTM, SVM, AdaBoost, RF, K-NN, LR, MLP	LSTM accuracy = 100%	Classification	LSTM demonstrated strong potential for automated mental health detection and may support timely identification of mental health conditions.	The reported 100% accuracy raises concerns regarding possible overfitting and requires validation using diverse and independent datasets.
2	[73]	MLP	Average MSE = 0.5300	Regression	Robust validation and exploratory data analysis enhance the reliability of the findings. Quantitative attributes provide useful information for analysing mental health during crises.	The dataset was collected primarily during the COVID-19 pandemic, which may limit generalizability to other periods and populations.

Continued on next page

S. No.	Reference	Models Used	Best Model / Result	Task Type	Advantages	Challenges
3	[86]	ANN, SVM, DT, RF	SVM accuracy = 98%	Binary Classification	HuDA_COVID highlights the potential of computational approaches for analysing mental health during the COVID-19 pandemic and addressing psychological needs.	High variability within the dataset resulted in comparatively lower performance of the DL model, particularly ANN, than traditional ML approaches.
4	[109]	LR, KNN, DT, RF, Bagging, DNN	LR, DT, Bagging = 100%	Binary Classification	Automated depression prediction can support early identification, reduce assessment time, and assist mental health intervention.	Automated models may not capture complex individual cases. Classifier performance may vary across demographic and cultural contexts. Data privacy and potential bias require careful consideration.

Leveraging a chord diagram given in Fig. 2.11, it visualizes the intricate connections between authors and the specific ML and DL techniques they've employed in their research, offering a comprehensive view of their contributions in the realm of this research. Fig. 2.12 shows how ML and DL can be used to detect anxiety and depression.

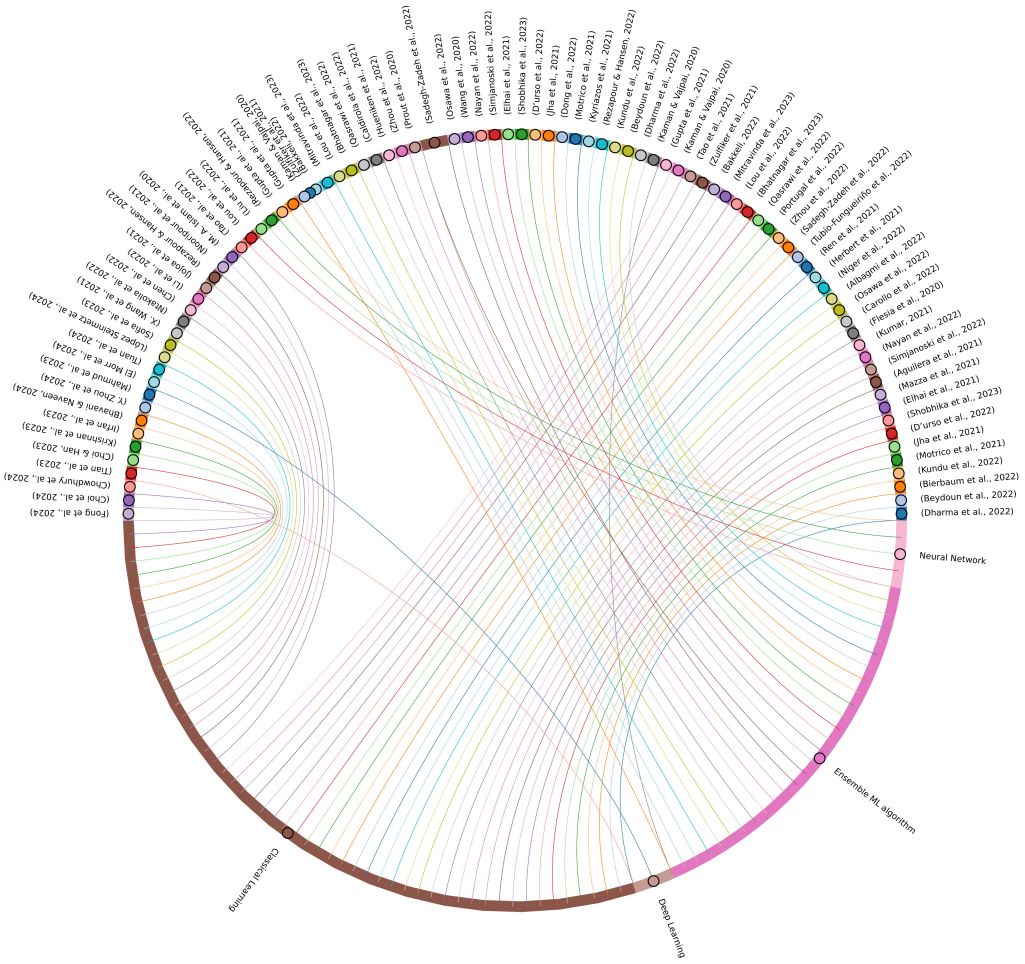


Figure 2.11: A chord diagram demonstrating the author's relationship to DL and ML approaches.

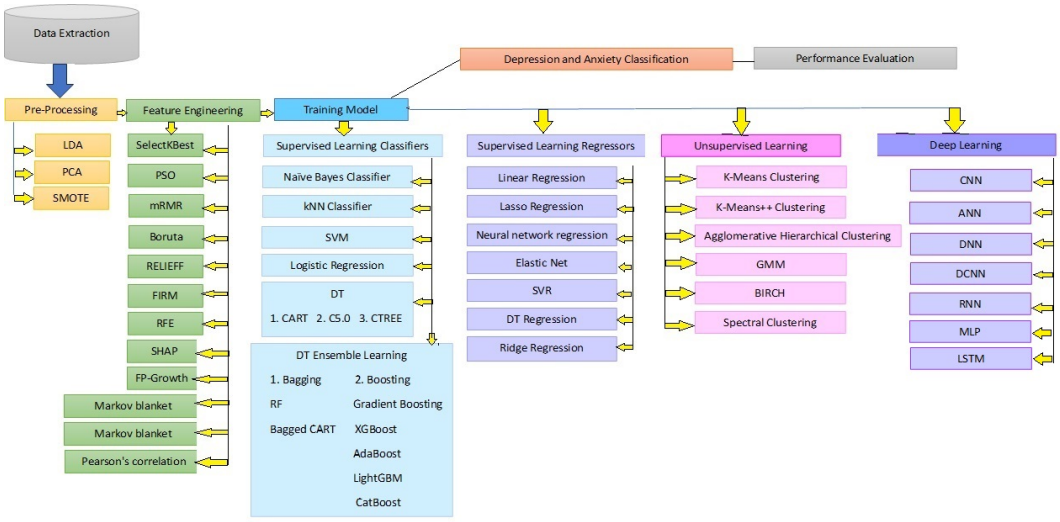


Figure 2.12: The application of ML and DL for depression and anxiety detection.

2.8.3 Hybrid ML and DL Models for Depression and Anxiety Detection

Depression and anxiety detection have been widely investigated using various approaches. Social media, text, and audio data have been used for detecting mental health problems. Another approach is the use of standardized psychological assessment instruments for the detection of depression and anxiety. Hybrid models combine two or more techniques to exploit their complementary strengths and improve predictive performance. Table 2.12 summarizes existing studies that have employed hybrid models for mental health detection.

Table 2.12: Overview of existing approaches in mental health assessment using hybrid models

S. No.	Reference	Models Used	Method	Results
1	[248]	CNN and LSTM models in a hybrid model	Social media	Accuracy = 92%
2	[152]	CNN and LSTM models in a hybrid model	Text + audio	Accuracy = 80%
3	[115]	Hybrid model using TF-IDF feature extraction and Logistic Regression as the classification algorithm	Twitter tweets	Accuracy = 99.4%
4	[272]	Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN)	Social media	Accuracy = 89.5%
5	[202]	Hybrid CNN-SVM	Social media	Accuracy = 68%
6	[138]	Hybrid CNN and Long Short-Term Memory (LSTM) model	Text + audio	Accuracy = 96.125%
7	[7]	Hybrid Random Forest (RF) and Artificial Neural Network (ANN) model	Psychological instrument	Accuracy = 74%
8	[262]	DT + KNN and RF + NN hybrid models	Psychological instrument	DT + KNN Accuracy = 86.69%; RF + NN Accuracy = 93.54%

2.9 Impact of Dataset Size on Machine Learning Model Performance

This section examines pertinent research on the connection between dataset size and ML model performance. It highlights key studies, their contributions, and the gaps addressed by the present research.

Sordo and Zeng [222] examined the relationship between classification accuracy and sample size. Their study used NB, SVM, and DT on a dataset of 8,500 text quotes from Brigham and Women's Hospital, focusing on patients' smoking status. Their results, which demonstrated a favourable correlation between training set size and classification accuracy, were consistent with earlier studies. As the dataset increased, both SVM and DT demonstrated notable improvements in performance, indicating more stable and reliable learning.

The work done by Imlawi and Alsharo [98] looked at how estimation methods affected

sample distribution variance, concluding that resampling methods exhibited lower variance when approximating classification accuracy. Further research has also suggested that larger datasets may help reduce the influence of assessment strategy choices on variance.

Similarly, the study by Dyar and Ytsma [60] supports the notion that increases in dataset size minimize the effect of differences in evaluation strategies. Their work demonstrated that prediction accuracy on unprocessed data consistently improved with larger training sets, indicating increased model robustness.

Shahinfar et al. [214] used an image dataset to thoroughly examine the performance of DL models with gradually increasing per-class (species) sample numbers. Their study provided valuable insights into how incremental growth in class-level data affects training stability and overall accuracy.

Similarly, Stockwell and Peterson [227] investigated how sample size affects species distribution models' accuracy. Their work highlighted how insufficient samples can significantly limit predictive performance, particularly in ecological modeling scenarios. Table 2.13 compares related work and their limitations.

Table 2.13: Comparative analysis of related work and their limitations

Authors	**Dataset**	**Contributions**	**Limitations**
[222]	Text Quotes	Increasing sample size leads to improved performance by overcoming local fluctuations and stabilizing accuracy.	Classifier performance can be negatively affected by limited training cases, especially in complex classification tasks.
[98]	Census Income Dataset	Resampling techniques effectively reduce variance in performance estimation, especially with smaller datasets (100–500 samples).	Resampling may be computationally expensive and not practical in time-sensitive or resource-limited environments.
[60]	Chemical Prediction Data (Rock Samples)	Highlights the importance of specifying training and testing sample sizes to accurately assess model performance and prevent overfitting.	Small training sets may lead to overfitting and poor generalization.
[214]	Image Dataset	Identifies a logarithmic relationship between model performance and the number of training images.	Challenges include incorrect labels, imbalanced samples, image quality variations, and camera trap inconsistencies, affecting robustness.
[227]	Atlas of the Distribution of Mexican Birds	Emphasizes selecting suitable modeling methods based on sample size and environmental variables for improved prediction accuracy.	Logistic regression models show unpredictable accuracy depending on environmental variable choice.

2.10 Evaluation Metrics

A range of performance measures is used to evaluate how well models predict anxiety and depression. ML and DL models employ evaluation metrics such as F1-score, AUC-ROC, Precision, Recall, Sensitivity, Accuracy, and Specificity. Nonetheless, statistical study requires metrics like Mean, Standard Deviation, Mean Square Error, and Root Mean Square Error. This section gives answers to research question 9.

Mean (μ)

The mean value is obtained by dividing the total of all the values in a data collection by the total number of items [141]. It is given by the following equation:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Where:

- The overall number of data points is N.
- Each distinct data point is represented by x_i.

Standard Deviation (σ)

The standard deviation is a statistical tool used to quantify the variance or dispersion in a set of data points. It offers a means of comprehending how dispersed or grouped the data values are around the mean of the data collection [141]. It is given by the equation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Where:

- The overall number of data points is N.
- Each distinct data point is represented by x_i.
- μ is the data set's mean (average).

Mean Square Error (MSE)

The expression can be defined as the total number of data points divided by the sum of the squared discrepancies between the predicted and actual target variables [94]. It is given by the equation:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Where:

- MSE stands for Mean Squared Error.
- n is the total number of data points or observations.
- Y_i reflects the i-th data point's actual (observed) value.
- $\hat{Y}_i$ shows the expected result for the i-th data point.

Root Mean Square Error (RMSE)

The RMSE is equal to the square root of the MSE. The MSE is assessed in the square of the target variable, whereas the RMSE is quantified in the same units as the target variable. Because of the way it is formulated, MSE effectively penalizes greater errors more harshly, much like the squared loss function it comes from [85, 94, 208]. It is given by:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}$$

Where:

- RMSE is the Root Mean Square Error.
- n is the total number of data points or observations.
- Y_i represents the actual (observed) value for the i-th data point.
- $\hat{Y}_i$ represents the predicted value for the i-th data point.

Accuracy

The percentage of successful hits about the total number of classifications is known as accuracy [18, 196]. The accuracy, defined as the ratio of correct forecasts to the total number of predictions, is determined by considering the number of false-positive (FP), false-negative (FN), true-positive (TP), and true-negative (TN) predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP}$$

Precision

Positive predictive value (PPV), which is another term for precision. It is the estimation of the precise proportion of correctly identified events among the favorably classified instances [82].

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall

Recall quantifies the model's capacity to identify positive samples [82, 241]. The recall is calculated using the number of FP, FN, and TP:

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-score

F1 score measures the accuracy of a model. It is the mean of the accuracy and recall rating of a model. Accuracy- it is a measure that calculates how often the model actually makes good predictions over the whole dataset [186]. The F1-score is another useful measure to be used in the evaluation of how the models will perform over datasets with imbalanced class distributions. It is calculated by the ratio of recall and precision:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Specificity

The percentage of genuine negatives that the test accurately detects is known as specificity [9].

$$\text{Specificity} = \frac{TN}{TN+FP}$$

Matthew's Correlation Coefficient (MCC)

The mean corrected correlation (MCC) is one metric used to evaluate the performance of binary classification models. This metric accounts for true positives, true negatives, false positives, and false negatives. MCC yields one number with +1 indicating that the model would make totally perfect predictions, 0 indicating that the model would make completely random predictions, and -1 showing an utter discrepancy between the predictions and the actual results.Thus, MCC makes it possible to assess a model's performance, particularly in datasets that are unbalanced and have different class sizes [156].

$$\text{MCC} = \frac{(TN \times TP) - (FN \times FP)}{(FP+TP)(FN+TP)(TN+FN)}$$

AUC-ROC

The average sensitivity of a test across all likely specificity values, or vice versa, is called a ROC curve. The overall diagnostic efficacy of a test can be evaluated using the AUC measure.

The range of accepted values is from 0 to 1, with 0 indicating a test that is entirely erroneous and 1 indicating an entirely accurate test [151].

2.11 Post-Pandemic Mental Health Crisis

Pandemics have historically had profound and long-lasting effects on society. Past global health crises, such as the H1N1 Swine Flu, HIV/AIDS, SARS, Ebola, and Zika virus outbreaks, have not only strained healthcare systems but also reshaped perceptions of contagious diseases and their broader societal consequences.

Human Immunodeficiency Virus (HIV) attacks CD4 cells, which are crucial for the immune system. This significantly impairs the body's ability to combat opportunistic infections, such as tuberculosis, fungal infections, bacterial infections, and various cancers [258]. Individuals who are HIV-positive or at risk often experience heightened mental health issues [191]. Thus, HIV impacts mental health across diverse populations, extending its effects beyond the immune system.

Severe Acute Respiratory Syndrome (SARS) severely affected healthcare workers as the virus rapidly spread in hospital settings, which was first identified in November 2002 in Guangdong Province of China. Research during the SARS outbreak highlighted significant psychological distress among healthcare professionals due to their frontline roles [158].

The global spread of Influenza H1N1 (Swine Flu) in 2009 posed significant challenges to healthcare systems. Reactions to pandemics often include stockpiling supplies and altering travel plans. Observations showed varying responses, with Malaysians displaying higher anxiety and Europeans underestimating H1N1 mortality, indicating a need for better communication about the flu vaccine [80].

The first Ebola epidemic happened in remote rainforest villages in Central Africa. However, the West African Ebola outbreak from 2014 to 2016 was unprecedented, covering Guinea, Sierra Leone, and Liberia, and had the highest case and mortality counts [257]. The West African Ebola outbreak from 2014 to 2016, the most severe since the virus was first discovered, had profound effects on mental health. Research in Sierra Leone revealed significant depression, anxiety, and PTSD among survivors, medical professionals, and affected families [103].

The Zika virus, which can cause congenital Zika syndrome, including microcephaly, is predominantly transmitted by Aedes mosquitoes, which are active during the day [259]. Pregnant women are at risk from the Zika virus, which is spread by Aedes mosquitoes and can result in congenital Zika syndrome, which can include microcephaly. The widespread

fear and panic among pregnant women globally highlight the virus's extensive psychological impact [69].

These historical pandemics underscore the complex challenges infectious diseases present and emphasize the importance of preparedness and support networks.

Long-term mental health issues have replaced short-term psychological anguish as a result of the COVID-19 pandemic's substantial and ongoing rise in mental health disorders worldwide. Emerging evidence indicates a rise in conditions such as depression, anxiety, PTSD, and sleep disturbances across diverse populations, including individuals who were not directly infected [11].

Post-pandemic mental health outcomes have been particularly severe among vulnerable groups such as healthcare workers, students, elderly individuals, and those experiencing economic or social instability. Studies suggest that psychological symptoms often persist long after recovery, indicating that mental health impacts are not merely immediate reactions but part of a prolonged post-COVID syndrome [11].

A key observation across the literature is the widespread use of standardized psychometric tools such as PHQ-9, GAD-7, and DASS-21 for assessing mental health conditions, as shown in Fig. 2.13. Although these instruments offer trustworthy clinical benchmarks, they mostly depend on self-reported data and frequently overlook more comprehensive contextual elements including social isolation, environmental stressors, and digital behaviour patterns. This restricts their capacity to offer a thorough comprehension of the dynamics of mental health in the post-pandemic age.

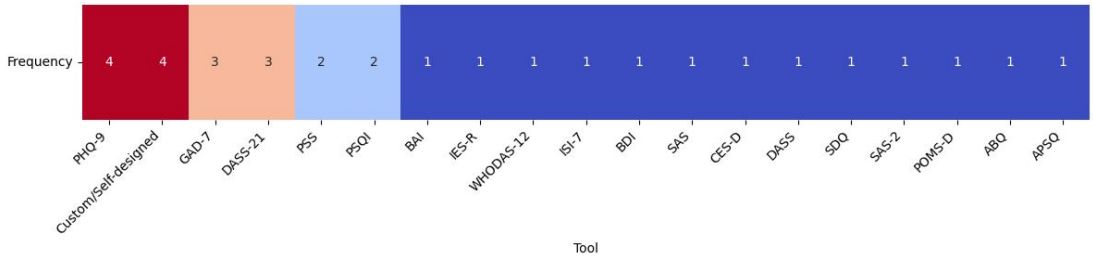


Figure 2.13: Frequency of different mental health assessment tools utilized in chosen studies for the assessment of post-COVID-19 psychological symptoms.

Furthermore, existing studies predominantly employ traditional statistical methods, including regression analysis and hypothesis testing, which are effective for identifying associations but lack the capability to model complex, high-dimensional data. As a result, there is limited focus on predictive and real-time assessment of mental health conditions.

Importantly, there is a notable underutilization of ML and DL techniques in post-pandemic

mental health research [11]. Given their ability to process large-scale, multi-modal data and enable early detection and personalized interventions, these approaches present a significant opportunity for advancing mental health assessment.

Therefore, the post-pandemic context highlights the need for data-driven, scalable, and real-time mental health monitoring systems. The employment of ML-based techniques in this study, which aims to improve the precision and efficacy of mental health prediction models, is directly motivated by this gap.

Summary

Two of the most common mental health conditions, depression and anxiety, were thoroughly reviewed in this chapter, along with the computational methods used for their automated diagnosis. It began by outlining the fundamental concepts, including the types, risk factors, and clinical manifestations of these disorders, highlighting the influence of genetic, neurochemical, traumatic, and psychosocial factors on their development.

A detailed overview of traditional screening and diagnostic techniques was provided, covering psychological instruments, neuroimaging methods such as MRI, EEG, HRV, facial expression analysis, biological markers, and social media-based sentiment analysis. Among these, self-report psychological instruments such as PHQ, BDI, GAD, and DASS were identified as widely used due to their practicality and validated effectiveness across diverse populations.

The chapter further presented a structured and systematic review of ML and DL approaches for mental health assessment, with particular emphasis on studies conducted during and after the COVID-19 pandemic. A wide range of models, including traditional classifiers, ensemble techniques, and deep learning architectures, were critically examined in terms of their methodologies, data sources, feature engineering strategies, and performance evaluation metrics.

This chapter highlighted a number of important obstacles despite notable progress, such as problems with data quality, class imbalance, model interpretability, overfitting, and generalisability across datasets. These drawbacks show that more reliable, scalable, and effective methods for predicting mental health are required.

Overall, this chapter establishes a strong theoretical and methodological foundation for the present research by integrating domain knowledge with existing computational techniques. The insights derived from this chapter help identify critical research gaps and directly motivate the development of the proposed models and experimental framework presented in the subsequent

chapter.

Chapter 3

Unveiling the Impact of Dataset Size on Machine Learning Models for Anxiety and Depression Prediction amid the COVID-19 Pandemic: Determining Optimal Data Collection Thresholds

3.1 Introduction

This chapter examines how dataset size affects ML models' ability to identify anxiety and depression during the COVID-19 epidemic. While previous chapters establish the relevance of ML in mental health assessment, this chapter specifically examines how variations in data availability affect model accuracy, robustness, and generalizability.

Using multiple mental health survey datasets, this chapter evaluates the performance of DT, RF, SVM, and KNN classifiers across progressively increasing dataset sizes. Data slicing and augmentation techniques are employed to simulate realistic data collection scenarios.

This chapter's main goals are to find practical criteria for accurate mental health prediction and to spot patterns in model performance as the dataset size grows. Fig 3.1a provides a visual depiction of the suggested study project. The critical review presented in Section 2.9 identified a gap in systematically examining the influence of dataset size on ML-based anxiety and depression prediction. Accordingly, this chapter investigates the effect of varying dataset

sizes on the performance of multiple ML models using consistent experimental settings. The insights derived from this analysis provide guidance for effective data collection strategies and inform the development of more advanced models presented in subsequent chapters.

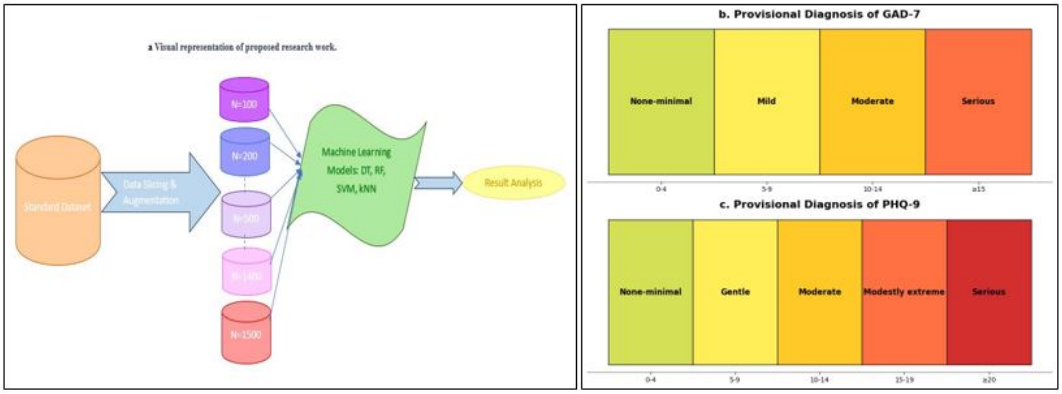


Figure 3.1: a Visual representation of proposed research work, b Provisional diagnosis and severity of GAD-7 & Figure 1c Provisional diagnosis and severity of PHQ-9

3.2 Proposed Methodology

To ensure generalizability and scalability, four ML models were applied across three different datasets as explained in Table 3.1. In addition to standardised psychological tests like the PHQ-9 and GAD-7 as explained in detail in section 2.5, which gauge the degree of anxiety and sadness associated with the COVID-19 epidemic, all datasets included demographic data.

3.2.1 Datasets Used

We used three datasets in this work, collected by Islam et al. [100], Qasrawi et al. [188], and Carollo et al. [38]. The GAD-7 and PHQ-9 tests were utilised in every study to gauge participants' levels of anxiety and depression. Higher levels of depression and anxiety were reported across participants. Table 3.1 describes the characteristics of the datasets in detail. Tables 3.2, 3.3 and 3.4 present anonymized illustrative snapshots of the datasets used in this study, showing representative feature variables, questionnaire responses, derived scores, and corresponding class labels for supervised learning.

Table 3.1: Description of datasets used

S. No.	Reference	Size (N)	Dataset Link	Description	Psychological Instruments
1	[100]	476	https://doi.org/10.7910/DVN/N5BUJR	Clinical survey data of students	GAD-7 and PHQ-9 questionnaire
2	[188]	3569	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FCDGEB	Clinical survey data of pregnant and postpartum women	GAD-7 and PHQ-9 questionnaire
3	[38]	435	https://rdr.ucl.ac.uk/articles/dataset/Self-Perceived_Loneliness_and_Depression_During_the_COVID-19_Pandemic_a_Two-Wave_Replication_Study/20183858	Respondents from the UK and Greece during the first lockdown	GAD-7, PHQ-9, and self-perceived loneliness scores

Table 3.2: Anonymized Illustrative Snapshot of the Open-Source Dataset D1

Age	Sex	GAD1	GAD2	GAD3	GAD Score	GAD Class	PHQ1	PHQ2	PHQ3	PHQ Score	PHQ Class
25	1	1	1	1	7	2	1	1	1	5	2
21	1	0	0	0	1	1	0	0	0	1	1
22	2	3	3	3	19	4	3	1	3	25	5
20	2	1	1	0	6	2	3	1	0	9	2
19	1	1	0	0	5	2	1	0	1	9	2
21	2	0	0	0	0	1	2	0	0	2	1
23	1	0	0	0	3	1	3	0	1	5	2

Table 3.3: Anonymized Illustrative Snapshot of the Open-Source Dataset D2

f_id	Preg.	Anx.1	Anx.2	Anx.3	Anx.4	Anx.5	Anx.6	Anx.7	Anx. Score	Anx. Class	Dep.1	Dep.2	Dep.3	Dep.4	Dep.5	Dep.6	Dep.7	Dep.8	Dep.9	Dep. Score	Dep. Class
2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	0	0	3	2	3	3	1	12	1	2	1	1	0	0	0	2	0	0	6	1
4	2	0	0	0	0	1	0	0	1	0	1	0	1	1	0	0	1	0	1	5	1
5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3.4: Anonymized Illustrative Snapshot of the Open-Source Dataset D3

Country	LoneTot	CASEStot	PHQtot	PHQ Class	GADtot	GAD Class	SleepTot	BeliefTot	ModActDiff	CHAOStot	spq22	RPQtot	MildActDiff	MildActTimeDiff
United Kingdom of Great Britain and Northern Ireland	65	18	15	2	9	1	11	39	-3	20	16	14	-7	-180
United Kingdom of Great Britain and Northern Ireland	39	47	6	1	18	2	10	37	3	12	2	6	2	0
United Kingdom of Great Britain and Northern Ireland	42	53	8	1	6	1	9	33	-4	13	7	14	-4	-30
United Kingdom of Great Britain and Northern Ireland	39	44	1	0	2	0	13	39	-1	7	4	6	-5	-90
United Kingdom of Great Britain and Northern Ireland	62	54	25	2	17	2	8	36	-6	14	7	4	2	60
United Kingdom of Great Britain and Northern Ireland	45	49	9	1	4	0	11	34	-2	13	4	5	-3	-180
United Kingdom of Great Britain and Northern Ireland	46	38	13	1	18	2	10	39	0	19	9	7	2	90
United Kingdom of Great Britain and Northern Ireland	47	47	4	0	1	0	11	40	6	13	12	4	-1	10

3.2.2 Mental Health Assessment

The GAD-7 scale was used to measure anxiety, and the PHQ-9 scale was used to detect depression as explained in detail in section 2.5. The GAD-7 is a widely used self-report instrument for assessing the severity of generalized anxiety disorder symptoms [226]. It is a simple and practical tool that helps mental health professionals and primary care physicians diagnose and monitor anxiety symptoms.

The GAD-7 can be used for initial screening, determining the severity of anxiety symptoms, monitoring treatment progress, and tracking symptom changes over time.

The seven items on the GAD-7 correspond to common symptoms of anxiety disorders. Each question asks about the frequency of specific anxiety-related symptoms experienced during the previous two weeks. Respondents rate each item on a scale from 0 to 3, where 0 indicates "not at all" and 3 indicates "nearly every day." Similar to the PHQ-9, each item contributes a score between 0 and 3, yielding a total score ranging from 0 to 21.

For this work, anxiety levels were categorized as follows:

- None–minimal: < 5
- Mild: 5–9
- Moderate: 10–14
- Severe: ≥ 15

Figure 3.1b summarizes the provisional diagnosis associated with the GAD-7 scale and the corresponding "Anxiety Class" used as the output variable for the ML models.

A common self-report tool for determining a person's degree of depression symptoms is the PHQ-9 [123]. The nine items on it ask how frequently a certain set of depression symptoms occurred over the previous two weeks; each item is graded from 0 (not at all) to 3 (almost every day). The overall score is between 0 and 27. Depression severity levels were categorised as mild (5–9), moderate (10–14), moderately severe (15–19), and severe (≥ 20) for the purposes of this work. Figure 3.1c summarizes the provisional diagnosis of the PHQ-9 scale and the corresponding "Depression Class" used as the output variable for the ML models.

3.2.3 Data Slicing and Data Augmentation

In the present work, datasets of different sizes, ranging from N=100 to N=1500, were constructed using data slicing and augmentation techniques. Data slicing was conducted through random

sampling, wherein, for each target sample size, A subset of the data was chosen at random. In the cases where the target sample size was smaller than the available data, straightforward random sampling was done. On the other hand, data augmentation was applied when the target sample size was greater than the data that was available. To get the required sample size for data augmentation, the original dataset's random samples were duplicated. This involved the repetition of the whole dataset several times until it reached close to the target size. After which, any deficit in reaching the exact sample size was completed by random sampling (See Algorithm 1).

```
 Algorithm 1: Data Slicing and Augmentation
Input: Original dataset D, Target sample sizes S
Output: Augmented datasets D1, D2, ..., Dk
for Si in S:
if Si <= n:
Di <- random_sample(D, Si)
else:
num_repeats <- Si // n
remainder <- Si % n
Di <- concatenate(repeat(D, num_repeats), random_sample(D, remainder))
return D1, D2, ..., Dk
```

For the data-slicing experiments, datasets of sizes ($N = 100, 200, \ldots, 1500$) were generated by randomly sampling observations from each original dataset. A fixed random seed (`random_state=42`) was used to ensure reproducibility. Each sample size was generated independently and was not intentionally constructed as a nested subset. When the requested sample size exceeded the number of available observations, resampling was used to obtain the required dataset size. Five-fold cross-validation was used across various subgroups to reduce the possibility of bias. This makes it easier to evaluate how augmentation affects model performance in a reliable way.

3.2.4 Data Preprocessing and Predictive Modelling

The datasets used in this work typically have a large number of missing, inconsistent, and superfluous variables. Therefore, in order for ML models to handle the data, data cleansing is required. Missing qualities for PHQ-9 and GAD-7 surveys were supplanted with 0, accepting

the absence of the side effect. The training data set has chosen the goal variables known as the anxiety class and the depression class in order to forecast mental health problems. In this undertaking's underlying stage, an underlying trial is conducted, utilizing an $80-20$ split. Here, 80% of the information fills in as the preparation dataset, while the leftover 20% fills in as the testing dataset. The acquired experimental outcomes are the foundation for performance evaluation. In this work, precision, recall, accuracy, and the F1 score are computed using the classification report. By methodically examining the optimal hyperparameters that would enhance the models' performance, we improved the methods used to construct the models. We have done a grid search for each model, using cross-validation techniques to consider a range of various hyperparameters; these have included the maximum depth in DT, the number of neighbors in KNN, and even the number of estimators in RF. From the Grid Search CV, we were able to pick the final set of parameters that gave the best trade-off between accuracy and overfitting. Hyperparameters selected for each classifier are summarized in Table 3.5.

Table 3.5: Hyperparameters for each classifier

Model	**Hyperparameters Tuning**
DT	max_depth = 5
	min_samples_leaf = 75
SVM	C = 0.1
	kernel = 'rbf'
	gamma = 'scale'
	random_state = 10
CNN	n_neighbors = 2
RF	n_estimators = 100
	min_samples_leaf = 50

3.2.4.1 Machine Learning Algorithms Used

Four supervised machine learning algorithms—DT, SVM, KNN, and RF—were used in this study to predict anxiety and depression. These models were thoroughly described in Section 2.8.1.3. The DT algorithm is a non-parametric technique that creates a hierarchical tree structure of nodes and leaves by recursively dividing the dataset according to criteria that maximise information gain or minimise variance [168]. SVM is very useful in high-dimensional spaces; in this work, an RBF kernel with a regularisation value $C = 0.1$ and gamma = 'scale' was employed. SVM looks for the best hyperplane that divides data points of various classes [102]. Euclidean distance is frequently employed to gauge similarity, and in this case, $k = 2$ was

used. The KNN method classifies a new instance based on the majority class of its nearest neighbours [253]. In order to increase generalisation and decrease overfitting, RF, an ensemble learning technique, combines several DTs. In this work, 100 trees with at least 50 samples per leaf were used [32].

3.2.5 Evaluation Metrics

To assess the predictive performance of the ML models in this work, we employed standard evaluation metrics commonly used in classification tasks. These metrics were defined in detail in Section 2.10. The metrics used include accuracy, precision, recall, and F1-score.

Together, these metrics offer a thorough understanding of how well the models identify anxiety and depression, and they serve as the foundation for evaluating the effectiveness of the four machine learning models used in this chapter.

3.3 Statistical Analysis of Model Performance

The statsmodels package (version 0.14.1) and the Python programming language were used to perform statistical analysis. The survey responses were analyzed and visualized using scripts written in Python. The ANOVA test is done, which calculates how dataset size influences the performance of different ML models while predicting anxiety and depression. To ascertain whether the mean value for the performance metrics—accuracy, precision, recall, and F1-score—across various dataset sizes differs significantly, the ANOVA test was used.

Hypotheses Formulation:

- **Null Hypothesis (H0):** "There is no significant relationship between the size of the dataset and the performance of ML models in terms of accuracy, precision, recall, and F1-score."
- **Research Hypothesis (H1):** "As the size of the dataset increases, the performance of ML models improves in terms of accuracy, precision, recall, and F1-score."

Then, for every model, the respective p-values were calculated along with F-statistics to judge the significance of observed differences in performance metrics. An F-statistic is a measure of the ratio of variance explained by the model to unexplained variance. When the p-value is smaller than the predetermined significance level $\alpha = 0.05$ and the F-statistic is high, it indicates that performance improves dramatically with increasing dataset size.

3.4 Results

Different ML models were used to assess the psychological well-being appraisals across dataset sizes from 100 to 1500. Four ML models are analysed to determine their predictive skills, accuracy, precision, recall, and F1-score. This demonstrates the models' ability to accurately classify anxiety and depression classes, identify genuine positives, avoid false negatives, and optimise recall and precision. Utilizing these markers offers a customized evaluation of how well a model performs based on the objectives and important aspects of the application.

3.4.1 Performance of Dataset 1 for GAD-7 and PHQ-9

Using GAD-7 and PHQ-9 data, the study evaluated how well ML classifiers predicted anxiety and depression levels, as shown in Table 3.6. Initially, the classifier's performance was not very good, with accuracies varying from 0.15 to 0.6. Precision, recall, and F1 scores showed that it was difficult to predict anxiety levels using the given data. As the number of samples grew, the classifiers got better, with SVM and RF consistently performing better than DT and kNN. When N was 200, the results improved, and the accuracy varied from 0.38 to 0.73. All classifiers achieved high accuracies, often exceeding 0.8, when N was 400 or more. On the other hand, especially with smaller sample sizes, kNN showed greater recall but poorer accuracy. This indicates that anxiety levels based on GAD-7 scores may be accurately predicted by ML algorithms, enabling early identification and management. DT fared worse than other classifiers, particularly in terms of accuracy and recall. This demonstrates that ML algorithms may be able to predict anxiety levels based on GAD-7 scores with accuracy, enabling early detection and intervention. A comparison of performance metrics GAD-7 and PHQ-9 for dataset 1 is depicted in Fig. 3.2.

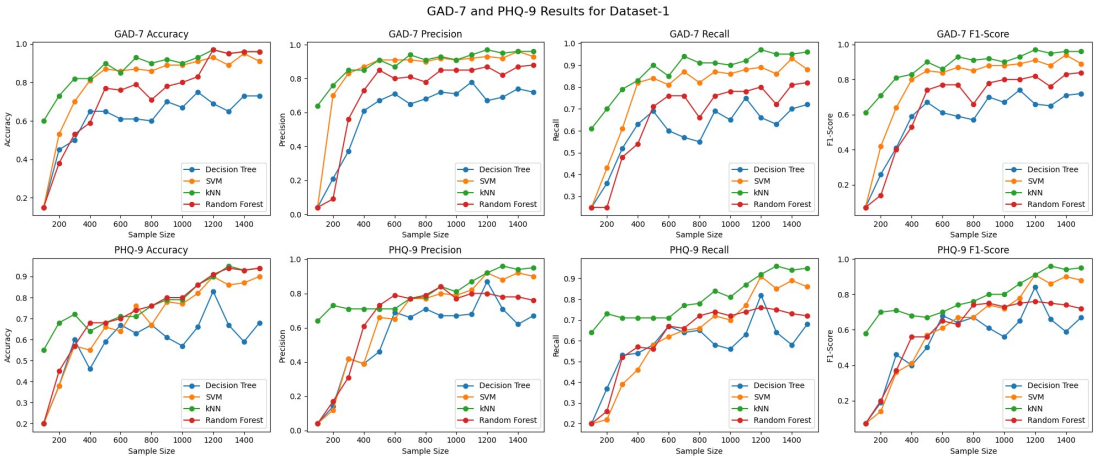


Figure 3.2: Comparison between the performance metrics of GAD-7 and PHQ-9 on varied dataset sizes for Dataset-1

The findings of the PHQ-9 questionnaire show that different ML classifiers perform differently in terms of predicting depression severity. All classifiers performed poorly with accuracies averaging 0.2 at a sample size of 100. The misclassification rates of ML models across different dataset sizes were calculated Table 3.7 for dataset 1. It has been observed that at a smaller sample size, the model lacks the power to learn meaningful patterns from the little data available. So, the misclassification rate was higher at a lower sample size. However, a decrease in the same has been observed as the sample size increases. Performance improved progressively as the sample size grew, with accuracy spikes noted at certain sample size increments. When compared to DT and kNN, SVM and RF consistently showed greater accuracies and other performance indicators at N=300. Although its precision was lower, kNN demonstrated competitive recall performance, indicating a propensity for more false positive predictions. Lower performance indicators for DT suggest that it is not as good at capturing intricate linkages in the PHQ-9 data.

It seems that many classifiers have distinct performance measures, emphasising how important it is to select models that are appropriate for the specific characteristics of the data. High predicted accuracy and the ability to capture non-linear correlations are the strengths of SVM and RF, although kNN could be more appropriate in situations where high recall is more important than precision. Because DT has intrinsic limits when processing non-linear patterns, it might not be as appropriate for complicated datasets like PHQ-9.

According to the plateau effect, additional increases in sample size may not significantly enhance predictive accuracy at a certain point. This emphasizes how important it is to optimize sample size, taking into account real-world factors like available resources and the desired

degree of forecast accuracy.

Table 3.6: Performance of Dataset–1 for GAD-7 and PHQ-9 using different machine learning models.

Sample Size (N)	DT				SVM				kNN				RF			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
GAD-7 Results for Dataset-1																
N=100	0.15	0.04	0.25	0.07	0.15	0.04	0.25	0.07	0.60	0.64	0.61	0.61	0.15	0.04	0.25	0.07
N=200	0.45	0.21	0.36	0.26	0.53	0.70	0.43	0.42	0.73	0.76	0.70	0.71	0.38	0.09	0.25	0.14
N=300	0.50	0.37	0.52	0.41	0.70	0.83	0.61	0.64	0.82	0.85	0.79	0.81	0.53	0.56	0.48	0.40
N=400	0.65	0.61	0.63	0.59	0.81	0.87	0.82	0.80	0.82	0.85	0.83	0.83	0.59	0.73	0.54	0.53
N=500	0.65	0.67	0.69	0.67	0.87	0.91	0.84	0.85	0.90	0.91	0.90	0.90	0.77	0.85	0.71	0.74
N=600	0.61	0.71	0.60	0.61	0.86	0.91	0.81	0.84	0.85	0.87	0.85	0.86	0.76	0.80	0.76	0.77
N=700	0.61	0.65	0.57	0.59	0.87	0.91	0.87	0.87	0.93	0.94	0.94	0.93	0.79	0.81	0.76	0.77
N=800	0.60	0.68	0.55	0.57	0.86	0.90	0.82	0.85	0.90	0.91	0.91	0.91	0.67	0.78	0.66	0.66
N=900	0.70	0.72	0.69	0.70	0.89	0.92	0.87	0.88	0.92	0.93	0.91	0.92	0.78	0.85	0.76	0.78
N=1000	0.67	0.71	0.65	0.67	0.89	0.91	0.86	0.88	0.90	0.91	0.90	0.90	0.80	0.85	0.78	0.80
N=1100	0.75	0.78	0.75	0.74	0.89	0.92	0.88	0.89	0.93	0.94	0.92	0.93	0.80	0.85	0.78	0.80

Sample Size (N)	DT				SVM				kNN				RF			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
N=1200	0.69	0.67	0.66	0.66	0.91	0.93	0.89	0.91	0.97	0.97	0.97	0.97	0.82	0.87	0.80	0.82
N=1300	0.65	0.69	0.63	0.65	0.89	0.92	0.86	0.88	0.95	0.95	0.95	0.95	0.76	0.82	0.72	0.76
N=1400	0.73	0.74	0.70	0.71	0.95	0.96	0.93	0.94	0.96	0.96	0.95	0.96	0.84	0.87	0.81	0.83
N=1500	0.73	0.72	0.72	0.72	0.91	0.93	0.88	0.89	0.96	0.96	0.96	0.96	0.84	0.88	0.82	0.84
PHQ-9 Results for Dataset-1																
N=100	0.20	0.04	0.20	0.07	0.20	0.04	0.20	0.07	0.55	0.64	0.63	0.58	0.20	0.04	0.20	0.07
N=200	0.38	0.14	0.37	0.19	0.38	0.12	0.22	0.14	0.68	0.73	0.75	0.70	0.45	0.17	0.26	0.20
N=300	0.60	0.42	0.53	0.46	0.57	0.42	0.39	0.36	0.72	0.71	0.75	0.71	0.55	0.31	0.52	0.37
N=400	0.46	0.39	0.54	0.40	0.55	0.39	0.46	0.41	0.64	0.71	0.71	0.68	0.57	0.61	0.57	0.56
N=500	0.59	0.46	0.58	0.50	0.70	0.66	0.58	0.57	0.68	0.71	0.69	0.67	0.58	0.73	0.56	0.56
N=600	0.67	0.69	0.67	0.68	0.68	0.65	0.62	0.61	0.68	0.71	0.71	0.70	0.67	0.79	0.67	0.65
N=700	0.63	0.66	0.64	0.64	0.76	0.77	0.65	0.67	0.71	0.77	0.73	0.74	0.69	0.77	0.66	0.63
N=800	0.67	0.71	0.65	0.67	0.74	0.77	0.66	0.67	0.76	0.78	0.77	0.76	0.71	0.79	0.72	0.74

Sample Size (N)	DT				SVM				kNN				RF			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
N=900	0.61	0.67	0.58	0.61	0.78	0.80	0.72	0.74	0.79	0.84	0.80	0.80	0.74	0.84	0.74	0.75
N=1000	0.57	0.67	0.56	0.56	0.78	0.79	0.70	0.72	0.79	0.81	0.80	0.80	0.73	0.77	0.72	0.73
N=1100	0.66	0.68	0.63	0.65	0.83	0.82	0.77	0.78	0.86	0.87	0.85	0.86	0.79	0.80	0.74	0.75
N=1200	0.83	0.87	0.82	0.84	0.90	0.92	0.91	0.91	0.90	0.92	0.91	0.91	0.76	0.80	0.76	0.76
N=1300	0.67	0.71	0.64	0.66	0.86	0.88	0.85	0.86	0.95	0.96	0.96	0.96	0.76	0.78	0.75	0.75
N=1400	0.59	0.62	0.58	0.59	0.90	0.92	0.89	0.90	0.93	0.94	0.94	0.94	0.73	0.78	0.73	0.74
N=1500	0.68	0.67	0.68	0.67	0.87	0.90	0.86	0.88	0.94	0.95	0.96	0.95	0.72	0.76	0.72	0.72

Table 3.7: Misclassification rates of machine learning models across different dataset sizes for Dataset–1.

Sample Size (N)	GAD-7 Dataset–1 Misclassification Rates				PHQ-9 Dataset–1 Misclassification Rates			
	DT	SVM	kNN	RF	DT	SVM	kNN	RF
N=100	0.85	0.85	0.40	0.85	0.80	0.80	0.45	0.80
N=200	0.55	0.47	0.27	0.62	0.62	0.62	0.32	0.55
N=300	0.50	0.30	0.18	0.47	0.40	0.43	0.28	0.45
N=400	0.35	0.19	0.18	0.41	0.54	0.45	0.36	0.43
N=500	0.35	0.13	0.10	0.23	0.41	0.30	0.32	0.42
N=600	0.39	0.14	0.15	0.24	0.33	0.32	0.32	0.33
N=700	0.39	0.13	0.07	0.21	0.37	0.24	0.29	0.31
N=800	0.40	0.14	0.10	0.33	0.33	0.26	0.24	0.29
N=900	0.30	0.11	0.08	0.22	0.39	0.22	0.21	0.26
N=1000	0.33	0.11	0.10	0.20	0.43	0.22	0.21	0.27
N=1100	0.25	0.11	0.07	0.20	0.34	0.17	0.14	0.21
N=1200	0.31	0.09	0.03	0.18	0.17	0.10	0.10	0.24
N=1300	0.35	0.11	0.05	0.24	0.33	0.14	0.05	0.24

Sample Size (N)	GAD-7 Dataset–1 Misclassification Rates				PHQ-9 Dataset–1 Misclassification Rates			
	DT	SVM	kNN	RF	DT	SVM	kNN	RF
N=1400	0.27	0.05	0.04	0.16	0.41	0.10	0.07	0.27
N=1500	0.27	0.09	0.04	0.16	0.32	0.13	0.06	0.28

3.4.2 Performance of Dataset 2 for GAD-7 and PHQ-9

The efficacy of machine learning classifiers in predicting anxiety levels across various sample sizes is demonstrated by the GAD-7 and PHQ-9 findings for Dataset-2, as shown in Table 3.8. A comparison of performance metrics GAD-7 and PHQ-9 for Dataset-2 is depicted in Fig. 3.3. All classifiers performed poorly with a sample size of 100, with accuracies ranging from 0.2 to 0.75. Table 3.9 shows misclassification rates of ML models across different dataset sizes for dataset 2. The same effect of decreasing misclassification rates on increasing the sample size has been observed for dataset 2, as seen in dataset 1. Compared to DT and KNN, SVM and RF consistently showed greater accuracy and performance metrics. As the sample size grew, this pattern persisted, with SVM and RF continuing to dominate in predicted accuracy. DT exhibited lower performance metrics, indicating its limitations in capturing complex relationships within the GAD-7 data.

Even though KNN did well in terms of recall, its accuracy was lower, suggesting a tendency to make more incorrect positive predictions. DT's performance indicators indicated a lag, highlighting its limited ability to detect non-linear correlations in the GAD-7 data. As per the plateau effect, further increases in sample size may not substantially improve predictive accuracy beyond a certain point. This underscores the importance of optimizing prediction performance using available data by adjusting model designs and selecting features, and by choosing the right sample size based on practical factors.

The findings from Dataset-2's PHQ-9 survey show how effectively ML models can forecast levels of depression. Across a sample of 100, all models showed moderate performance, with accuracy levels between 0.65 and 0.8. All of the classifiers improved their performance noticeably as the sample size increased. At N=300, RF and SVM consistently outperformed DT and kNN, showing higher accuracies and better overall performance. DT's performance measures had limitations in identifying complex correlations observed in the PHQ-9 data when compared to other methods.

Although its accuracy was lower, KNN demonstrated competitive recall performance, indicating a propensity for more false positive predictions. The fact that DT underperformed in performance metrics suggests that it has a limited capacity to identify non-linear correlations. The results show a plateau effect, which implies that increases in sample size beyond a certain point would not result in appreciable gains in prediction ability. To maximize prediction performance within the restrictions of the available data, researchers must concentrate on fine-tuning model topologies, feature selection, and other components of the work. This emphasizes

the significance of optimizing sample size based on practical concerns.

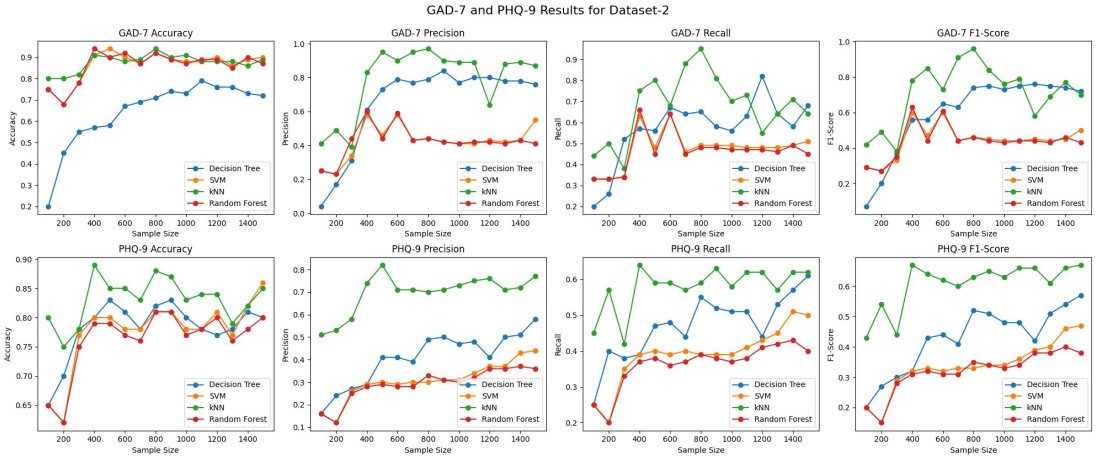


Figure 3.3: Comparison between the performance metrics of GAD-7 and PHQ-9 on varied dataset sizes for Dataset-2

Table 3.8: Performance of Dataset–2 for GAD-7 and PHQ-9 using different machine learning models.

Sample Size (N)	DT				SVM				kNN				RF			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
GAD-7 Results for Dataset–2																
N=100	0.20	0.04	0.20	0.07	0.75	0.25	0.33	0.29	0.80	0.41	0.44	0.42	0.75	0.25	0.33	0.29
N=200	0.45	0.17	0.26	0.20	0.68	0.23	0.33	0.27	0.80	0.49	0.50	0.49	0.68	0.23	0.33	0.27
N=300	0.55	0.31	0.52	0.37	0.78	0.34	0.34	0.33	0.82	0.39	0.38	0.38	0.78	0.44	0.34	0.35
N=400	0.57	0.61	0.57	0.56	0.91	0.58	0.63	0.60	0.91	0.83	0.75	0.78	0.94	0.60	0.66	0.63
N=500	0.58	0.73	0.56	0.56	0.94	0.46	0.48	0.47	0.90	0.95	0.80	0.85	0.90	0.44	0.45	0.44
N=600	0.67	0.79	0.67	0.65	0.90	0.58	0.64	0.60	0.88	0.90	0.68	0.73	0.92	0.59	0.64	0.61
N=700	0.69	0.77	0.66	0.63	0.87	0.43	0.46	0.44	0.89	0.95	0.88	0.91	0.87	0.43	0.45	0.44
N=800	0.71	0.79	0.72	0.74	0.92	0.44	0.49	0.46	0.94	0.97	0.95	0.96	0.92	0.44	0.48	0.46
N=900	0.74	0.84	0.74	0.75	0.89	0.42	0.49	0.45	0.90	0.90	0.81	0.84	0.89	0.42	0.48	0.44
N=1000	0.73	0.77	0.72	0.73	0.88	0.41	0.49	0.44	0.91	0.89	0.70	0.76	0.87	0.41	0.47	0.43
N=1100	0.79	0.80	0.74	0.75	0.88	0.41	0.48	0.44	0.88	0.89	0.73	0.79	0.89	0.42	0.47	0.44
N=1200	0.76	0.80	0.76	0.76	0.90	0.43	0.48	0.45	0.88	0.64	0.55	0.58	0.89	0.42	0.47	0.44
N=1300	0.76	0.78	0.75	0.75	0.86	0.42	0.48	0.44	0.88	0.88	0.64	0.69	0.85	0.41	0.46	0.43
N=1400	0.73	0.78	0.73	0.74	0.89	0.43	0.49	0.45	0.86	0.89	0.71	0.77	0.90	0.43	0.49	0.46
N=1500	0.72	0.76	0.72	0.72	0.90	0.55	0.51	0.50	0.89	0.87	0.64	0.70	0.87	0.41	0.45	0.43
PHQ-9 Results for Dataset–2																
N=100	0.65	0.16	0.25	0.20	0.65	0.16	0.25	0.20	0.80	0.51	0.45	0.43	0.65	0.16	0.25	0.20

Sample Size (N)	DT				SVM				kNN				RF			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
N=200	0.70	0.24	0.40	0.27	0.62	0.12	0.20	0.15	0.75	0.53	0.57	0.54	0.62	0.12	0.20	0.15
N=300	0.78	0.27	0.38	0.30	0.77	0.26	0.35	0.29	0.78	0.58	0.42	0.44	0.75	0.25	0.33	0.28
N=400	0.80	0.29	0.39	0.32	0.80	0.29	0.39	0.32	0.89	0.74	0.64	0.67	0.79	0.28	0.37	0.31
N=500	0.83	0.41	0.47	0.43	0.80	0.30	0.40	0.33	0.85	0.82	0.59	0.64	0.79	0.29	0.38	0.32
N=600	0.81	0.41	0.48	0.44	0.78	0.29	0.39	0.32	0.85	0.71	0.59	0.62	0.77	0.28	0.36	0.31
N=700	0.78	0.39	0.44	0.41	0.78	0.30	0.40	0.33	0.83	0.71	0.57	0.60	0.76	0.28	0.37	0.31
N=800	0.82	0.49	0.55	0.52	0.81	0.30	0.39	0.33	0.88	0.70	0.59	0.63	0.81	0.33	0.39	0.35
N=900	0.83	0.50	0.52	0.51	0.81	0.31	0.39	0.34	0.87	0.71	0.63	0.65	0.81	0.31	0.38	0.34
N=1000	0.80	0.47	0.51	0.48	0.78	0.31	0.39	0.34	0.83	0.73	0.58	0.63	0.77	0.30	0.37	0.33
N=1100	0.78	0.48	0.51	0.48	0.78	0.34	0.41	0.36	0.84	0.75	0.62	0.66	0.78	0.32	0.38	0.34
N=1200	0.77	0.41	0.44	0.42	0.81	0.37	0.43	0.39	0.84	0.76	0.62	0.66	0.80	0.36	0.41	0.38
N=1300	0.78	0.50	0.53	0.51	0.77	0.37	0.45	0.40	0.79	0.71	0.57	0.61	0.76	0.36	0.42	0.38
N=1400	0.81	0.51	0.57	0.54	0.82	0.43	0.51	0.46	0.82	0.72	0.62	0.66	0.78	0.37	0.43	0.40
N=1500	0.80	0.58	0.61	0.57	0.86	0.44	0.50	0.47	0.85	0.77	0.62	0.67	0.80	0.36	0.40	0.38

Table 3.9: Misclassification rates of machine learning models across different dataset sizes for Dataset–2

Sample Size (N)	GAD-7 Dataset–2 Misclassification Rates				PHQ-9 Dataset–2 Misclassification Rates			
	DT	SVM	kNN	RF	DT	SVM	kNN	RF
N=100	0.80	0.25	0.20	0.25	0.35	0.35	0.20	0.84
N=200	0.55	0.32	0.20	0.32	0.30	0.38	0.25	0.88
N=300	0.45	0.22	0.18	0.22	0.22	0.23	0.22	0.75
N=400	0.43	0.09	0.09	0.06	0.20	0.20	0.11	0.72
N=500	0.42	0.06	0.10	0.10	0.17	0.20	0.15	0.71
N=600	0.33	0.10	0.12	0.08	0.19	0.22	0.15	0.72
N=700	0.31	0.13	0.11	0.13	0.22	0.22	0.17	0.72
N=800	0.29	0.08	0.06	0.08	0.18	0.19	0.12	0.67
N=900	0.26	0.11	0.10	0.11	0.17	0.19	0.13	0.69
N=1000	0.27	0.12	0.09	0.13	0.20	0.22	0.17	0.70
N=1100	0.21	0.12	0.12	0.11	0.22	0.22	0.16	0.68
N=1200	0.24	0.10	0.12	0.11	0.23	0.19	0.16	0.64
N=1300	0.24	0.14	0.12	0.15	0.22	0.23	0.21	0.64

Sample Size (N)	GAD-7 Dataset–2 Misclassification Rates				PHQ-9 Dataset–2 Misclassification Rates			
	DT	SVM	kNN	RF	DT	SVM	kNN	RF
N=1400	0.27	0.11	0.14	0.10	0.19	0.18	0.18	0.63
N=1500	0.28	0.10	0.11	0.13	0.20	0.14	0.15	0.64

3.4.3 Performance of Dataset 3 for GAD-7 and PHQ-9:

Interesting trends in the predicted efficacy of ML classifiers across a range of sample sizes are shown by the GAD-7 findings for Dataset-3, as shown in Table 3.10. All classifiers performed moderately with a sample size of 100, with accuracies ranging from 0.3 to 0.55. Table 3.11 shows misclassification rates of ML models across different dataset sizes for dataset 3. The same effect of decreasing misclassification rates on increasing the sample size has been observed for Dataset 3, as seen in Datasets 1 and 2. Fig. 3.5 and Fig. 3.6 depict misclassification rates for ML models for GAD-7 and PHQ-9, respectively. The efficacy of DT, SVM, and RF in capturing the underlying patterns in the GAD-7 scores was demonstrated by their near-perfect accuracy, precision, recall, and F1-score at N=300. Reduced kNN performance metrics point to issues with the GAD-7 data's complexity. All classifiers performed well as the sample size grew, with DT, SVM, and RF consistently surpassing kNN on a range of criteria. A comparison of performance metrics GAD-7 and PHQ-9 for Dataset-3 is depicted in Fig. 3.4.



Figure 3.4: Comparing the GAD-7 and PHQ-9 performance measures on various sizes of Dataset-3

The PHQ-9 findings for Dataset-3 demonstrate how the prediction effectiveness of ML classifiers varies with sample size. With accuracies ranging from 0.2 to 0.6, all classifiers performed poorly when there was a sample size of 100. All classifiers, however, performed better as the sample size reached 300; DT, SVM, and RF all received good marks. These models were better able to capture the underlying patterns in the PHQ-9 scores with a moderately sized dataset. As the sample size increased, further improvements were observed across all classifiers. SVM and RF consistently achieved near-perfect performance at larger sample sizes (N=600 and above), demonstrating their ability to reliably predict anxiety and depression levels using the GAD-7 and PHQ-9 questionnaires, respectively.

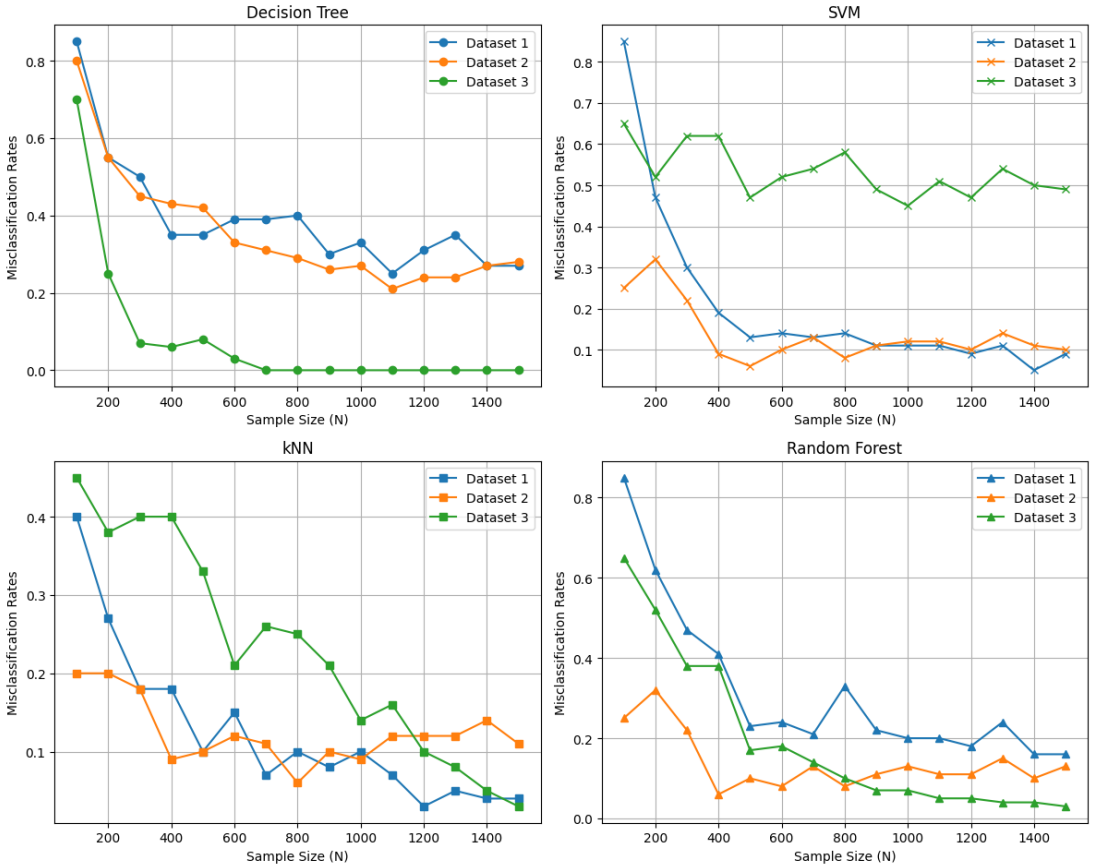


Figure 3.5: Comparative analysis of misclassification rates for different machine learning models for GAD-7

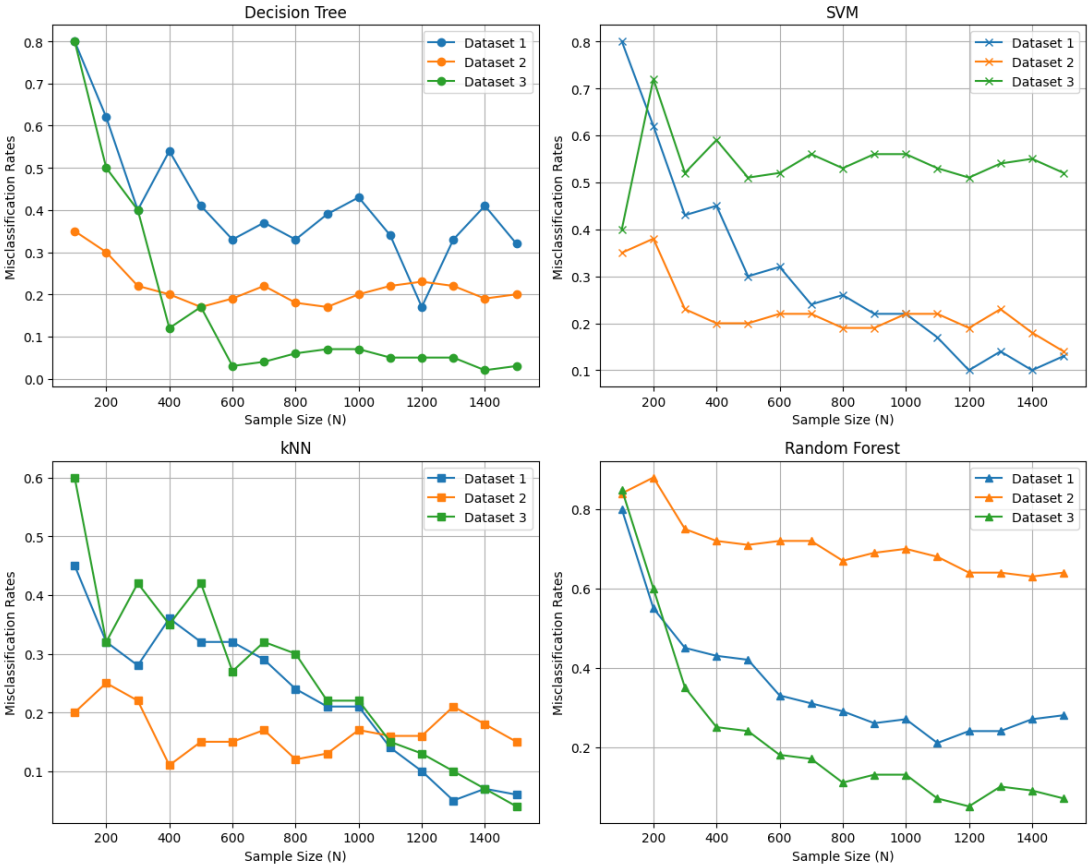


Figure 3.6: Comparative analysis of misclassification rates for different machine learning models for PHQ-9

Table 3.10: Performance of Dataset–3 for GAD-7 and PHQ-9 using different machine learning models

Sample Size (N)	GAD-7 Results for Dataset–3								PHQ-9 Results for Dataset–3							
	DT				SVM				kNN				RF			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
N=100	0.30	0.07	0.25	0.12	0.35	0.09	0.25	0.13	0.20	0.05	0.25	0.08	0.15	0.04	0.25	0.07
N=200	0.75	0.38	0.50	0.42	0.48	0.12	0.25	0.16	0.50	0.20	0.40	0.27	0.40	0.20	0.27	0.21
N=300	0.93	0.65	0.75	0.69	0.38	0.10	0.25	0.14	0.60	0.24	0.40	0.30	0.65	0.40	0.53	0.43
N=400	0.94	0.67	0.75	0.70	0.38	0.09	0.25	0.14	0.88	0.51	0.60	0.54	0.75	0.43	0.50	0.46
N=500	0.92	0.66	0.75	0.70	0.53	0.19	0.28	0.22	0.83	0.52	0.60	0.55	0.76	0.45	0.53	0.48
N=600	0.97	0.95	0.96	0.95	0.48	0.15	0.26	0.18	0.97	0.76	0.80	0.78	0.82	0.58	0.60	0.58
N=700	1.00	1.00	1.00	1.00	0.46	0.16	0.26	0.19	0.96	0.74	0.80	0.77	0.83	0.61	0.65	0.63
N=800	1.00	1.00	1.00	1.00	0.42	0.15	0.26	0.17	0.94	0.74	0.80	0.76	0.89	0.65	0.71	0.68
N=900	1.00	1.00	1.00	1.00	0.51	0.20	0.29	0.23	0.93	0.72	0.80	0.75	0.87	0.62	0.66	0.64
N=1000	1.00	1.00	1.00	1.00	0.55	0.23	0.30	0.26	0.93	0.72	0.80	0.75	0.87	0.62	0.66	0.64
N=1100	1.00	1.00	1.00	1.00	0.49	0.21	0.29	0.24	0.95	0.74	0.80	0.76	0.93	0.71	0.78	0.74

Sample Size (N)	GAD-7 Results for Dataset–3								PHQ-9 Results for Dataset–3							
	DT				SVM				kNN				RF			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
N=1200	1.00	1.00	1.00	1.00	0.53	0.24	0.31	0.26	0.95	0.87	0.87	0.86	0.95	0.73	0.78	0.75
N=1300	1.00	1.00	1.00	1.00	0.46	0.20	0.30	0.24	0.95	0.91	0.91	0.91	0.90	0.70	0.76	0.72
N=1400	1.00	1.00	1.00	1.00	0.50	0.23	0.32	0.26	0.98	0.96	0.96	0.96	0.91	0.70	0.76	0.73
N=1500	1.00	1.00	1.00	1.00	0.51	0.48	0.35	0.30	0.97	0.94	0.93	0.92	0.93	0.89	0.93	0.91

Table 3.11: Misclassification rates of machine learning models across different dataset sizes for Dataset–3

Sample Size (N)	GAD-7 Dataset–3 Misclassification Rates				PHQ-9 Dataset–3 Misclassification Rates			
	DT	SVM	kNN	RF	DT	SVM	kNN	RF
N=100	0.70	0.65	0.45	0.65	0.80	0.40	0.60	0.85
N=200	0.25	0.52	0.38	0.52	0.50	0.72	0.32	0.60
N=300	0.07	0.62	0.40	0.38	0.40	0.52	0.42	0.35
N=400	0.06	0.62	0.40	0.38	0.12	0.59	0.35	0.25
N=500	0.08	0.47	0.33	0.17	0.17	0.51	0.42	0.24
N=600	0.03	0.52	0.21	0.18	0.03	0.52	0.27	0.18
N=700	0.00	0.54	0.26	0.14	0.04	0.56	0.32	0.17
N=800	0.00	0.58	0.25	0.10	0.06	0.53	0.30	0.11
N=900	0.00	0.49	0.21	0.07	0.07	0.56	0.22	0.13
N=1000	0.00	0.45	0.14	0.07	0.07	0.56	0.22	0.13
N=1100	0.00	0.51	0.16	0.05	0.05	0.53	0.15	0.07
N=1200	0.00	0.47	0.10	0.05	0.05	0.51	0.13	0.05
N=1300	0.00	0.54	0.08	0.04	0.05	0.54	0.10	0.10

Sample Size (N)	GAD-7 Dataset–3 Misclassification Rates				PHQ-9 Dataset–3 Misclassification Rates			
	DT	SVM	kNN	RF	DT	SVM	kNN	RF
N=1400	0.00	0.50	0.05	0.04	0.02	0.55	0.07	0.09
N=1500	0.00	0.49	0.03	0.03	0.03	0.52	0.04	0.07

3.4.4 Results of Statistical Analysis on Model Performance

The survey responses were statistically analyzed using the ANOVA test, which quantifies the improvement in performance metrics as the dataset size increases, as shown below in Table 3.12.

Table 3.12: Significant p-values (at the 5% level) and F-statistic from ANOVAs of effects of dataset size on performance metrics

GAD-7 ANOVA Test Results								
Metric	Accuracy		Precision		Recall		F1 Score	
Model	F Statistic	P value	F Statistic	P value	F Statistic	P value	F Statistic	P value
DT	16.22	< 0.01	3.30	0.04	12.48	< 0.01	6.22	< 0.01
SVM	38.73	< 0.01	67.59	< 0.01	32.55	< 0.01	41.20	< 0.01
kNN	5.47	< 0.01	5.12	0.03	7.04	< 0.01	4.99	0.01
RF	4.48	0.04	5.90	< 0.01	6.26	0.01	4.72	0.02
PHQ-9 ANOVA Test Results								
Metric	Accuracy		Precision		Recall		F1 Score	
Model	F Statistic	P value	F Statistic	P value	F Statistic	P value	F Statistic	P value
DT	10.26	< 0.01	4.38	0.02	9.01	< 0.01	5.25	< 0.01
SVM	25.16	< 0.01	22.21	< 0.01	22.23	< 0.01	20.97	< 0.01
kNN	5.28	0.03	9.17	< 0.01	12.75	< 0.01	6.73	< 0.01
RF	13.79	< 0.01	13.04	< 0.01	14.57	< 0.01	10.98	< 0.01

Together, the F-statistic and p-value values demonstrate statistically significant correlations between model performance metrics and dataset size. Since the null hypothesis predicted that there would be no meaningful correlation between the size of the dataset and the model's performance, it is rejected in this instance.

To strengthen the statistical analysis presented in this thesis, the assumptions underlying the parametric analysis were subsequently examined. In particular, the Shapiro–Wilk test was used to assess the normality of the performance observations. The Shapiro–Wilk test indicated significant departures from normality in few cases therefore, Friedman test was used to further assess the results. The Friedman test was used to compare the three datasets at corresponding sample-size levels. Significant differences were observed for all classifiers and performance metrics for both PHQ-9 and GAD-7 (p ¡ .05). For PHQ-9, the Friedman χ^2 statistic values ranged from 10.964 to 22.068, while for GAD-7 they ranged from 6.533 to 26.678 as shown in Table 3.13. However, the Friedman test assesses differences among datasets and does not directly determine the direction of the relationship between dataset size and performance.

Table 3.13: Friedman Test Results for PHQ-9 and GAD-7

Scale	Model	Accuracy		Precision		Recall		F1-score	
		χ^2	p	χ^2	p	χ^2	p	χ^2	p
PHQ-9	DT	20.035	< 0.001	14.271	< 0.001	21.655	< 0.001	18.345	< 0.001
PHQ-9	SVM	20.800	< 0.001	22.068	< 0.001	22.068	< 0.001	20.933	< 0.001
PHQ-9	KNN	20.800	< 0.001	10.964	0.004	20.793	< 0.001	19.356	< 0.001
PHQ-9	RF	20.133	< 0.001	19.356	< 0.001	18.610	< 0.001	19.172	< 0.001
GAD-7	DT	24.966	< 0.001	20.276	< 0.001	22.448	< 0.001	24.138	< 0.001
GAD-7	SVM	20.414	< 0.001	24.400	< 0.001	26.678	< 0.001	24.400	< 0.001
GAD-7	KNN	10.897	0.004	10.034	0.007	16.933	< 0.001	16.933	< 0.001
GAD-7	RF	9.254	0.010	6.533	0.038	6.759	0.034	6.933	0.031

Note: $\alpha = 0.05$. The Friedman test compares the three datasets at corresponding sample-size levels.

Therefore, Spearman rank correlation with Holm adjustment was used as the primary non-parametric analysis to examine the association between dataset size and model performance. Linear regression was additionally used to quantify the change in performance per 100 additional observations, together with R^2 and 95% confidence intervals for the slope.

As summarized in Table 3.14, PHQ-9 showed significant positive associations in 37 of 48 model–dataset–metric combinations after Holm correction, comprising 6/12 for Accuracy, 10/12 for Precision, 10/12 for Recall, and 11/12 for F1-score. For GAD-7, 35 of 48 combinations were significant, comprising 8/12 for Accuracy and 9/12 for each of Precision, Recall, and F1-score. Overall, 72 of 96 combinations (75%) showed a significant positive association with dataset size. The relationships were generally stronger for Dataset 1 and Dataset 3, whereas Dataset 2 showed weaker or non-significant associations in several cases.

Table 3.14: Summary of Spearman Rank Correlation Between Dataset Size and Model Performance After Holm Adjustment

GAD-7 Spearman Correlation Results					
Model	Dataset	Accuracy	Precision	Recall	F1 Score
DT	Dataset 1	↑	↑	↑	↑
	Dataset 2	↑	↑	↑	↑
	Dataset 3	↑	↑	↑	↑
SVM	Dataset 1	↑	↑	↑	↑
	Dataset 2	ns	ns	ns	ns
	Dataset 3	ns	↑	↑	↑
kNN	Dataset 1	↑	↑	↑	↑
	Dataset 2	ns	ns	ns	ns
	Dataset 3	↑	↑	↑	↑
RF	Dataset 1	↑	↑	↑	↑
	Dataset 2	ns	ns	ns	ns
	Dataset 3	↑	↑	↑	↑
PHQ-9 Spearman Correlation Results					
Model	Dataset	Accuracy	Precision	Recall	F1 Score
DT	Dataset 1	ns	↑	↑	↑
	Dataset 2	ns	↑	↑	↑
	Dataset 3	↑	↑	↑	↑
SVM	Dataset 1	↑	↑	↑	↑
	Dataset 2	ns	↑	↑	↑
	Dataset 3	ns	ns	ns	ns
kNN	Dataset 1	↑	↑	↑	↑
	Dataset 2	ns	ns	ns	↑
	Dataset 3	↑	↑	↑	↑
RF	Dataset 1	↑	↑	↑	↑
	Dataset 2	ns	↑	↑	↑
	Dataset 3	↑	↑	↑	↑

The regression results generally supported the positive trends identified by Spearman correlation. For example, PHQ-9 DT-Dataset 3 Accuracy increased by approximately 3.77 percentage points per 100 additional observations, while KNN-Dataset 3 and RF-Dataset 3 also showed strong positive relationships. Regression was used as complementary evidence to quantify the magnitude of change rather than as the primary inferential test.

Statistical Analysis was appropriate because the performance observations for the different models were obtained repeatedly across the same set of sample-size conditions. Thus, the

observations across sample sizes were not treated as completely independent, and the repeated-measures design accounts for the dependence among measurements obtained under the same experimental conditions. The data trend supports that as the dataset size increases, the ML models' performance improves for many metrics and models with low p-values.

3.5 Discussion

This section discusses the analysis of the results of GAD-7, PHQ-9, and the statistical test (ANOVA) across three different datasets. Various ML classifiers such as DT, SVM, kNN, and RF were assessed for their predictive performance using sample sizes ranging from N=100 to N=1500.

3.5.1 GAD-7 Results

Sample size affected the classifiers' ability to predict GAD-7 accurately in all datasets. For instance, in Dataset-1, small samples made it hard for classifiers to achieve good levels of precision and accuracy. Thus, predicting anxiety levels using the GAD-7 questionnaire is not simple in a small dataset. In essence, SVM and RF outperformed others significantly as sample sizes increased, especially above N=500. Thereby, when working on larger data sets, SVM and RF classifiers can capture more details about the measurement of anxiety by the GAD-7 questionnaire. The same trend was evident in Dataset-2, which indicated that higher sample sizes were associated with superior performance of SVM and RF classifiers. Notably, for these classifiers, nearly flawless results were obtained with large sample sizes (N=1000 and above), hence their reliability in accurately predicting anxiety levels from the use of the GAD-7 questionnaire. Therefore, this means that if given enough data, ML classifiers may be able to understand subtleties about assessing anxiety via GAD-7.

In Dataset 3, both SVM and RF classifiers had a better performance on this data, in comparison with other classifiers, showing similar behavior for a range of sample sizes. It should be mentioned that these classifiers did a reasonable job of predicting anxiety levels based on the GAD-7 questionnaire. In terms of prediction accuracy, they performed about as well as they could have for a bigger sample size of N=600 and up. These findings suggest that SVM and RF can improve the anxiety assessment process and enable the development of more sophisticated intervention strategies.

3.5.2 PHQ-9 Results

The analysis of the results of PHQ-9 across all three datasets also revealed interesting variations in prediction accuracies for different sample sizes. The initial trouble of the classifier in Dataset-1 to achieve high accuracy and precision with a smaller number of samples showed the difficulty of accurately predicting levels of depression using the PHQ-9 questionnaire for a small dataset. However, SVM and RF classifiers performed better as the sample sizes increased, especially around N=300. Larger sample sizes (N=800 and above) have achieved almost perfect recall, accuracy, precision, and F1 scores.

Similar patterns were also present in Dataset 2, in which SVM and RF classifiers outperformed all others at both the set minimum sample size and all higher sample sizes. It should also be noted that almost perfect prediction accuracy was shown by the classifiers at larger sample sizes (N=600 and above). This can be taken to mean that the PHQ-9 questionnaire was able to accurately predict depressive levels.

The performance across different sample sizes in both the SVM and RF classifiers of Dataset-3 was excellent. At N=1000 and above, they achieved an almost perfect classification report. This is indicative of how ML classifiers, especially SVM and RF, are capable of enhancing the evaluation of depression, enabling the development of more focused treatment approaches.

All dataset/algorithm combinations experienced overfitting at some point of increasing dataset size.

3.5.3 Statistical Results

This work emphasizes that the size of a dataset is especially important in the performance enhancement of ML models adopted for assessment related to mental health. For the GAD-7 dataset and PHQ-9 dataset, our ANOVA study shows distinct relationships between dataset size and the fluctuation of several performance indicators, including accuracy, precision, recall, and F1-score across multiple models. The SVM model performs best when the dataset is larger, as seen by the highest F-statistics and lowest p-values across all measures. This result also agrees with the intuition that larger datasets can capture more diverse patterns, which should translate into better generalization and improved model performance. Also, the gain from the latter two datasets was for all other models, too, statistically significant.

3.5.4 Cross-Validation analysis

5-fold cross-validation results are presented in Table 3.15. All of the models' cross-validation scores have been found to rise with dataset size, indicating that the models are doing well in terms of generalisation, as shown in Fig. 3.7 and Fig. 3.8. The characteristics of Datasets 1 and 2 differ from those of Datasets 3 used in our work (Table 3.1). We also repeated the approach on the Wine dataset to ensure generalisability "(https://www.kaggle.com/datasets/yasserh/wine-quality-dataset)", and similar results were obtained, as attached in Table 3.16. The similarity in results for datasets 1, 2, 3, and the Wine dataset indicates that variation in the performance of classifiers is not totally dependent on the characteristics of GAD-7 and PHQ-9, but is mainly dependent on the dataset size taken. Larger sample sizes yield better classifier performances consistently, the model learns the pattern efficiently, and generalizes well.

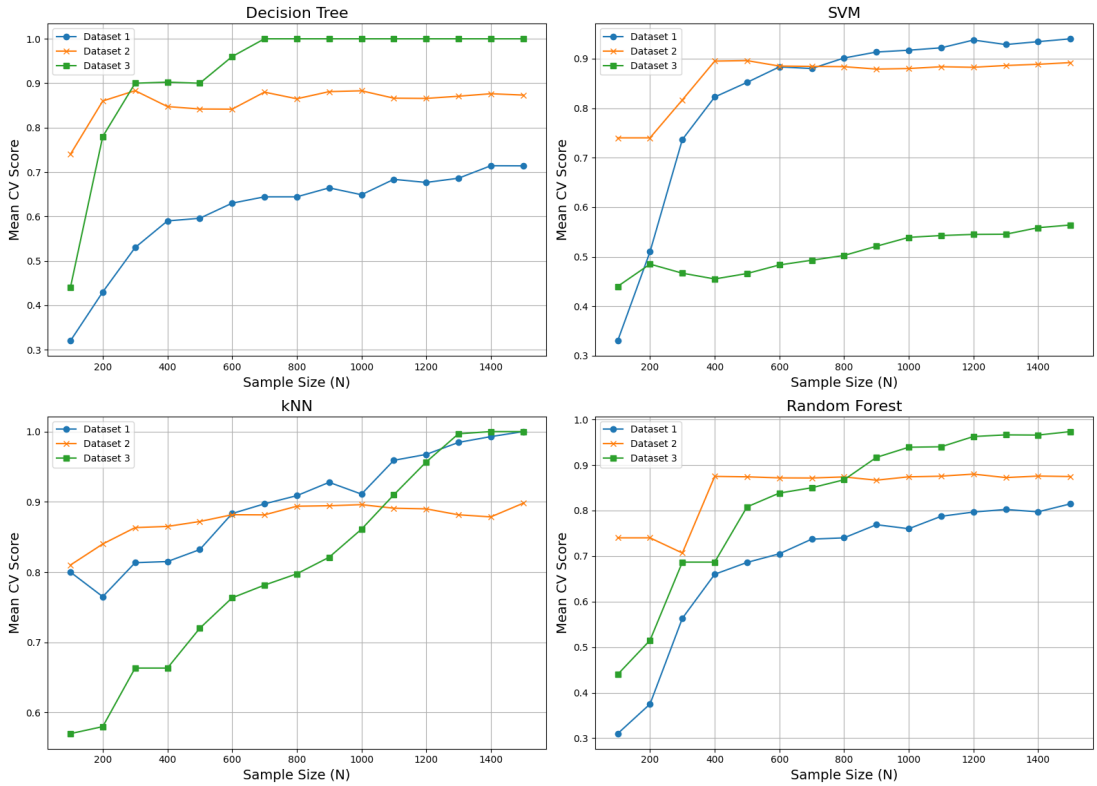


Figure 3.7: Comparative analysis of cross-validation scores for different machine learning models for GAD-7

To summarize the results, the findings of this work have great implications concerning mental health diagnostics, especially in times of pandemics. During pandemics and in general global crises, traditional care in the mental health field is constrained due to limited resources, increased demand, and social distancing measures.

ML models can serve to help bridge gaps in care delivery, especially for the underserved

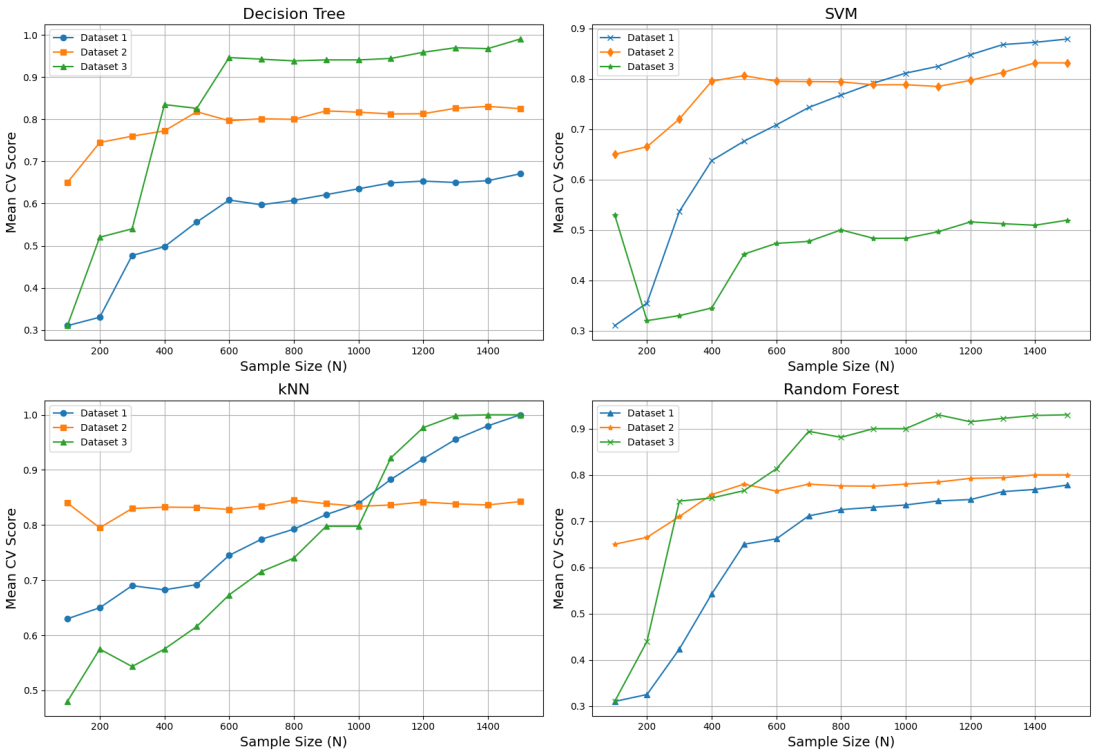


Figure 3.8: Comparative analysis of cross-validation scores for different machine learning models for PHQ-9

or areas lacking real mental health infrastructure, by enabling rapid, remote assessment. The integration of ML tools with existing telehealth platforms could support the timely identification of individuals at risk, helping healthcare systems prioritize interventions and improve outcomes. During pandemics, where stressors such as isolation, economic uncertainty, and health anxieties are pervasive, the use of ML-based diagnostics can alleviate the psychological impact through early detection and scalable screening solutions.

Further, the potential of SVM and RF classifiers maintaining peak performance across datasets attests to the generalizability of these methods and therefore lends these solutions greater flexibility and potential for adoption in wider contexts beyond pandemics, into routine screening programs of mental health or other public health crises.

Table 3.15: CV scores of all datasets for GAD-7 and PHQ-9 using different machine learning models

GAD-7

Sample Size (N)	DT			SVM			kNN			RF		
	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
N=100	0.32	0.74	0.44	0.33	0.74	0.44	0.80	0.81	0.57	0.31	0.74	0.44
N=200	0.43	0.86	0.78	0.51	0.74	0.48	0.76	0.84	0.58	0.37	0.74	0.51
N=300	0.53	0.88	0.90	0.74	0.82	0.47	0.81	0.86	0.66	0.56	0.71	0.69
N=400	0.59	0.85	0.90	0.82	0.89	0.45	0.81	0.86	0.66	0.66	0.87	0.69
N=500	0.60	0.84	0.90	0.85	0.89	0.46	0.83	0.87	0.72	0.69	0.87	0.80
N=600	0.63	0.84	0.96	0.88	0.88	0.48	0.88	0.88	0.76	0.70	0.87	0.84
N=700	0.64	0.88	1.00	0.88	0.88	0.49	0.90	0.88	0.78	0.73	0.87	0.85
N=800	0.64	0.86	1.00	0.90	0.88	0.50	0.90	0.89	0.80	0.74	0.87	0.87
N=900	0.66	0.88	1.00	0.91	0.88	0.52	0.93	0.89	0.82	0.77	0.87	0.92
N=1000	0.65	0.88	1.00	0.92	0.88	0.54	0.91	0.89	0.86	0.76	0.87	0.94
N=1100	0.68	0.87	1.00	0.92	0.88	0.54	0.96	0.89	0.91	0.79	0.87	0.94

Sample Size (N)	DT			SVM			kNN			RF		
	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
GAD-7 (continued)												
N=1200	0.68	0.86	1.00	0.94	0.88	0.54	0.97	0.89	0.96	0.80	0.88	0.96
N=1300	0.69	0.87	1.00	0.93	0.89	0.54	0.98	0.88	1.00	0.80	0.87	0.97
N=1400	0.71	0.88	1.00	0.93	0.89	0.57	0.99	0.88	1.00	0.80	0.87	0.96
N=1500	0.71	0.87	1.00	0.94	0.89	0.56	1.00	0.90	1.00	0.81	0.87	0.97
PHQ-9												
N=100	0.31	0.65	0.31	0.31	0.65	0.53	0.63	0.84	0.48	0.31	0.74	0.31
N=200	0.33	0.74	0.52	0.35	0.66	0.32	0.65	0.79	0.57	0.32	0.74	0.44
N=300	0.48	0.76	0.54	0.54	0.72	0.33	0.69	0.83	0.54	0.42	0.71	0.74
N=400	0.50	0.77	0.83	0.64	0.79	0.34	0.68	0.83	0.57	0.54	0.87	0.75
N=500	0.56	0.82	0.83	0.68	0.81	0.45	0.69	0.83	0.62	0.65	0.87	0.77
N=600	0.61	0.80	0.95	0.71	0.79	0.47	0.74	0.83	0.67	0.66	0.87	0.81
N=700	0.60	0.80	0.94	0.74	0.79	0.48	0.77	0.83	0.71	0.71	0.87	0.89

GAD-7 (continued)

Sample Size (N)	DT			SVM			kNN			RF		
	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
N=800	0.61	0.80	0.94	0.77	0.79	0.50	0.80	0.84	0.74	0.72	0.87	0.88
N=900	0.62	0.82	0.94	0.79	0.79	0.48	0.82	0.84	0.80	0.73	0.87	0.90
N=1000	0.63	0.82	0.94	0.81	0.79	0.48	0.84	0.83	0.80	0.73	0.87	0.90
N=1100	0.65	0.81	0.94	0.82	0.78	0.50	0.88	0.84	0.92	0.74	0.87	0.93
N=1200	0.65	0.81	0.96	0.85	0.80	0.51	0.92	0.84	0.98	0.74	0.88	0.91
N=1300	0.65	0.83	0.97	0.87	0.81	0.51	0.95	0.84	1.00	0.76	0.87	0.92
N=1400	0.65	0.83	0.97	0.87	0.83	0.51	0.98	0.84	1.00	0.77	0.87	0.93
N=1500	0.67	0.83	0.99	0.89	0.83	0.52	1.00	0.84	1.00	0.78	0.87	0.93

Table 3.16: Performance of wine dataset for different machine learning models

Sample Size (N)	DT					SVM					kNN					RF				
	Accuracy	Precision	Recall	F1-score	CV score	Accuracy	Precision	Recall	F1-score	CV score	Accuracy	Precision	Recall	F1-score	CV score	Accuracy	Precision	Recall	F1-score	CV score
N=100	0.4	0.13	0.33	0.19	0.4	0.4	0.13	0.33	0.19	0.4	0.8	0.83	0.82	0.81	0.82	0.4	0.13	0.33	0.19	0.4
N=200	0.85	0.86	0.85	0.86	0.83	0.4	0.13	0.33	0.19	0.39	0.85	0.89	0.86	0.86	0.845	0.8	0.8	0.75	0.76	0.81
N=300	0.9	0.9	0.9	0.9	0.89	0.33	0.11	0.33	0.17	0.4	0.9	0.9	0.9	0.9	0.9	0.97	0.97	0.97	0.97	0.97
N=400	0.9	0.91	0.92	0.91	0.9	0.42	0.26	0.33	0.24	0.38	0.92	0.93	0.93	0.93	0.94	0.95	0.95	0.96	0.96	0.96
N=500	0.92	0.92	0.93	0.92	0.9	0.4	0.27	0.33	0.24	0.4	0.93	0.94	0.91	0.92	0.99	1	1	1	1	0.98
N=600	0.92	0.92	0.93	0.92	0.9	0.4	0.27	0.33	0.24	0.4	0.93	0.94	0.91	0.92	0.99	1	1	1	1	0.97
N=700	0.92	0.92	0.92	0.92	0.93	0.54	0.39	0.46	0.41	0.5	1	1	1	1	1	0.98	0.97	0.98	0.98	0.98
N=800	0.9	0.91	0.92	0.91	0.93	0.51	0.36	0.42	0.37	0.51	0.98	0.98	0.97	0.98	1	0.99	0.98	0.99	0.99	0.99
N=900	0.93	0.95	0.93	0.93	0.94	0.56	0.39	0.46	0.41	0.52	1	1	1	1	1	1	1	1	1	1
N=1000	0.92	0.94	0.91	0.92	0.95	0.55	0.37	0.44	0.4	0.51	1	1	1	1	1	1	1	1	1	0.99
N=1100	0.93	0.92	0.94	0.93	0.95	0.55	0.37	0.48	0.41	0.52	1	1	1	1	1	0.99	0.98	0.99	0.99	0.99
N=1200	0.92	0.94	0.92	0.92	0.95	0.51	0.36	0.47	0.39	0.53	1	1	1	1	1	0.99	1	1	1	0.99
N=1300	0.94	0.94	0.95	0.94	0.95	0.53	0.37	0.49	0.41	0.55	1	1	1	1	1	1	1	1	1	0.99
N=1400	0.95	0.95	0.97	0.96	0.96	0.55	0.37	0.48	0.41	0.55	1	1	1	1	1	0.99	0.99	0.99	0.99	0.99
N=1500	0.97	0.97	0.97	0.97	0.96	0.53	0.36	0.47	0.41	0.55	1	1	1	1	1	0.98	0.98	0.99	0.98	0.99

3.5.5 Limitations

It is essential to accept some limitations of our work, such as the explicit population under investigation, the assumption of questionnaire validity, and the likelihood of bias in the datasets. The varied performance of dataset-3 could be because of the absence of questions in the dataset for GAD-7 and PHQ-9. Further exploration could investigate different information sources and other emotional well-being evaluation instruments for a more extensive understanding.

Online-based methods of data collection may also have caused biases, such as self-selection bias, where those who are more at ease with their use of technology or perhaps apprehensive about their mental states might have reported higher rates. In addition, those with active anxiety or depression may have been more likely to participate in such studies, biasing the sample toward higher symptom severity and limiting generalizability to the general population.

It's vital to take note that while artificial intelligence can be a significant guide in diagnosing depression and anxiety, it shouldn't supplant the skill of psychological well-being experts. All things considered, it can act as a significant screening device, distinguishing people who might profit from additional clinical evaluation and conduct. When developing and implementing such models in healthcare settings, ethical concerns and the privacy of patients are also of the utmost importance.

3.6 Summary

This chapter used data from the GAD-7 and PHQ-9 questionnaires to investigate how dataset size affects machine learning models' ability to detect anxiety and depression. Three separate datasets with sample sizes ranging from N=100 to N=1500 were used for experimental assessments using DT, SVM, KNN, and RF classifiers. Precision, Accuracy, recall, F1-score, cross-validation, and statistical significance testing were used to evaluate performance.

The findings show that model reliability and generalisation are significantly influenced by dataset size. Across all datasets, classifier performance improved consistently as the sample size increased, with SVM and RF models showing superior scalability and robustness. Performance stabilization was observed beyond specific dataset thresholds, indicating a plateau effect where further increases in data yielded marginal improvements. Statistical analysis using ANOVA confirmed that the observed performance gains with increasing dataset size were statistically significant, particularly for the SVM model.

Cross-validation analysis further validated the generalizability of the models, with consistent

performance trends observed across multiple datasets, including an external benchmark dataset. These findings indicate that classifier performance is influenced more by dataset size than by the specific characteristics of the psychological assessment instruments.

Overall, this chapter provides empirical evidence for identifying minimum dataset size requirements necessary for stable and reliable ML-based mental health prediction. The insights gained highlight the importance of adequate data collection strategies and motivate the development of advanced and hybrid modeling approaches, which are explored in the subsequent chapter.

Chapter 4

Exploring Deep Learning Architectures for Mental Health Assessment: A Comparative Analysis

4.1 Introduction

Using standardised psychological questionnaires, this chapter investigates the use of DL architectures for automated anxiety and depression assessments. Building upon the ML-based analyses presented in the previous chapters, this chapter investigates whether DL models can capture more complex, non-linear relationships within mental health data and thereby improve predictive performance.

Because DL approaches can learn hierarchical feature representations, they are especially well-suited for analysing high-dimensional and structured data. These models have the ability to find subtle patterns in questionnaire replies that conventional machine learning techniques would miss in the context of mental health assessment. This chapter focuses on analysing anxiety and depression using information from proven screening tools such as the GAD-7 and PHQ-9.

Five DL architectures—NN, CNN, RNN, LSTM networks, and TabNet—are used and assessed for multiclass categorisation of anxiety and depression severity levels in order to do this. In terms of prediction accuracy, robustness, and generalisation capacity, the comparative analysis seeks to evaluate each model's advantages and disadvantages. The comparative analysis seeks to evaluate each model's advantages and disadvantages in terms of prediction accuracy,

robustness, and generalisation capacity. The results of this chapter advance knowledge of the function of deep learning in mental health diagnostics and offer guidance for the creation of better and hybrid modelling techniques in later chapters. The critical review presented in Section 2.8.2 identified a gap in the systematic comparative evaluation of DL architectures for anxiety and depression prediction. Accordingly, this chapter investigates the performance of multiple DL models using consistent datasets and experimental settings. The results of this chapter advance knowledge of DL's function in mental health diagnostics and offer guidance for the creation of enhanced and hybrid modelling techniques in later chapters.

4.2 Proposed Methodology

To achieve generalizability and scalability, we applied four DL models on three datasets.

4.2.1 Datasets Used

The first dataset was gathered by Islam et al., and their findings highlight the psychological toll of the epidemic by showing a considerable prevalence of elevated anxiety and sadness [100]. High levels of depression and anxiety were found among the participants, with 15% of the 476 students reporting moderately severe depression and 18.1% reporting severe anxiety. According to Carollo et al.'s analysis of the second dataset, which included 435 respondents in the first wave, it may be especially important to address two symptoms when implementing lockdown limits: self-perceived loneliness and sad symptoms [38]. Qasrawi et al. gathered the third dataset, which included measurements of maternal depression and anxiety among 3569 expectant and recent mothers [188]. Moderate to high levels of depression were reported. Table 4.1 provides an overview of the datasets used in this work. All these studies employed a traditional approach to detecting mental health problems using binary classification, paving the way for deeper analysis of the mental health problems using DL. Detailed characteristics and anonymized illustrative snapshots of these datasets are provided in Chapter 3 (Table 3.2, Table 3.3 and Table 3.4).

Table 4.1: An overview of the datasets utilised in earlier research.

Reference	**Dataset Size**	**DL Models used?**
[100]	$N = 476$	$\times$
[38]	$N = 435$	$\times$
[188]	$N = 3569$	$\times$

4.2.2 Mental Health Assessment

In addition to collecting demographic information from respondents, each dataset administered tests like the PHQ-9 and GAD-7 to measure the level of anxiety and depression caused by COVID-19, as explained in detail in section 2.5. PHQ-9 and GAD-7 questions were fed into ML models and the output variable was "depression_class" and "anxiety_class". "anxiety_class" measured the severity of anxiety with values ranging from 0 to 2, 0 meaning no anxiety to 2 meaning severe anxiety. Similarly, "depression_class" measured the severity of depression ranging from 0 to 2, 0 meaning no depression to 2 meaning severe depression as seen in Fig. 4.1.

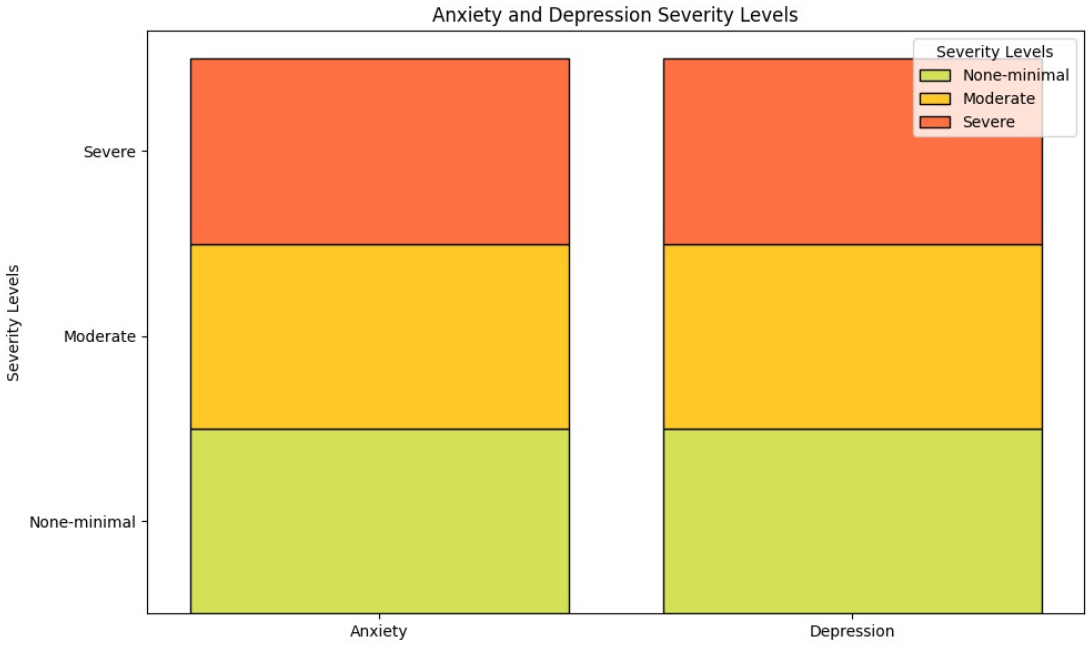


Figure 4.1: Score-Based Distribution of Anxiety and Depression Severity Levels

4.2.3 DL Models

The assessment of psychological instruments that will revolutionize the capacity to predict anxiety and hopelessness will heavily rely on DL. DL can then analyze complexities in data to predict with better accuracy and profile more personalized insights. DL can aggregate multi-sourced data to offer a more holistic picture of mental health.

In this chapter, five DL models—NN, CNN, RNN, LSTM networks, and TabNet were employed for the prediction of depression and anxiety. These models were discussed in detail in 2.8.2. Table 4.2 summarizes a succinct analysis of the benefits, drawbacks linking various DL techniques and Table 4.3 provides details of configuration and hyperparameters of the DL

models.

Table 4.2: Advantages and disadvantages of the models used to forecast anxiety and sadness

Model	Strength	Weakness
NN	Can learn complex relationships	Can be prone to overfitting
CNN	Can automatically learn features	May require specialized hardware for training
RNN	Handles sequential data well	Struggles with long-term dependencies
LSTM	Effective for time series forecasting	Can be slower to train than simpler RNNs
TabNet	Can handle high-dimensional data	May require hyperparameter tuning for optimal performance

Table 4.3: Configuration and hyperparameters of the deep learning models

Model	**Architecture / Units**	**Activation**	**Optimizer**	**Batch / Epochs**	**Regularization / Early Stopping**
NN	Dense: $9 \to 5 \to 3$	ReLU, ReLU, Softmax	Adam	Default / 5	None / No
CNN	Conv1D: 32, $k = 3 \to$ Dense: $6 \to 3$	ReLU, ReLU, Softmax	Adam	Default / 3	None / No
RNN	SimpleRNN: $9 \to$ Dense: $6 \to 3$	ReLU, ReLU, Softmax	Adam	Default / 5	None / No
LSTM	LSTM: $9 \to$ Dense: $2 \to 3$	ReLU, ReLU, Softmax	Adam	Default / 3	None / No
TabNet	Default configuration	Default	Default	20 / 10	Default / Patience = 10

4.2.4 Evaluation metrics

The accuracy with which a model can complete a task determines its performance. Each DL model's performance during training and evaluation on the three datasets was evaluated separately using accuracy, F1-score, precision, and recall. Moreover, a thorough analysis of the

outcomes is conducted using the AUC-ROC. Common classification tasks focus on precision and recall measures. These metrics were defined in detail in 2.10.

4.3 Results

The results for the five DL on the three different datasets bring out very important observations on their performance and the nature of the datasets, as shown in tables 4.4, 4.5, 4.6, 4.7, 4.8 and 4.9.

4.3.1 Analysis of DL models on Dataset 1

Results for dataset 1 in DL models were not very variable in the prediction of anxiety and depression, as shown in Fig. 4.2 The accuracy was highest for LSTM and RNN models, with accuracy more than 0.9 in the case of both conditions being taken into consideration: anxiety and depression. LSTM: anxiety - 0.91, depression - 0.89; RNN: anxiety - 0.91 and depression - 0.88, as shown in Tables 4.4 and 4.5. The CNN also exhibited well, having accuracy ratings between 0.79 and 0.81 for depression and anxiety, respectively. The accuracies for the NN and TabNet models were average, with better performance for depression prediction. While both NN & TabNet showed high precision for depression, their recall of anxiety prediction was poor. Figs. 4.3 and 4.4 display the confusion matrix that shows how well the DL models performed on Dataset 1 for the categorisation of anxiety and depression, respectively.

Table 4.4: Analysis of DL models on dataset 1 for anxiety

Dataset 1 Anxiety	**Accuracy**	**Precision**	**Recall**	**F1-score**
NN	0.74	0.85	0.59	0.64
CNN	0.81	0.88	0.71	0.75
RNN	0.91	0.93	0.88	0.89
LSTM	0.91	0.93	0.88	0.89
TabNet	0.84	0.89	0.77	0.8

Table 4.5: Analysis of DL models on dataset 1 for depression

Dataset 1 Depression	Accuracy	Precision	Recall	F1-score
NN	0.76	0.9	0.6	0.65
CNN	0.79	0.82	0.69	0.72
RNN	0.88	0.88	0.87	0.87
LSTM	0.89	0.89	0.92	0.9
TabNet	0.93	0.95	0.89	0.91

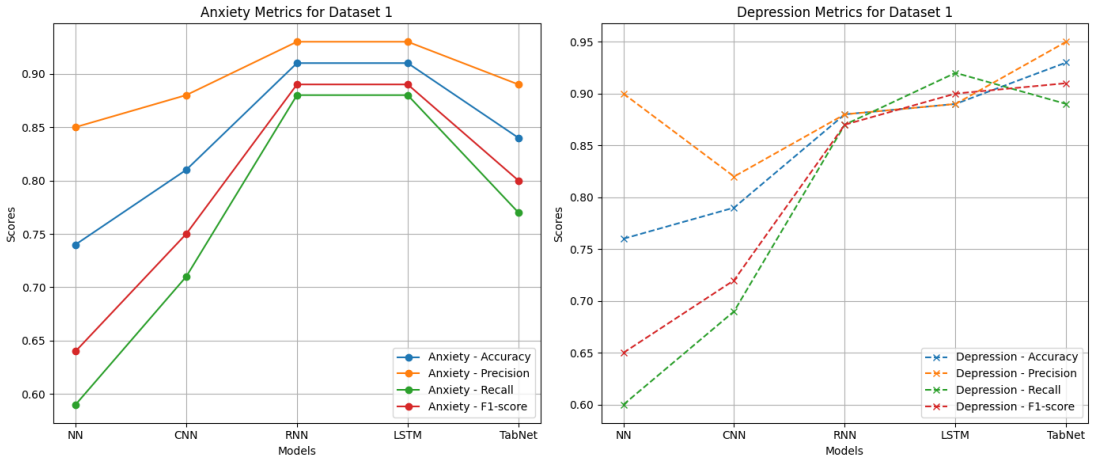


Figure 4.2: Performance of DL models for anxiety and depression on dataset 1.

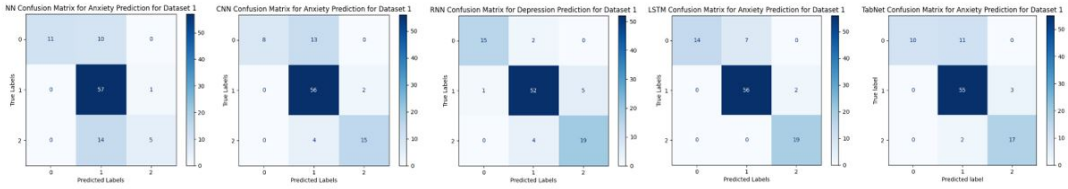


Figure 4.3: Performance of DL models for depression on dataset 1 using the Confusion Matrix.

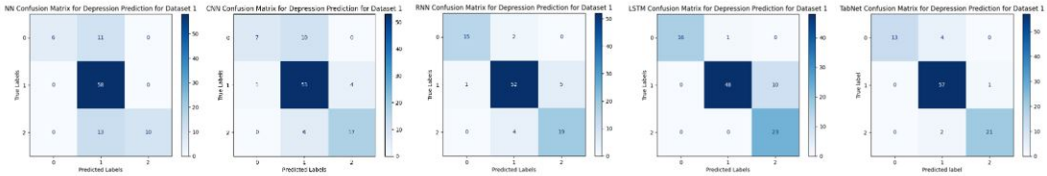


Figure 4.4: Performance of DL models for depression on dataset 1 using the Confusion Matrix.

4.3.2 DL model performance using Dataset 2

The evaluation of DL models in Dataset 2 were very different compared to Dataset 1, as shown in Fig. 4.5. The best-achieved accuracy of the CNN was about 0.9 for anxiety and 0.82 for depression, as shown in Tables 4.6 and 4.7. The TabNet model also demonstrated good performance, specifically for anxiety prediction: accuracy equal to 0.9, precision 0.87. The RNN and LSTM models generated lower accuracy for both conditions. While the LSTM model gave a moderate accuracy for depression, the poorest performance of all measures was that of the RNN model. Compared to Dataset 1, the NN also showed reduced accuracy in Dataset 2, with an accuracy of 0.69 for anxiety but 0.9 for depression. Figs. 4.6 and 4.7 present the confusion matrix illustrating the performance of the DL models for anxiety and depression classification, respectively, on Dataset 2.

Table 4.6: Analysis of DL models on dataset 2 for anxiety

Dataset 2 Anxiety	**Accuracy**	**Precision**	**Recall**	**F1-score**
NN	0.69	0.59	0.6	0.59
CNN	0.9	0.93	0.78	0.82
RNN	0.59	0.39	0.44	0.41
LSTM	0.85	0.58	0.63	0.6
TabNet	0.9	0.87	0.92	0.89

Table 4.7: Analysis of DL models on dataset 2 for depression

Dataset 2 Depression	**Accuracy**	**Precision**	**Recall**	**F1-score**
NN	0.9	0.6	0.65	0.76
CNN	0.82	0.69	0.72	0.79
RNN	0.88	0.87	0.87	0.88
LSTM	0.89	0.92	0.9	0.89
TabNet	0.78	0.77	0.81	0.77

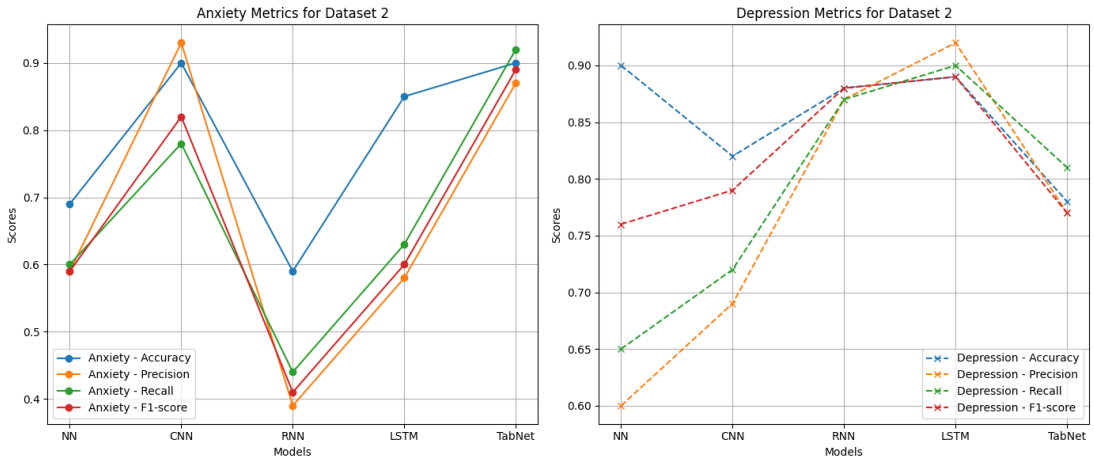


Figure 4.5: Performance of DL models for anxiety and depression on dataset 2.

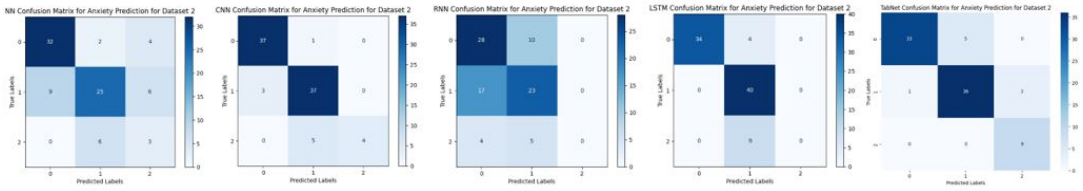


Figure 4.6: Performance of DL models for depression on dataset 2 using the Confusion Matrix.

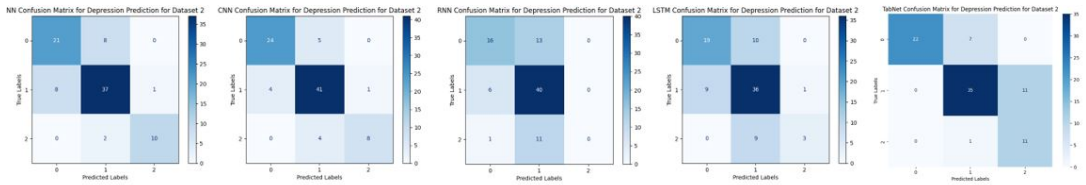


Figure 4.7: Performance of DL models for depression on dataset 2 using the Confusion Matrix.

4.3.3 Performance analysis of DL models on Dataset 3

On Dataset 3, many DL models were pretty accurate in their prediction of anxiety and depression. More specifically, RNN and LSTM did very well with an accuracy above 0.95 in both conditions: RNN: anxiety/depression—0.95; LSTM: anxiety/depression— 0.95, as Tables 4.8 and 4.9 show. The CNN showed excellent performance, with accuracy above 0.93 in anxiety and major depression. Another interesting observation is that the accuracy was high in the case of models based on neural networks and TabNet for depression, but relatively lower in the case of anxiety prediction as shown in Fig. 4.8. Figs. 4.9 and 4.10 present the confusion matrix illustrating

the performance of the DL models for anxiety and depression classification, respectively, on Dataset 3.

Table 4.8: Analysis of DL models on dataset 3 for anxiety

Dataset 3 Anxiety	Accuracy	Precision	Recall	F1-score
NN	0.93	0.6	0.62	0.61
CNN	0.95	0.62	0.64	0.63
RNN	0.95	0.62	0.64	0.63
LSTM	0.95	0.62	0.65	0.63
TabNet	0.93	0.61	0.63	0.62

Table 4.9: Analysis of DL models on dataset 3 for depression

Dataset 3 Depression	Accuracy	Precision	Recall	F1-score
NN	0.94	0.93	0.86	0.88
CNN	0.93	0.93	0.76	0.79
RNN	0.96	0.92	0.92	0.92
LSTM	0.95	0.94	0.94	0.94
TabNet	0.93	0.88	0.89	0.88

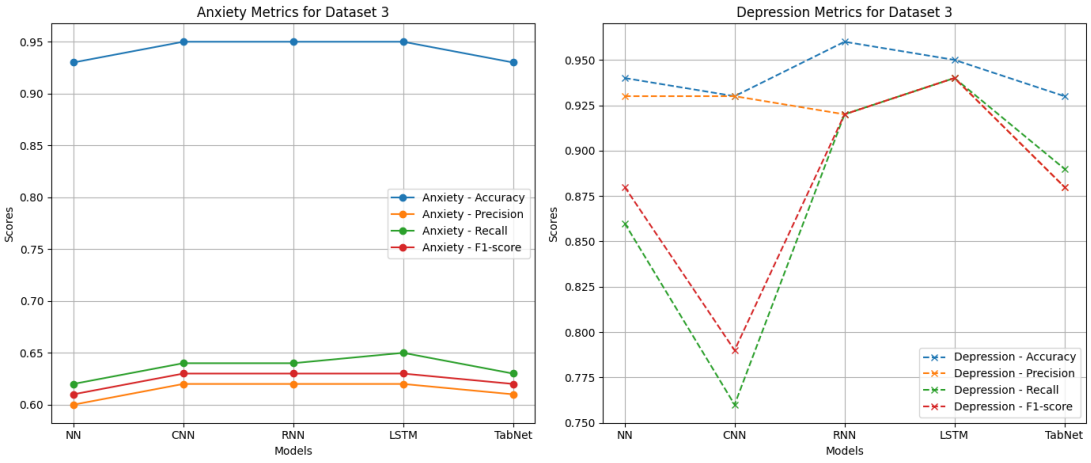


Figure 4.8: Performance of DL models for anxiety and depression on dataset 3.

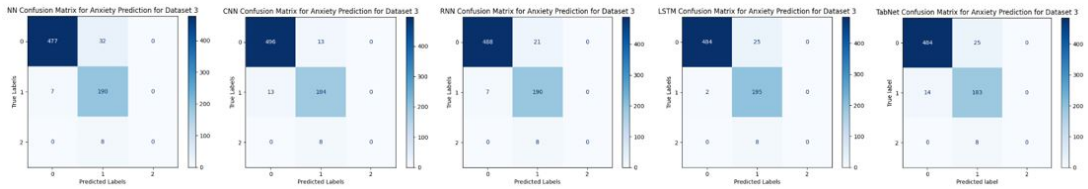


Figure 4.9: Performance of DL models for depression on dataset 3 using Confusion Matrix.

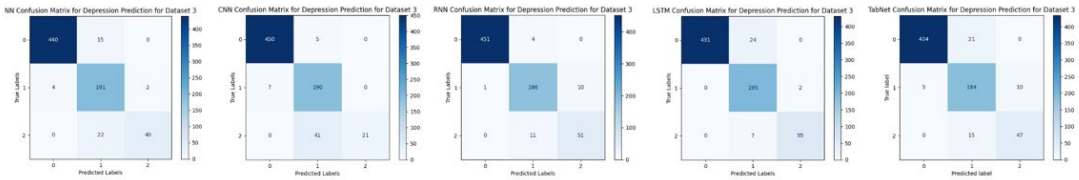


Figure 4.10: Performance of DL models for depression on dataset 3 using Confusion Matrix.

CNN and LSTM had outstanding results for anxiety, where in all datasets, the accuracy was high. Some variability was depicted by RNN; it showed a higher accuracy of 95% in Dataset 3 but lower in Dataset 2. TabNet and CNN performed well for depression prediction; their accuracies vary as per the specification in Fig. 4.11.

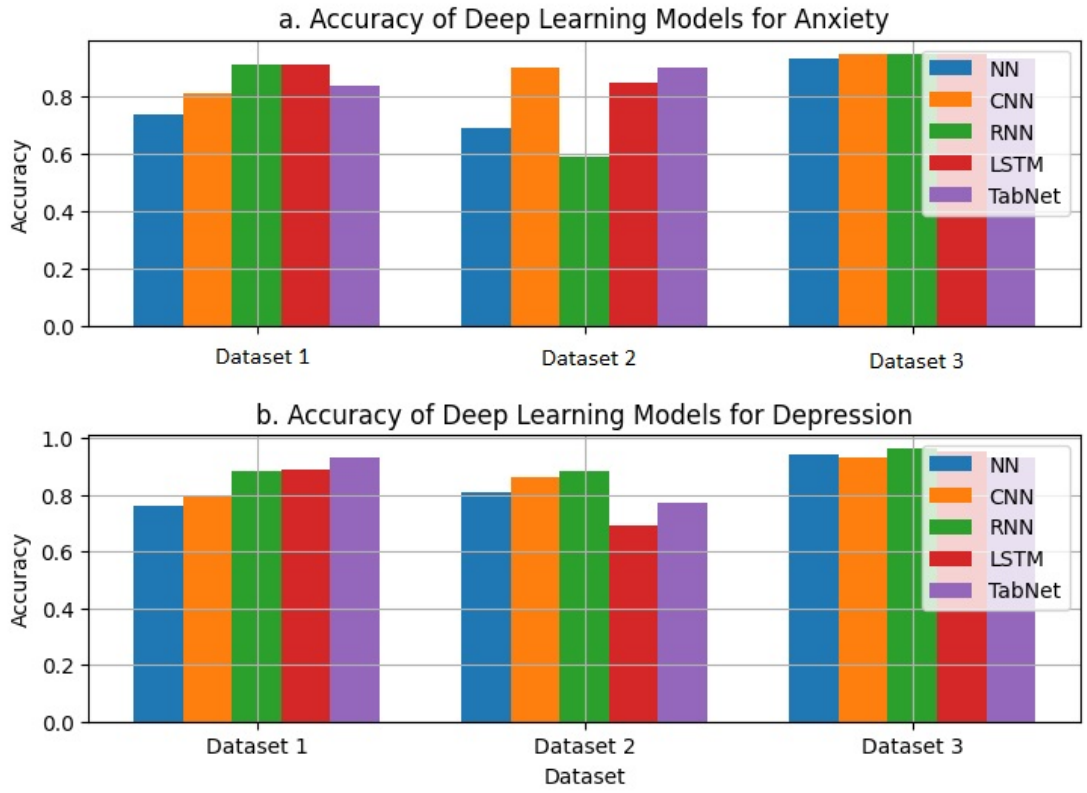


Figure 4.11: a. Evaluation of the accuracy of many DL models for anxiety in comparison. b. Evaluation of the accuracy of many DL models for depression in comparison.

4.4 Discussion

Performance metrics for CNN, LSTM, NN, RNN, and TabNet models on three datasets; additional validation on the AUC-Roc curve is shown. Generally, CNN does well in all

cases, LSTM and RNN do well but only in random instances, while NN uniformly does well but is lower in performance. As mentioned earlier, very high performance was seen across all three datasets, meaning that LSTM and RNN can go up to almost perfect scores as shown in Fig. 4.12. This work also found that CNN and TabNet were powerful in discriminating the prediction of depression across diverse datasets, as shown in Fig.4.13. LSTM and TabNet revealed very competitive performance, especially in Classes 1 and 2. NN and RNN had variable performance across classes. On Dataset 2, TabNet showed very strong performance, while CNN and LSTM were only moderate in performance. On Dataset 3, CNN, and TabNet did not drop performance for Class 1 and Class 2, while RNN continued to output variable yet with considerably high accuracy.

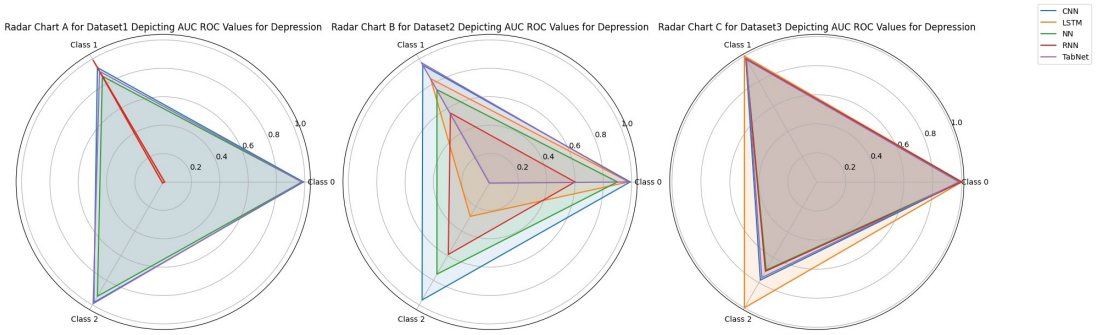


Figure 4.12: Evaluating the ability to predict depression using ROC Curves for Anxiety across all datasets using Radar Plot

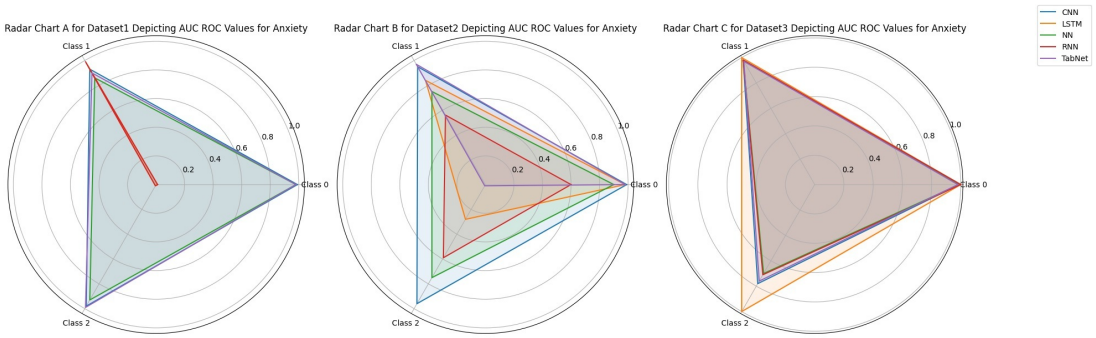


Figure 4.13: Evaluating the ability to predict depression using ROC Curves for Anxiety across all datasets using Radar Plot

4.5 Summary

Using data from standardised psychological questionnaires, this chapter provided a thorough analysis of DL architectures for the assessment of anxiety and depression. Five DL

were systematically analyzed across three independent datasets to examine their predictive capabilities and robustness.

The experimental results demonstrated that DL models are capable of achieving high predictive performance in mental health assessment tasks. CNNs demonstrated robust and consistent performance on all datasets, demonstrating their efficacy in identifying discriminative patterns in questionnaire-based data. Recurrent models, including LSTM and RNN, achieved near-perfect performance in several instances; however, their results showed variability across datasets and class distributions. TabNet emerged as a particularly competitive model, especially in its ability to discriminate between different severity classes of depression, indicating its suitability for tabular mental health data. In contrast, traditional neural networks displayed relatively consistent but comparatively lower performance, emphasizing the advantages of more advanced architectures.

The AUC–ROC analysis further validated the discriminative power of the DL models, confirming their effectiveness in distinguishing between multiple severity classes of anxiety and depression. Performance variations across datasets underscored the influence of dataset-specific characteristics on model behavior, reinforcing the importance of selecting and tuning models based on data properties rather than relying on a single architecture.

Overall, the findings of this chapter demonstrate that DL approaches offer substantial advantages over conventional ML methods for mental health assessment, particularly in handling complex and high-dimensional questionnaire data. However, the observed diversity between models emphasises the need for more investigation into interpretability-driven techniques, ensemble strategies, and hybrid architectures. These insights provide a strong foundation for the subsequent chapter, which focuses on advanced and hybrid intelligence-based approaches aimed at improving robustness, transparency, and clinical applicability of automated mental health assessment systems.

Chapter 5

Enhancing Mental Health Diagnosis with DA-TabSVM: A Multi-Class Hybrid Approach for Detecting Depression and Anxiety

5.1 Introduction

The need for precise, scalable, and data-driven mental health assessment tools has been highlighted by the rising incidence of anxiety and depression, especially in the wake of COVID-19. While earlier chapters demonstrated the effectiveness of traditional ML models and DL architectures for mental health prediction, each approach exhibits inherent limitations when applied independently. Conventional ML models often struggle to capture complex, nonlinear feature interactions, whereas DL models, despite their expressive power, may suffer from reduced interpretability and sensitivity to dataset characteristics.

This chapter focuses on hybrid intelligence-based models that use the complementary advantages of traditional ML and DL techniques to address these issues. Hybrid models offer a balanced framework by combining the interpretability and robustness of traditional algorithms with the representational capacity of DL architectures. Table 5.1 lists the comparison of hybrid models with traditional ML and DL models. Such integration is particularly valuable for mental health assessment, where questionnaire-based data are structured, high-dimensional, and often exhibit subtle interdependencies across features.

In this chapter, improved hybrid classification models, including DT+KNN and RF+NN, are evaluated for the prediction of depression and anxiety across multiple datasets. Furthermore, a novel hybrid architecture, termed DA-TabSVM, is introduced. This model integrates TabNet's capability for handling tabular data with the strong generalization and decision boundary optimization of SVMs, enabling effective multiclass classification of mental health conditions.

The proposed models are evaluated using standardized psychological instruments, namely the GAD-7, PHQ-9, and SCL-90, across three independent datasets. The objective of this chapter is to assess whether hybrid modeling strategies can achieve superior predictive accuracy, robustness, and generalizability compared to standalone ML and DL approaches. The findings presented in this chapter contribute to advancing hybrid intelligence frameworks for reliable and clinically relevant mental health assessment.

Table 5.1: Hybrid Model Comparison with Conventional ML and DL Models

Aspect	**Hybrid Models**	**Traditional ML Models**	**DL Models**	**Advantages of Hybrid ML Models**
Model Structure	Integrates the advantages of many ML/DL methods.	Uses a single ML algorithm (e.g., SVM, KNN).	Relies on deep neural networks (e.g., CNNs, RNNs).	Hybrid models leverage complementary strengths, improving overall performance.
Performance	Often achieves higher accuracy and robustness.	May struggle with complex patterns or data variations.	Can achieve high accuracy but requires large datasets.	Hybrid models can outperform single models by integrating diverse learning methods.
Flexibility	Adaptable to different types of data and tasks.	Limited by the characteristics of the chosen algorithm.	High adaptability but computationally intensive.	Provides flexibility by combining algorithms suited for specific data characteristics.
Interpretability	Better interpretability by combining interpretable models.	Generally interpretable depending on the algorithm.	Often lacks interpretability due to complex architectures.	Balances performance and transparency by integrating interpretable ML with high-performing DL models.
Computational Cost	Balanced cost due to selective model combination.	Generally low to moderate, depending on complexity.	High computational cost, especially for large models.	Offers a balance between performance and efficiency, unlike resource-intensive DL models.
Data Requirements	Can optimize performance even with smaller datasets.	Performance degrades significantly with limited data.	Requires large datasets for optimal performance.	Reduces data constraints by utilising the advantages of both ML and DL techniques.
Overfitting Control	Better control of overfitting through model blending.	Susceptible to overfitting if not tuned properly.	High risk of overfitting without sufficient data or regularization.	Combines models with different biases, reducing overfitting risk.

The studies summarized in Table 2.12 indicate that hybrid models have been explored

using diverse data sources and combinations of ML and DL techniques. However, the reported results vary considerably across datasets, data sources, model combinations, and experimental settings. For example, the reported accuracy ranges from 68% for the hybrid CNN-SVM approach [202] to 99.4% for the TF-IDF and Logistic Regression approach [115]. Furthermore, most existing studies rely on social media, textual, audio, or multimodal data, whereas comparatively fewer studies have investigated hybrid approaches using standardized psychological assessment instruments. The limited use of psychological assessment data and the variation in model combinations indicate a need for further investigation of hybrid approaches for structured depression and anxiety assessment. The chapter is, therefore, meant to give a novel model that raises the performance of the hybrid models. It integrates new approaches and techniques so that it will raise the performance metrics of the already available hybrid models and present an improved model for predicting mental health conditions.

5.2 Proposed Methodology

In detailing the study design, data collection techniques, and analysis procedures to present and interpret findings, the methodology provides an organised strategy and tools used to realise the goals of this work.

5.2.1 Dataset Used

The first dataset was gathered by Islam et al., and the findings highlight the psychological toll of the epidemic by showing a notable prevalence of elevated anxiety and sadness [100]. Out of 476 students, 15% reported moderately severe depression and 18.1% severe anxiety, revealing high levels of depression and anxiety among the participants. Qasrawi et al. gathered the second dataset used in this study, which assessed maternal anxiety and depression in 3569 pregnant and postpartum women [188]. While 22.4% and 7.7% of people, respectively, reported having moderate to high levels of anxiety, 11.6% and 8.7% of people, respectively, reported having moderate to high levels of depression. The third dataset was gathered by [255] in order to use the SCL-90 scale to forecast medical personnel's mental health, as explained in detail in section 2.5. Table 5.2 discusses the datasets used in this work. This work proposed DA-TabSVM, a new hybrid model, which integrates TabNet and SVM to further uplift the performance quality in mental health assessment for the accurate detection of anxiety and depression from complex datasets. Anonymized illustrative snapshots of the first two datasets, are presented in Chapter 3

(Table 3.3 and Table 3.4, respectively). These snapshots illustrate the representative features, questionnaire responses, derived scores, and class labels used for supervised learning. Since the dataset 3 is newly introduced in this chapter, an anonymized illustrative snapshot of this dataset is presented in Table 5.3.

Table 5.2: Description of datasets used

S.No.	Reference	Dataset Size	Hybrid Models used?
1	[100]	N=476	×
2	[188]	N=3569	×
3	[255]	N=5108	✓

Table 5.3: Anonymized Illustrative Snapshot of the Dataset 3

Facility LEVEL_DEPRESS	Gender	Nationality	Profession	Work Area	Years Exp.	S1	S2	S3	S4	S5	S6	S7	ANXIETY	LEVEL_ANXIETY	DEPRESSION
2 2	2	2	1	1	1	3	3	3	1	2	2	2	16	3	53.01
2 2	2	2	2	1	1	0	0	0	1	1	1	0	3	1	52
2 2	1	2	2	1	1	3	3	3	3	3	3	2	20	3	54
2 2	2	3	2	1	1	3	3	3	3	3	3	2	20	3	52
2 2	1	2	2	1	1	1	1	2	1	2	1	3	11.02	3	54
2 1	2	1	2	3	1	2	1	1	1	1	0	0	6	2	37

5.2.2 Mental Health Assessment

Anxiety was measured using the GAD-7 scale, while depression was identified using the PHQ-9 scale as explained in detail in section 2.5. The SCL-90 scale was used by dataset 3.

5.2.3 ML Algorithms

In this chapter, classical supervised ML models, namely, DT, SVM, KNN, and RF, were employed as part of the proposed hybrid framework for depression and anxiety prediction. The theoretical foundations, working principles, advantages, and limitations of these models have already been discussed in detail in Section 2.8.1.3. Here, these algorithms are utilized to establish baseline learning behavior and to capture structured patterns from psychometric survey data within the hybrid approach.

5.2.4 DL Models

To enhance representational learning and capture complex nonlinear relationships, DL models, including NN, CNN, RNN, LSTM, and TabNet, are integrated into the hybrid framework. The detailed architectures and learning mechanisms of these models were presented in Section 2.8.2. In the current chapter, these DL models are applied to learn higher-level abstractions from the data and to complement traditional ML models by improving robustness, generalization, and predictive performance.

5.2.5 Hybrid Models

In order to enhance prediction performance and generalisation capacity, a hybrid model will combine the capabilities of several machine learning models. The various model combinations in a hybrid method aim to capitalise on the distinct advantages of each component model while effectively eliminating their weak links. By effectively eliminating each component model's weak points, the various model combinations in a hybrid method aim to capitalise on its own advantages.

5.2.5.1 DT+KNN

To enhance classification performance, Yadav and Bokhari (2024) employ a hybrid model comprising DT and KNN, [262]. To find and model the patterns and connections between features and the target variable, first train a DT classifier. The use of these predictions from the DT as features for the next KNN model is based on both the training and test data. The output of the predictions from the DT is reshaped and fed for training in a KNN Classifier for the final prediction. The KNN model will operate based on the specified number of neighbours and leaf size for determining its classification decision. The accuracy of the final forecasts and the creation of a classification report will thus serve as a gauge of this hybrid approach's efficacy. This hybrid method uses the merits of both models: handling complex boundaries in decisions by the DT and the local information by KNN to improve prediction. This approach should provide better accuracy of classification by using the different advantages of each model, and hence is more likely to provide a robust solution to the problem of classification being undertaken. The hybrid function is:

$$f(x) = kNN(DT(x)) \tag{5.1}$$

where $DT(x)$ is the output of the DT model used as input features for KNN.

Table 5.4: Hyperparameters for DT+kNN

Model	**Hyperparameters**
DT	`max_depth`: 10
kNN	`n_neighbors`: 5, `leaf_size`: 1

5.2.5.2 RF+NN

In order to enhance classification problem performance, a hybrid approach is created in this study by combining RF and NN models provided by Yadav & Bokhari, 2024 [262]. An RF classifier will first identify intricate linkages and patterns in the dataset. A NN with dense layers and ReLU activations that enables the model to learn from the output of these probability estimates in refining classification tasks will also be created. The RF's output yields class probability estimates that are used as input features for the neural network. Additionally, a NN with thick layers and ReLU activations will be created so that the model can improve classification tasks by learning from the output of these probability estimates. The Adam optimiser and sparse categorical cross-entropy as the loss function will be used to create the model. The model will be trained over 20 epochs to optimise its parameters on the output features provided by the RF. It is appropriate for multi-class classification tasks. Lastly, the hybrid model's performance is evaluated by forecasting the test set's class probabilities, which are subsequently transformed into class labels using the argmax function. The classification report provides detailed information on how accurate and performant the model is, showing how the combination of RF and NN methodologies can leverage their individually unique strengths in leading to a more robust and accurate classification outcome. The hybrid function is:

$$f(x) = NN(RF(x)) \tag{5.2}$$

where $RF(x)$ generates probabilities fed into a Neural Network for final classification.

Table 5.5: Hyperparameters for RF+NN

Model	**Hyperparameters**
RF	`random_state`: 5
NN	Dense Layers: [64, 32, 3] Activation Functions: [ReLU, Softmax] Optimizer: Adam Loss Function: Sparse Categorical Crossentropy Epochs: 20

5.2.5.3 DA-TabSVM

The approach is based on creating a hybrid model from the predictions of two different ML techniques: TabNet and SVM [265]. First of all, train TabNet Classifier on the training data; due to the attention mechanisms applied, it works well on complex and structured data. Then, generate probabilistic predictions. It is trained together with an SVM model, which has high performance in high-dimensional spaces and gives its own set of probabilistic predictions. The combination of these predictions concatenates to a new feature space, merging with the strengths of these models. A Logistic Regression model will then be trained on this combined feature set to make the final classification as shown in Fig. 5.1. By utilising the complementary capabilities of TabNet and SVM, this stacking enables the final estimator to achieve potentially better overall model performance. This hybrid model's efficacy is assessed using the test set. This work shows how the hybrid model can integrate predictions stemming from diverse algorithms in a way that improves classification accuracy and comprehensively assesses the performance through the classification report. The hybrid function is:

$$f(x) = LR([TabNet(x), SVM(x)]) \tag{5.3}$$

where $TabNet(x)$ and $SVM(x)$ are the respective model outputs combined and classified with Logistic Regression.

The two base learners in the DA-TabSVM framework employ distinct optimization objectives. TabNet is trained using a multiclass classification loss, represented by the cross-entropy objective:

$$\mathscr{L}_{\text{TabNet}}(\theta) = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \log p_{ic} \tag{5.4}$$

where N denotes the number of training samples, C denotes the number of classes, y_{ic}

represents the true class indicator, and p_{ic} denotes the probability predicted by TabNet for class c.

The SVM base learner uses a margin-based objective, commonly expressed using the hinge loss:

$$\mathscr{L}_{\text{SVM}} = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\max\left(0, 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right) \tag{5.5}$$

where $\mathbf{w}$ and b represent the SVM parameters, C controls the regularization strength, and the second term penalizes observations that violate the desired classification margin.

The two objectives are optimized independently rather than jointly. TabNet and SVM are trained separately to generate their respective probabilistic predictions. These predictions are subsequently concatenated to form the meta-level feature space, which is provided to the Logistic Regression model for final classification. Thus, the two loss functions are not simultaneously optimized within a common objective, and no direct conflict between the TabNet and SVM objectives arises during the training of the base learners.

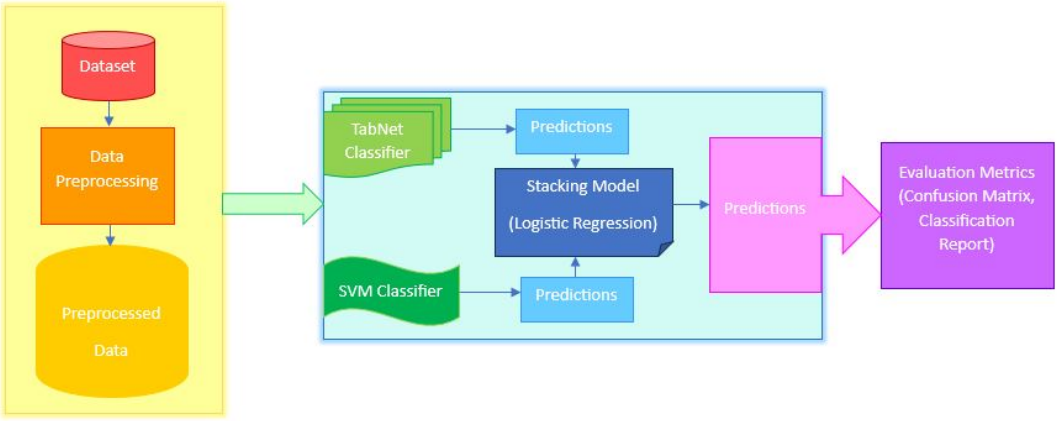


Figure 5.1: Proposed DA-TabSVM architecture.

Table 5.6: Hyperparameters for DA-TabSVM

Model	**Hyperparameters**
SVM	`probability`: True
TabNet	`max_epochs`: 5, `batch_size`: 20, `virtual_batch_size`: 12, `patience`: 10

In the current work, the best hyperparameters for each model are identified through a systematic search using GridSearchCV. This is the best setting of parameters for the DT+kNN

model according to the search: the maximum depth in the DT was set to 10, and KNN was configured with 5 neighbors and a leaf size of 1, as shown in Table 5.4.

For RF+NN, a random state of 5 in the RF was considered for optimization, with a Neural Network architecture that had dense layers of sizes 64, 32, and 3, with ReLU and Softmax as activation functions, respectively, as shown in Table 5.5. Using the Adam optimiser and the Sparse Categorical Cross-entropy loss function across 20 epochs was the ideal configuration. For the DA-TabSVM model, the SVM was set to output probabilities, and TabNet tuned the parameters with a maximum of 5 epochs, batch size of 20, virtual batch size of 12, and set patience at 10 as shown in Table 5.6. The GridSearchCV did the thorough hyperparameter tuning that had to be done to ensure each configuration within this model setting was optimized to the best performance. Table 5.7 gives an algorithm for the working of the proposed hybrid models.

Table 5.7: Algorithm of working of Hybrid Models

Hybrid Models (DT+KNN/RF+NN)	**DA-TabSVM**
Input: Training data: (X, y) Test data: (X_{test}, y_{test})	**Input:** Training data: (X, y) Test data: (X_{test}, y_{test})
Output: Classification accuracy	**Output:** Classification accuracy
Steps: 1. Split Data: Create training and validation sets from the training data. 2. Train ML Model 1: Create ML Model 1 and train on the training set. 3. Generate Features: Predict class labels using ML Model 1; reshape predictions for ML Model 2. 4. Train ML Model 2: Train ML Model 2 on ML Model 1 predictions and original labels. 5. Make Predictions: Predict labels for the test data. 6. Evaluate Performance: Calculate accuracy; optionally, generate classification report.	1. Split Data: Divide training data into training and validation sets. 2. Train TabNet: Create TabNetClassifier (tabnet_model), train with early stopping, evaluate on testing data. 3. Train SVM: Create SVC (svm_model) with probability=True; train on training data. 4. Create Stacking Ensemble: Create StackingClassifier (hybrid_model). 5. Define base estimators: ('tabnet', svm_model) and ('svm', svm_model). 6. Set final estimator: LogisticRegression; stack_method='predict_proba'. 7. Train Hybrid Model: Fit the hybrid model using training data to train it. 8. Evaluate Hybrid Model: Make predictions based on test data, compute accuracy, and produce a categorisation report.

5.2.6 Evaluation metrics

The accuracy with which a model can complete a task determines its performance. Each model's performance during training and evaluation on the three datasets was evaluated separately using accuracy, F1-score, precision, and recall. Moreover, a thorough analysis of the outcomes is conducted using the AUC-ROC. Common classification tasks focus on precision and recall measures. These metrics were explained in detail in Section 2.10.

5.3 Results

In this section, the results of the proposed model are discussed.

5.3.1 Performance analysis of ML and DL models on Depression

Variations in accuracy, precision, recall, and F1-score are found when comparing the performance of certain conventional ML and DL models for depression prediction on three datasets. With accuracies of 0.98, 0.98, and 0.99 for Datasets 1, 2, and 3, respectively, the SVM outperformed conventional models, as Table 5.8. While doing this, SVM also maintained high precision and recall, hence making it one of the top-ranked traditional models. RF followed closely with accuracies within a range of 0.97 to 0.98, hence proving its strength in many classification tasks. Similarly, KNN outdid others with the accuracies of 0.97, 0.96, and 0.97 on these datasets, though with somewhat variable recalls to fully capture all the relevant instances. NN showed good performances, especially on Datasets 2 and 3 with 0.97 and 0.99 accuracies, respectively, although that of Dataset 1 is relatively lower, which stands at an accuracy of 0.90. In contrast, DT and TabNet models are weaker, presenting a low accuracy with several oscillations among the datasets as shown in Fig. 5.2.

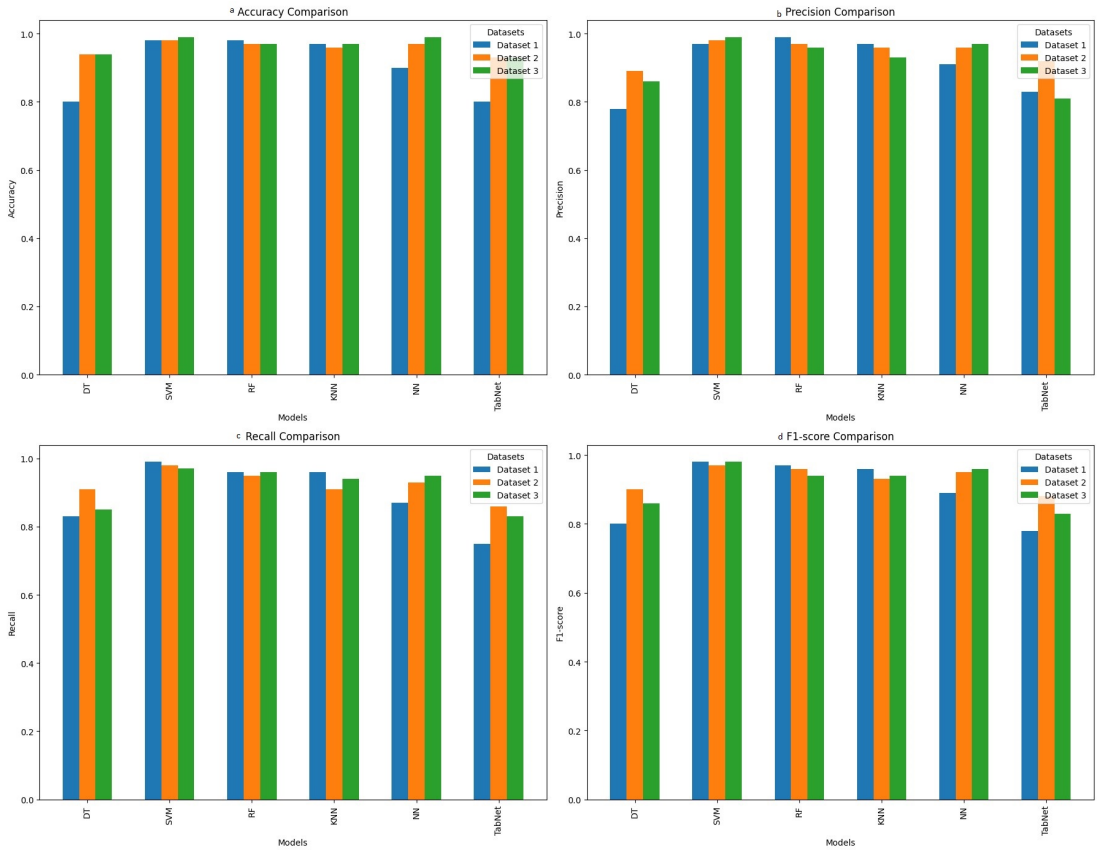


Figure 5.2: ML model's performance in detecting depression (a) Accuracy; (b) Precision; (c) Recall; (d) F1-score

By combining the best aspects of several available algorithms, hybrid models fared better than others. As demonstrated in Fig. 5.3, the RF+NN hybrid model outperformed all other models, producing accuracy values of 0.95, 0.97, and 0.98 on the three datasets, respectively. This is because it combines RF's robustness in handling features with the capability of NN for modeling complex relationships. Precision and recall levels rise as a result, achieving a favourable balance between the two. Computational DT+kNN hybrid ensured that the accuracies for DT and kNN stood at 0.88, 0.94, and 0.95, respectively, which attested to an effective synergy between the simplicity and interpretability of DT with the instance-based learning of KNN.

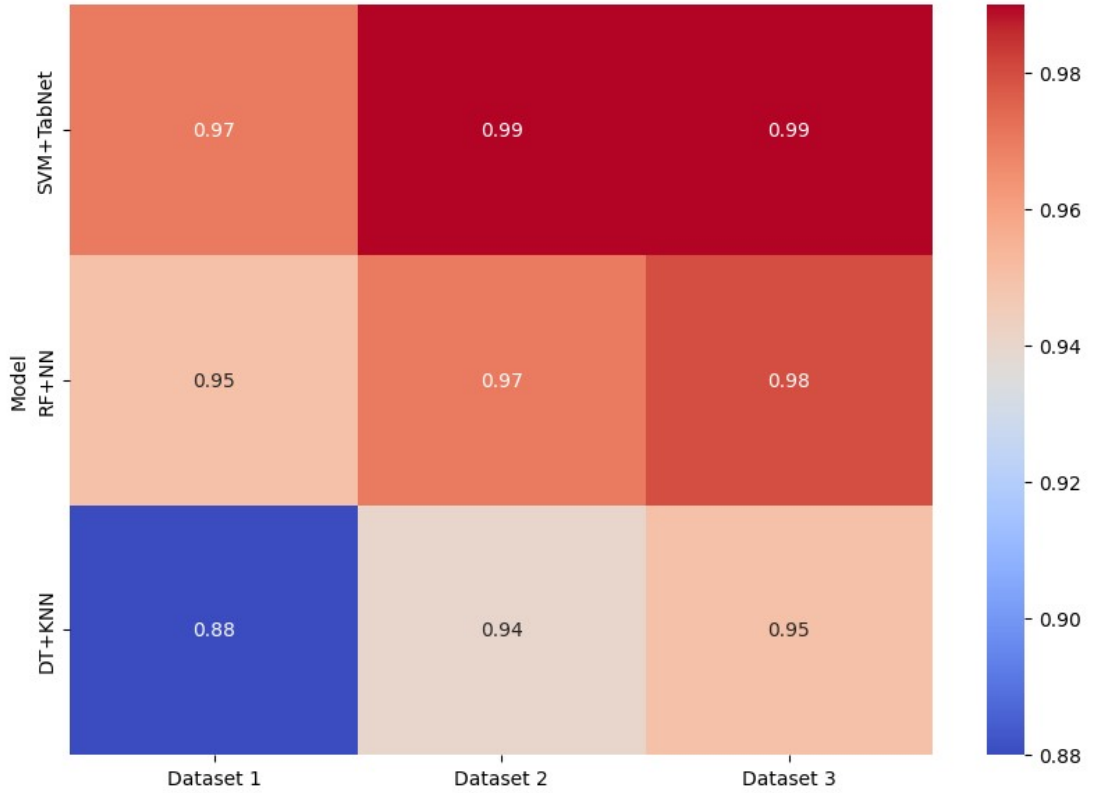


Figure 5.3: Comparison of the performance of hybrid models in identifying depression.

Particularly noteworthy was the proposed DA-TabSVM hybrid, which delivered highly consistent high performance, achieving the highest accuracies of 0.97, 0.99, and 0.99 across all datasets. This hybrid effectively combined strong SVM classification with adaptive feature selection in TabNet, thereby enhancing its depression case identification with high interpretability and precision. The superior recall rates further underlined the effectiveness of the hybrid in the minimum misclassification cardinal for a clinical application where missing true cases have great consequences. Fig. 5.4 displays the confusion matrix that shows how well the DA-TabSVM model performs for depression classification across all datasets.

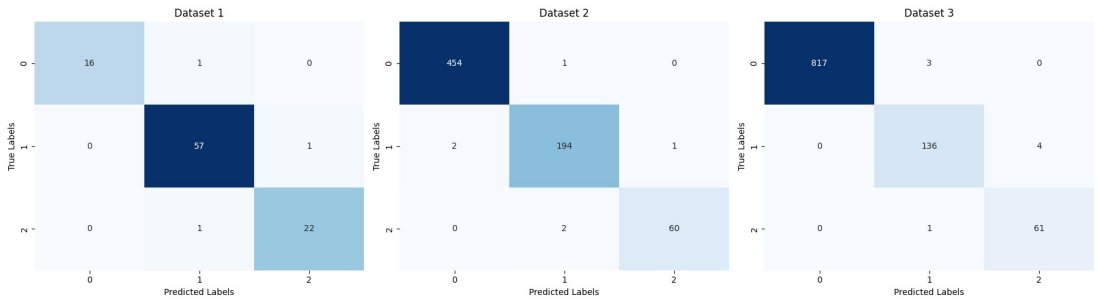


Figure 5.4: Performance of DA-TabSVM for depression using Confusion Matrix.

Table 5.8: Performance analysis of all the models to detect depression.

Model	**Dataset 1**				**Dataset 2**				**Dataset 3**			
	Accuracy	**Precision**	**Recall**	**F1-score**	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Accuracy**	**Precision**	**Recall**	**F1-score**
DT	0.80	0.78	0.83	0.80	0.94	0.89	0.91	0.90	0.94	0.86	0.85	0.86
SVM	0.98	0.97	0.99	0.98	0.98	0.98	0.97	0.97	0.99	0.97	0.99	0.98
RF	0.98	0.99	0.96	0.97	0.97	0.97	0.95	0.96	0.97	0.96	0.92	0.94
KNN	0.97	0.97	0.96	0.96	0.96	0.96	0.91	0.93	0.97	0.96	0.93	0.94
NN	0.90	0.91	0.87	0.89	0.97	0.96	0.93	0.95	0.99	0.97	0.95	0.96
TabNet	0.80	0.83	0.75	0.78	0.93	0.92	0.86	0.88	0.93	0.86	0.81	0.83
DT+kNN	0.88	0.89	0.84	0.87	0.94	0.89	0.88	0.88	0.95	0.90	0.86	0.88
RF+NN	0.95	0.97	0.91	0.94	0.97	0.97	0.96	0.96	0.98	0.96	0.93	0.94
DA-TabSVM	0.97	0.97	0.96	0.97	0.99	0.99	0.98	0.98	0.99	0.97	0.98	0.98

5.3.2 Evaluation of ML and DL models' performance in relation to anxiety

The three datasets provided are used to report the accuracy, precision, recall, and F1 score of classical models in predicting anxiety. It is noted among the traditional models that the SVM showed higher accurate performance, with 0.98, 0.99, and 1.00 for Datasets 1, 2, and 3 as shown in Table 5.9. The SVM exhibited high values of precision and recall in classifying. The RF algorithm also resulted in good results, with accuracy ranging from 0.97 to 0.98 across the datasets, showing that it is also capable of properly handling the anxiety data complexity. KNN has also shown good performance, where accuracy ranged between 0.95 and 0.98, with a recall that fluctuated a bit, thus showing a possible weak point in catching all anxiety cases consistently. NN showed performances in the range of good-performance Dataset 3 with an accuracy of 1.00, but was much less consistent in Dataset 1 and Dataset 2, with accuracies of 0.89 and 0.98, respectively. The DT and especially the TabNet models had more humble results, showing less precision and recall, with the TabNet model revealing poor performance—in particular, in Dataset 2—underlining the difficulty of identifying anxiety with good reliability as shown in Fig. 5.5.

Hybrid models, on their part, however, tend to perform better by combining the strengths of two algorithms. The RF+NN hybrid model returns high results: 0.88, 0.97, and 0.98 across the three datasets as shown in Fig. 5.6. This hybrid leverages effectively the ability of RF in managing diverse feature sets and NN in modeling complex patterns toward enhanced precision and recall compared to individual models. However, it was competitive with accuracies of 0.84, 0.97, and 0.96, respectively. This illustrates the ability to be balanced in interpretability and performance through the combination of the rule-based structure of a DT and instance-based learning from KNN.

Especially, the DA-TabSVM hybrid reached the best accuracies: 0.97, 0.99, and 1.00 in such datasets. This hybrid model combined SVM's strength in being able to classify effectively and TabNet's adaptive feature selection to result in high precision and recall. The DA-TabSVM model consistently performed in all datasets, thus really underlining robust and flexible character when dealing with fine-grained features harboring the essence of anxiety data, hence obtaining higher recall and precision than the other models. Fig. 5.7 presents the confusion matrix illustrating the performance of the DA-TabSVM model for anxiety classification on all datasets.

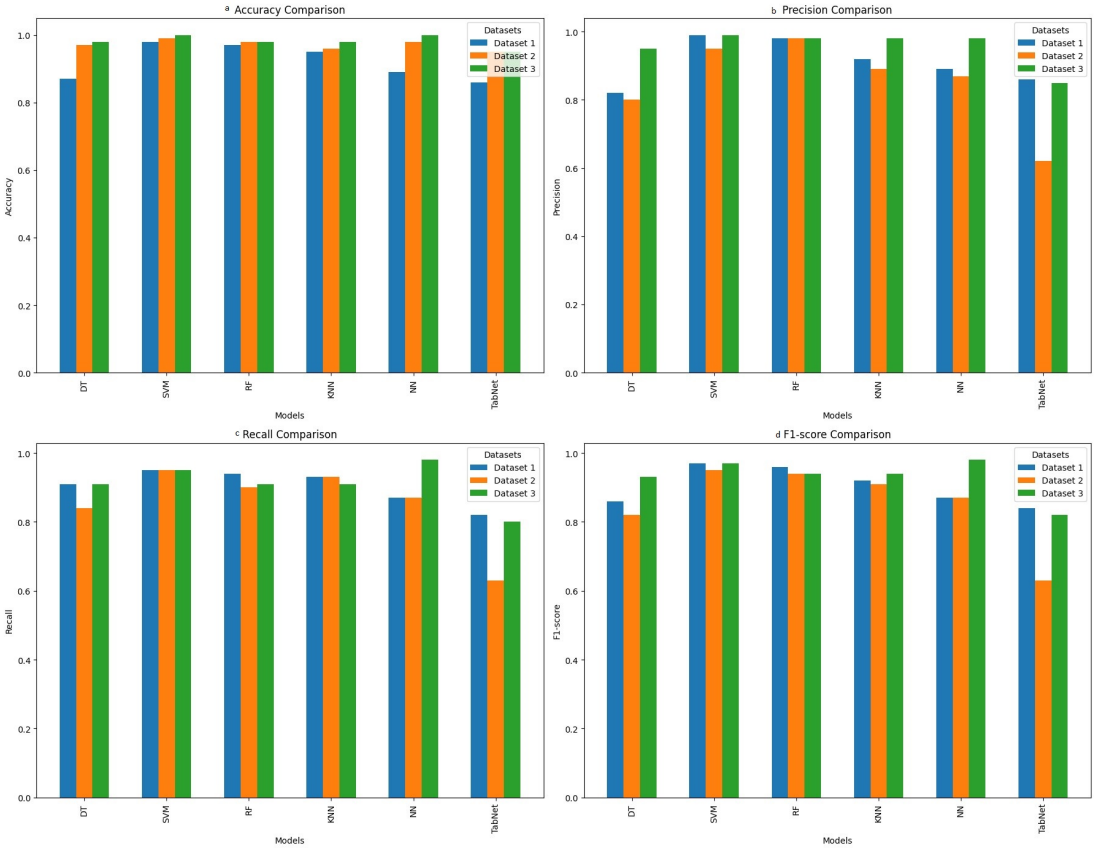


Figure 5.5: Comparison of ML models' performance to detect anxiety (a) Accuracy; (b) Precision; (c) Recall; (d) F1-score.

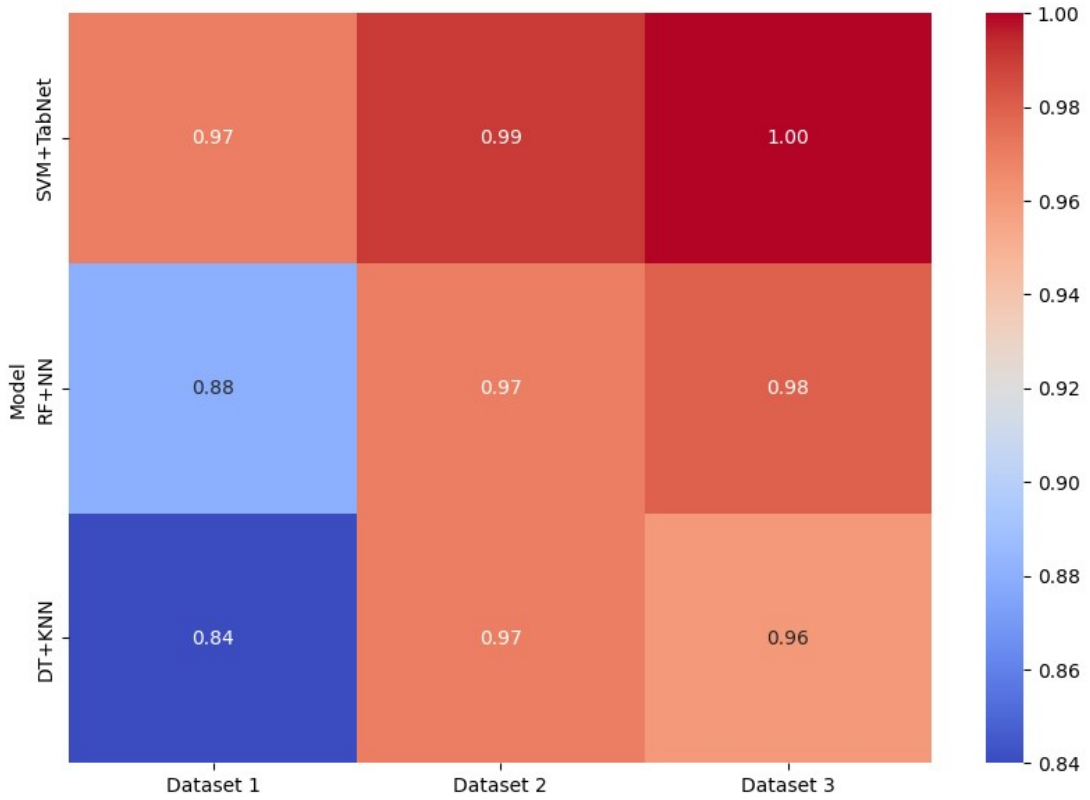


Figure 5.6: Comparison of hybrid models' performance to detect anxiety.

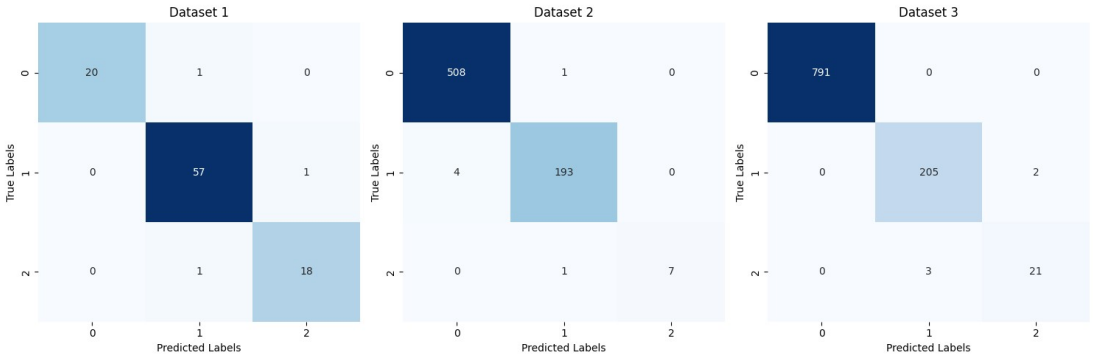


Figure 5.7: Performance of DA-TabSVM for anxiety using Confusion Matrix.

Table 5.9: Performance analysis of all the models to detect anxiety.

Model	**Dataset 1**				**Dataset 2**				**Dataset 3**			
	Accuracy	**Precision**	**Recall**	**F1-score**	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Accuracy**	**Precision**	**Recall**	**F1-score**
DT	0.87	0.82	0.91	0.86	0.97	0.80	0.84	0.82	0.98	0.95	0.91	0.93
SVM	0.98	0.99	0.95	0.97	0.99	0.95	0.95	0.95	1.00	0.99	0.95	0.97
RF	0.97	0.98	0.94	0.96	0.98	0.98	0.90	0.94	0.98	0.98	0.91	0.94
KNN	0.95	0.92	0.93	0.92	0.96	0.89	0.93	0.91	0.98	0.98	0.91	0.94
NN	0.89	0.89	0.87	0.87	0.98	0.87	0.87	0.87	1.00	0.98	0.98	0.98
TabNet	0.86	0.86	0.82	0.84	0.95	0.62	0.63	0.63	0.95	0.85	0.80	0.82
DT + kNN	0.84	0.79	0.82	0.80	0.97	0.93	0.91	0.92	0.96	0.90	0.94	0.92
RF + NN	0.88	0.91	0.83	0.86	0.97	0.98	0.89	0.92	0.98	0.95	0.96	0.96
DA-TabSVM	0.97	0.97	0.96	0.97	0.99	0.99	0.92	0.95	1.00	0.97	0.96	0.96

5.4 Discussion

Overall, the findings show that hybrid models, which combine complimentary strengths to improve accuracy, precision, recall, and F1 scores, typically outperform classic ML and DL models. Hybrids represent a more balanced approach; thus, they are the most effective in complex tasks such as anxiety prediction, where subtle pattern understanding is needed along with reliable classification. The performance of the hybrid models are higher on both datasets, standing for their potential to become a significant, though powerful avenue for the assessment of mental health beyond a model's traditional application.

The suggested hybrid approach consistently achieved the greatest degree of precision, accuracy, recall, and F1-score, outperforming or matching traditional ML and other hybrid models, according to the performance analysis of the DA-TabSVM model in identifying depression and anxiety using three datasets. The performance of DA-TabSVM reached up to 97% in terms of accuracy on Dataset 1, 99% on Dataset 2, and 99% on Dataset 3 for depression detection and similarly high when detecting anxiety with a range reaching up to an accuracy of 100% on Dataset 3, as shown in Fig. 5.8 and Fig. 5.9.

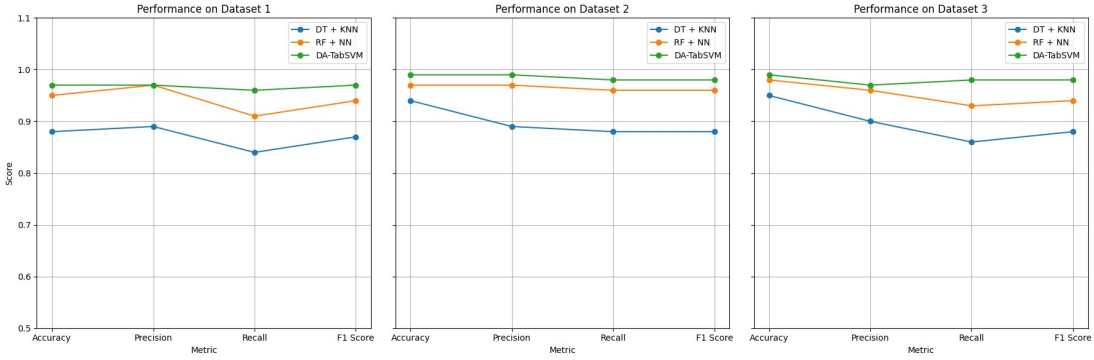


Figure 5.8: Comparison of accuracy of hybrid models' performance to detect depression (a) Dataset 1; (b) Dataset 2; (c) Dataset 3

RF+NN and DT+kNN demonstrated competitive performances, with RF+NN's maximum accuracy reaching 98% on Dataset 3. However, for the most part, these models were less consistent than DA-TabSVM, especially in precision and recall. As the relative comparison goes, the embedding of TabNet with SVM lets DA-TabSVM be uniquely privileged in terms of enhanced feature learning and classifications, hence returning overall improved performances on mental health assessment.

A comparison of the hybrid models, such as DT+KNN and RF+NN, using three datasets, tells a lot about their performances concerning the baseline results by Yadav & Bokhari, 2024, as shown in Table 5.10. First is the fact that on Dataset 1, both DT+KNN and RF+NN had an

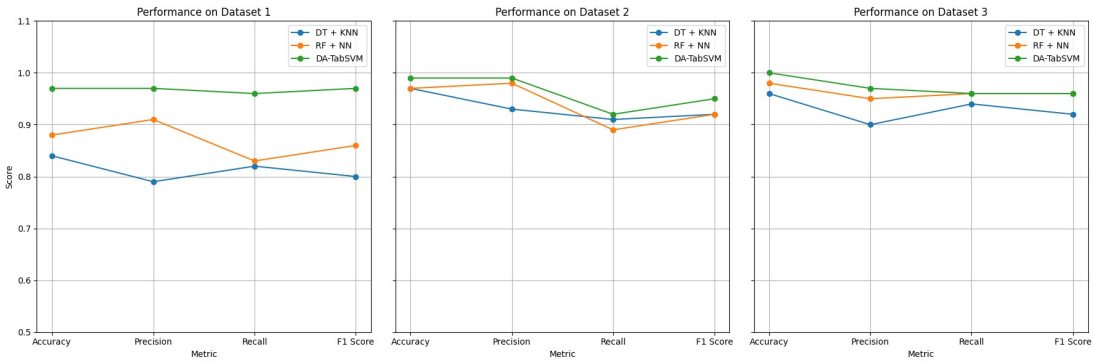


Figure 5.9: Comparison of accuracy of hybrid models' performance to detect anxiety (a) Dataset 1; (b) Dataset 2; (c) Dataset 3

accuracy of 0.88, outperforming 0.8669 as reported by Yadav & Bokhari. That is to say, the models proposed in this work are slightly improved from the baseline in the current dataset. In Dataset 2, RF+NN outperformed DT+KNN, with an accuracy of 0.97 against the 0.94 accuracy of DT+KNN. Both outdid the baseline reported by Yadav & Bokhari with an accuracy of 0.9354, which means there was an improvement with the ensembles implemented here. Dataset 3 reinforces the fact that RF+NN yielded an accuracy of 0.98 while DT + KNN had an accuracy of 0.95. Compared to a baseline accuracy of 0.95 by Yadav & Bokhari, DT + KNN realizes this, while RF+NN presents an appreciable boost for this dataset. In summary, these results point to the high performance of the RF + NN hybrid model, which, particularly, enhances the accuracy across the datasets concerning both DT + KNN and baseline findings as shown in Fig. 5.10.

Three models, DT+KNN, RF+NN, and DA-TabSVM, have shown heterogeneous effectiveness in detecting depression and three classes for three different datasets. AUC-ROC scores showed that the performance of the DT+KNN model was consistent, where all its performances ranged between 0.86 and 0.98, with the highest performance of Class 0 in Dataset 2 being 0.98, as shown in Table 5.11. On the contrary, RF+NN, as well as DA-TabSVM manifested optimal stability in all three classes across the datasets, wherein a perfect score of 1.00 was achieved. The scores for the detection of anxiety were all high, with minor variation across datasets in DT+KNN at 0.89 to 1.00, while RF+NN and DA-TabSVM retained near-perfect scores across all classes and datasets. This suggests that RF+NN and DA-TabSVM perform better in terms of accuracy and robustness in the detection of depression and anxiety across various data settings, even though DT+KNN performs well, as shown in Table 5.12. Fig. 5.11 and Fig. 5.12 depict pictorial representations of AUC-ROC curves for depression and anxiety, respectively.

Table 5.10: Accuracy Comparison of Hybrid Models DT+kNN and RF+NN

Dataset	DT+KNN	RF+NN	Yadav & Bokhari, 2024
Dataset 1	0.88	0.88	0.8669
Dataset 2	0.94	0.97	0.9354
Dataset 3	0.95	0.98	0.95

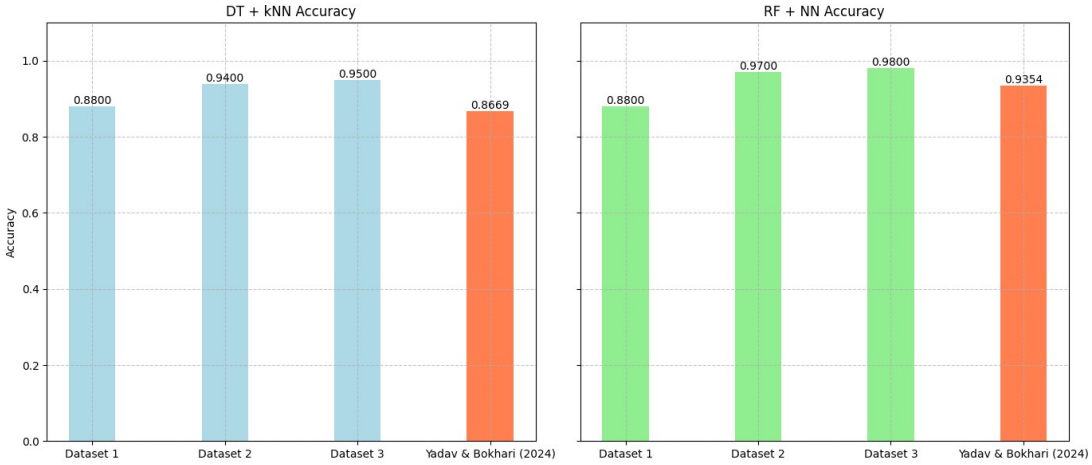


Figure 5.10: Comparison of accuracy across DT+KNN and RF+NN.

Table 5.11: AUC-ROC Comparison of Hybrid Models for Depression

Class	**DT+KNN**			**RF+NN**			**DA-TabSVM**		
	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3
Class 0	0.86	0.98	0.92	1	1	1	1	1	1
Class 1	0.88	0.93	0.88	0.97	1	1	1	1	1
Class 2	0.91	0.87	0.90	0.98	1	1	1	1	1

Table 5.12: AUC-ROC Comparison of Hybrid Models for Anxiety

Class	**DT+KNN**			**RF+NN**			**DA-TabSVM**		
	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3
Class 0	1.00	0.97	0.95	0.99	1	1	0.99	1	1
Class 1	0.98	0.96	0.93	1	1	1	1	1	1
Class 2	0.99	0.89	0.96	1	1	1	1	1	1

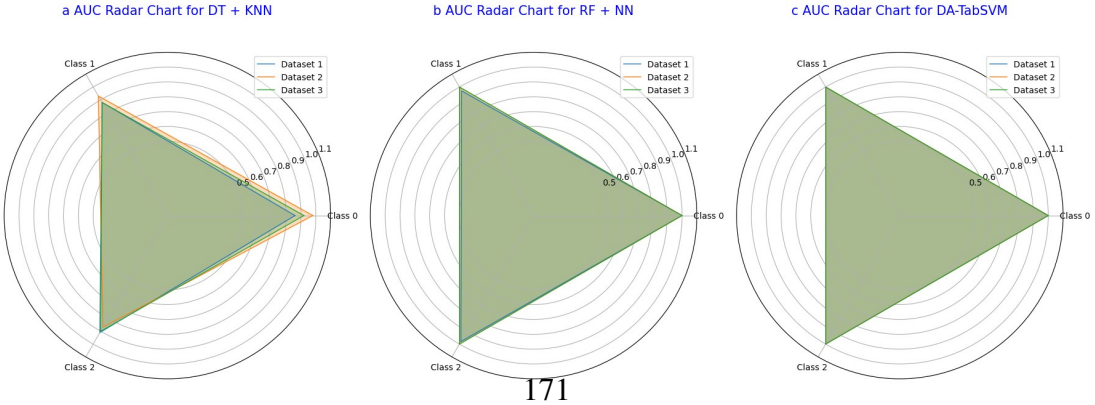


Figure 5.11: Comparison of hybrid models' AUC-ROC scores for depression detection.

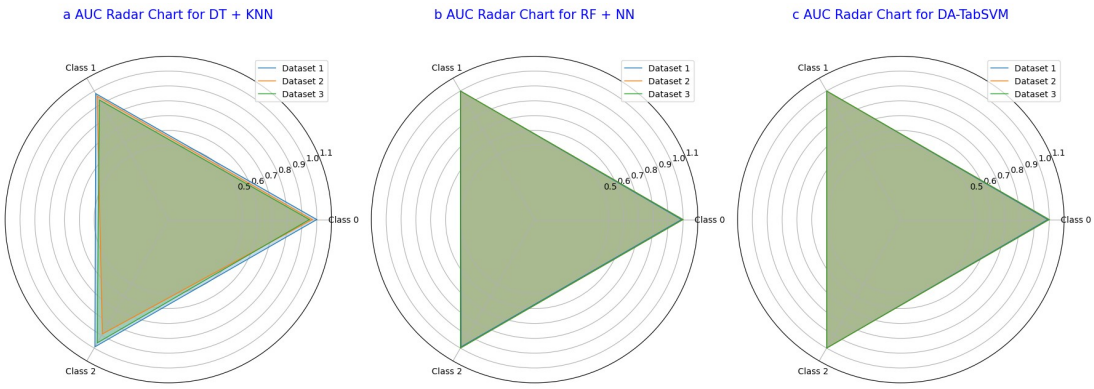


Figure 5.12: Comparison of hybrid models' AUC-ROC scores for anxiety detection.

5.5 Summary

Using data from standardised psychological assessments, this chapter provided a thorough analysis of hybrid intelligence-based models for the automated diagnosis of anxiety and depression.The chapter examined the efficacy of merging complementing algorithms to improve prediction performance and resilience, building on the shortcomings found in single ML and DL techniques.

Three separate datasets from the GAD-7, PHQ-9, and SCL-90 instruments were used to assess improved hybrid models, notably DT+KNN and RF+NN, in addition to a novel hybrid architecture, DA-TabSVM. The experimental findings demonstrated that hybrid approaches consistently achieved superior performance compared to traditional models, with DA-TabSVM exhibiting the most stable and accurate results across datasets and mental health conditions. The integration of TabNet's feature learning capabilities with the classification strength of SVMs enabled effective multiclass prediction while maintaining high recall, accuracy, precision, and F1-scores.

Overall, this chapter demonstrates that hybrid intelligence frameworks provide a strong and trustworthy method for assessing mental health, especially when handling complicated, high-dimensional questionnaire data. The findings demonstrate the value of algorithmic synergy in enhancing robustness and generalisation across a variety of datasets. These findings provide a solid basis for later chapters that delve deeper into sophisticated modelling techniques meant to improve the interpretability, transparency, and practicality of automated mental health diagnostic systems. These observations provide a solid basis for the next chapters that go deeper into sophisticated modelling techniques meant to improve the interpretability, transparency, and practicality of automated mental health diagnostic systems.

Chapter 6

A Novel Adaptive Sigma KNN Model for Depression and Anxiety Detection Following the COVID-19 Pandemic

6.1 Introduction

Building upon the findings of the previous chapter, which demonstrated the effectiveness of hybrid models such as DA-TabSVM for mental health classification, this chapter introduces a novel ASKNN model to address the remaining limitations related to neighborhood sensitivity, local data heterogeneity, and classification stability.

Although the hybrid models explored earlier improved predictive performance and generalization capability, challenges associated with adaptive neighborhood learning and local decision optimization continued to affect classification performance in heterogeneous psychometric datasets.

Because ML-based methods can identify intricate patterns in vast amounts of psychological and behavioural data, they have become attractive substitutes for automated mental health screening in recent years. Among these approaches, instance-based learning algorithms such as the KNN classifier have gained attention because of their simplicity, interpretability, and effectiveness on tabular medical datasets. However, conventional KNN methods suffer from several limitations, including sensitivity to noise, uniform treatment of neighbors, and dependence on fixed distance parameters, which can negatively affect classification performance in heterogeneous mental health data.

The critical review presented in Section 2.8.1.3 identified limitations in conventional KNN and existing KNN-based approaches, particularly the use of fixed distance measures for heterogeneous psychological assessment data. This chapter presented a novel ASKNN model intended to enhance multi-class categorisation of anxiety and depression severity levels in order to address these issues. The proposed approach dynamically adjusts the influence of neighboring instances using an adaptive sigma parameter, enabling better handling of local data density variations and enhancing classification stability. Because ASKNN uses adaptive neighbourhood weighting, in contrast to conventional KNN variations, it is especially well-suited for complicated and unbalanced mental health datasets.

This chapter presents a comprehensive evaluation of ASKNN using both self-collected and publicly available mental health datasets. The model's performance is assessed using standard evaluation metrics, along with rigorous statistical validation through non-parametric significance testing and effect size analysis. By systematically comparing ASKNN with multiple KNN-based variants, this chapter aims to demonstrate the effectiveness of adaptive neighborhood learning for mental health classification and to establish ASKNN as a robust alternative for automated depression and anxiety detection.

The results of this chapter assist the creation of data-driven tools that are in line with international public health agendas and advance machine learning approaches for accurate and comprehensible mental health evaluation.

6.2 Proposed Methodology

The approach utilised in this work is described in this section, along with the dataset, preprocessing procedures, suggested model, evaluation measures, and statistical analytic methods used to gauge model performance.

6.2.1 Datasets Used

To assess the effectiveness of the suggested ASKNN model, a primary dataset was gathered for this work. To increase the work's validity and generalisability, other benchmark datasets were also used.

6.2.1.1 Proposed Dataset

To assess the effectiveness of the suggested ASKNN model, a primary dataset was gathered for this work. A survey took place between August 2022 and February 2023. Online forums, blogs, and social media groups of university students were sent links to anxiety and depression questionnaires. No payment or compensation was requested, and informed consent clarified the aim of the work to determine the prevalence of anxiety and depression post-pandemic.

Ethics All methods were carried out according to relevant guidelines and regulations. All experimental protocols were approved by the Ethics Committee of the Department of Software Engineering at Delhi Technological University, Delhi, India (Approval ID: DTU/SE/2024/MOM/01). The research was carried out in compliance with the requirements provided by the Checklist for Reporting Results of Internet ESurveys (CHERRIES) [67]. Informed consent was obtained from all participants before their inclusion in this work.

Data Collection Instruments Standardized psychometric scales, including the BDI and BAI, were used to measure depression and anxiety symptoms as explained in detail in section 2.5. The range of the overall score is 0 to 63. The total scores were divided into four groups as shown in Fig. 6.1 and Fig. 6.2. The tentative diagnoses for BDI and BAI are included in Table 6.1 and 6.2, respectively. Although these tools have shown validity and reliability in a variety of populations, the sample used in this study explicitly validated their psychometric qualities.

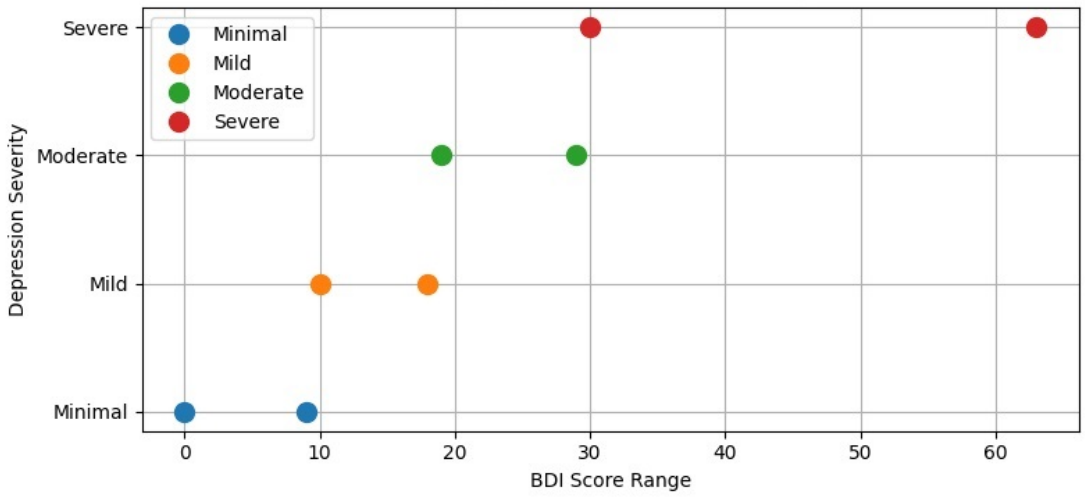


Figure 6.1: BDI score ranges and depression severity levels.

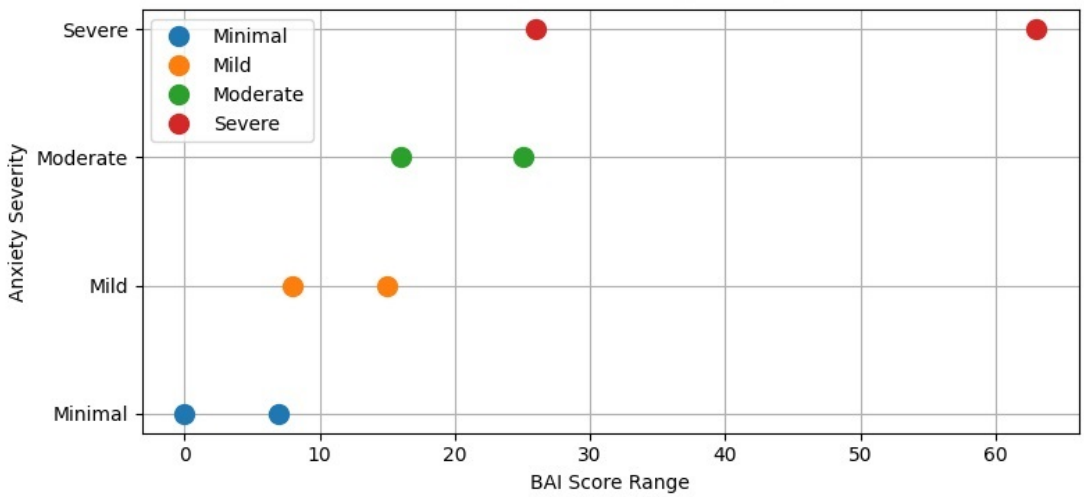


Figure 6.2: BAI score ranges and anxiety severity levels.

Table 6.1: A provisional diagnosis of BDI

BDI Score	Depression Severity
0–9	Minimal Level of Depression
10–18	Mild Depression
19–29	Moderate Depression
30–63	Severe Depression

Table 6.2: A provisional diagnosis of BAI

BAI Score	Anxiety Severity
0–7	Minimal Level of Anxiety
8–15	Mild Anxiety
16–25	Moderate Anxiety
26–63	Severe Anxiety

The dataset includes multiple features corresponding to questionnaire responses, where each feature represents a specific psychological indicator. The responses were encoded numerically to facilitate computational analysis and ML model training.

The gathered dataset is the main source for assessing the suggested model's efficacy under actual circumstances, guaranteeing contextual applicability and relevance.

Study type and population English-literate Indian college and university students who were at least eighteen years of age were given a web-based cross-sectional survey. Excluded

were students who were not enrolled in any Indian educational institution. Two surveys were administered: one for anxiety and one for depression.

- Depression responses: 1596 received, 1552 included in the analysis (Table 6.3).
- Anxiety responses: 1356 received, 1355 included in the analysis (Table 6.4).

The sample is significant based on the sample size calculation:

$$x = Z(c/100)2r(100-r) \tag{6.1}$$

Where r is the estimated percentage of the population that exhibits the characteristic of interest, (100 - r) is the percentage of the population that does not exhibit the characteristic, Z(c/100) is the Z-score linked to the desired confidence level, and x is the necessary sample size [68]. Approximately 412.3 million students are presently enrolled in higher education in Delhi [2]. We employed a sample of 1356 respondents with a 50% response distribution, a 5% margin of error, and a 95% confidence level in order to account for this demography. With an additional 10 to 15% non-respondent error, the minimum sample required for 95% confidence would be 384, bringing the total to 423 to 442. Our sample is therefore noteworthy since $1356 > 442$. Fig. 6.3 and Fig. 6.4 depict how many responses were received for each category (0,1,2,3) for every question from the BDI and BAI surveys, respectively, using a circos plot [126].

Table 6.3: Results of questions based on the BDI

BDI Questions N=1552	**NOT AT ALL = 0**	%	**SEVERAL DAYS = 1**	%	**MORE THAN HALF DAYS = 2**	%	**NEARLY EVERY DAY = 3**	%
Sadness	794	51.16	591	38.08	83	5.35	84	5.41
Pessimism	907	58.44	408	26.29	159	10.24	78	5.03
Past Failure	875	56.38	383	24.68	223	14.37	71	4.57
Loss of Pleasure	761	49.03	540	34.79	178	11.47	73	4.7
Guilty Feelings	859	55.35	536	34.54	104	6.7	53	3.41
Punishment Feelings	1078	69.46	330	21.26	62	3.99	82	5.28
Self-Dislike	1029	66.3	309	19.91	139	8.96	75	4.83
Self-Criticalness	834	53.74	468	30.15	166	10.7	84	5.41
Suicidal Thoughts or Wishes	1306	84.15	170	10.95	33	2.13	43	2.77
Crying	1053	67.85	216	13.92	86	5.54	197	12.69
Agitation	925	59.6	434	27.96	107	6.89	86	5.54
Loss of Interest	752	48.45	517	33.31	186	11.98	97	6.25
Indecisiveness	929	59.86	421	27.13	110	7.09	92	5.93

BDI Questions N=1552	**NOT AT ALL = 0**	%	**SEVERAL DAYS = 1**	%	**MORE THAN HALF DAYS = 2**	%	**NEARLY EVERY DAY = 3**	%
Worthlessness	1103	71.07	275	17.72	117	7.54	57	3.67
Loss of Energy	798	51.42	548	35.31	132	8.51	74	4.77
Changes in Sleeping Pattern	569	36.66	640	41.24	249	16.04	94	6.06
Irritability	955	61.53	446	28.74	88	5.67	63	4.06
Changes in Appetite	776	50	531	34.21	148	9.54	97	6.25
Concentration Difficulty	651	41.95	537	34.6	268	17.27	96	6.19
Tiredness or Fatigue	821	52.9	521	33.57	129	8.31	81	5.22
Loss of Interest in Sex	1272	81.96	155	9.99	59	3.8	66	4.25

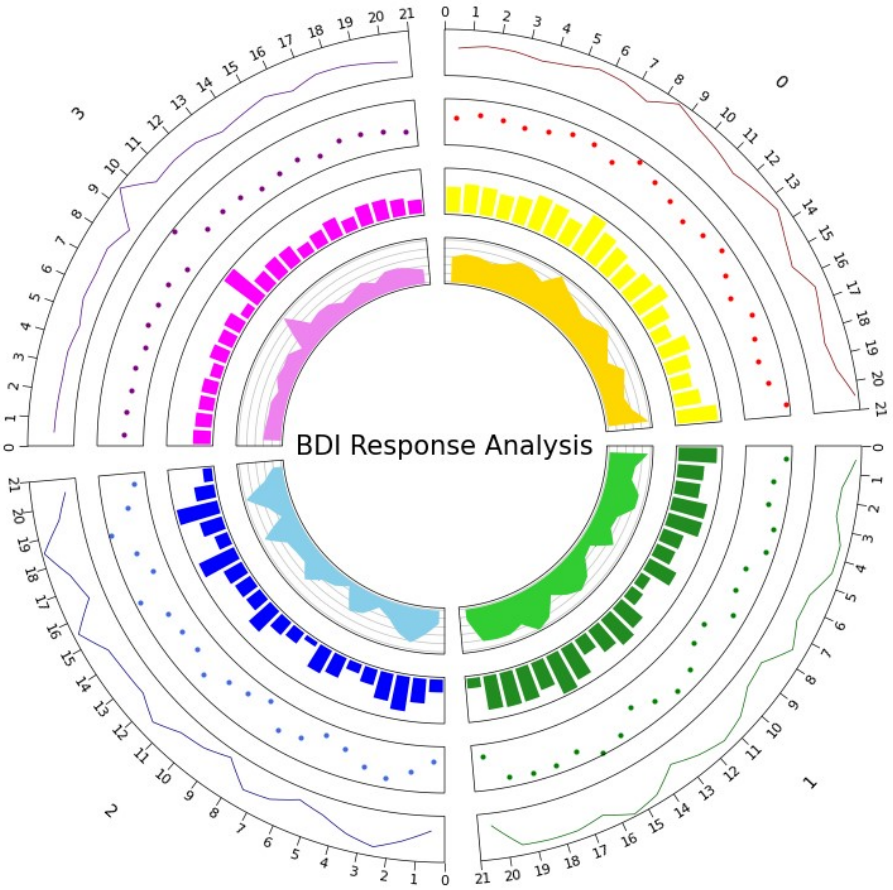


Figure 6.3: BDI Response Analysis Using Circos Plot

Table 6.4: Results of questions based on the BAI

BAI N=1355	**0-Not at all**	%	**1-Mildly, but it didn't bother me much**	%	**2-Moderately – it wasn't pleasant at times**	%	**3-Severely – it bothered me a lot**	%
Numbness or tingling	841	62.07	358	26.42	121	8.93	35	2.58
Feeling hot	725	53.51	378	27.9	167	12.32	85	6.27
Wobbliness in legs	915	67.53	273	20.15	133	9.82	34	2.51
Unable to relax	587	43.32	418	30.85	247	18.23	103	7.6
Fear of worst happening	501	36.97	378	27.9	320	23.62	156	11.51
Dizzy or lightheaded	759	56.01	381	28.12	171	12.62	44	3.25
Heart pounding / racing	713	52.62	389	28.71	197	14.54	56	4.13
Unsteady	816	60.22	349	25.76	138	10.18	52	3.84
Terrified or afraid	721	53.21	403	29.74	163	12.03	68	5.02
Nervous	491	36.24	468	34.54	278	20.52	118	8.71
Feeling of choking	1021	75.35	209	15.42	92	6.79	33	2.44
Hands trembling	936	69.08	277	20.44	102	7.53	40	2.95
Shaky / unsteady	941	69.45	278	20.52	99	7.31	37	2.73

BAI N=1355	0-Not at all	%	1-Mildly, but it didn't bother me much	%	2-Moderately – it wasn't pleasant at times	%	3-Severely – it bothered me a lot	%
Fear of losing control	809	59.7	325	23.99	147	10.85	74	5.46
Difficulty in breathing	968	71.44	247	18.23	98	7.23	42	3.1
Fear of dying	1022	75.42	205	15.13	74	5.46	54	3.99
Scared	821	60.59	360	26.57	119	8.78	55	4.06
Indigestion	811	59.85	349	25.76	145	10.7	50	3.69
Faint / lightheaded	975	71.96	272	20.07	73	5.39	35	2.58
Face flushed	984	72.62	255	18.82	85	6.27	31	2.29
Hot / cold sweats	847	62.51	332	24.5	121	8.93	55	4.06

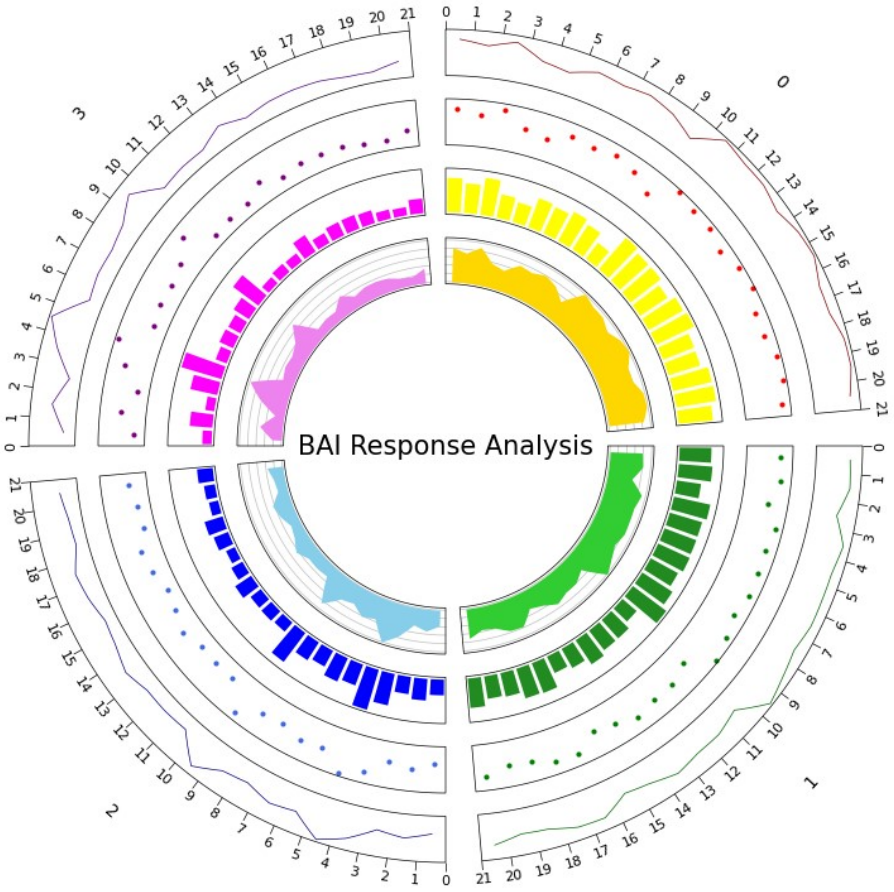


Figure 6.4: BAI Response Analysis Using Circos Plot

6.2.1.2 Benchmark Datasets used

Table 6.5 provides a general overview of the datasets used in this study, including those specifically related to anxiety and depression as well as other medical illnesses for comparison. The table contains important information such as dataset names, number of features used, dataset sizes, and respective sources where available. Our main datasets (D1a and D1b) target depression and anxiety, with 21 features and sample sizes of 1,552 and 1,356, respectively. Furthermore, public datasets (D2a–D5b) for depression and anxiety are provided for model comparison. The datasets differ in feature number and sample sizes, from the smaller ones such as D4a (N=435) to the larger ones such as D5a (N=5,108).

Furthermore, we incorporate additional medical benchmark datasets (D6–D13) for conditions such as diabetes, heart disease, chronic renal disease, breast cancer, and appendicitis

to assess the generalisability of the suggested approach. The datasets are of varying sizes and complexities, with feature numbers between 5 and 24 and sample sizes between 106 and 802. We can evaluate the adaptability of the suggested model for uses other than mental health prediction thanks to the incorporation of these datasets. Since more features typically produce more interesting data for predictive modelling, the number of features used in each dataset is carefully chosen to give valuable analysis. Larger datasets help improve model training and generalization, whereas smaller datasets are helpful for testing performance in low-data situations.

Our dataset is the main object of this work, providing real-world evidence for depression and anxiety in students. The use of publicly available datasets also adds more strength to the evaluation, allowing for solid comparisons. The incorporation of medical datasets also emphasizes the generalizability of the suggested approach in healthcare prediction tasks. Table 6.5 provides an overview of these datasets, clearly describing their main features and sources.

Tables 6.6 and 6.7 present anonymized illustrative snapshots of the primary depression and anxiety datasets (D1a and D1b), respectively, showing representative demographic attributes, questionnaire responses, derived scores, and corresponding classes used in ML models in this work. Detailed characteristics and anonymized illustrative snapshots of other datasets are provided in Chapter 3 and 5 (Table 3.2, Table 3.3, Table 3.4 and Table 5.3).

Table 6.5: Overview of datasets employed in this work, comprising primary depression and anxiety datasets, open-source benchmark datasets, and other medical datasets for comparative work

S. No.	Reference	Dataset Name	Number of features used	Dataset Size	Link
D1a	[12]	Depression	21	N=1552	https://www.nature.com/articles/s41598-025-24252-9
D1b	[12]	Anxiety	21	N=1356	https://www.nature.com/articles/s41598-025-24252-9
D2a	[100]	Depression	9	N=476	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/N5BUJR
D2b	[100]	Anxiety	7	–	–

Continued on next page

Table 6.5 continued from previous page

S. No.	Reference	Dataset Name	Number of features used	Dataset Size	Link
D3a	[188]	Depression	9	N=3569	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FCDGEB
D3b	[188]	Anxiety	7	–	–
D4a	[38]	Depression	14	N=435	https://rdr.ucl.ac.uk/articles/dataset/Self-Perceived_Loneliness_and_Depression_During_the_COVID-19_Pandemic_a_Two-Wave_Replication_Study/20183858
D4b	[38]	Anxiety	14	–	–
D5a	[252]	Depression	13	N=5108	https://github.com/Hu-Li/mental-health-dataset

Continued on next page

Table 6.5 continued from previous page

S. No.	Reference	Dataset Name	Number of features used	Dataset Size	Link
D5b	[252]	Anxiety	9	–	–
D6	[148]	Diabetes	8	N=768	https://www.kaggle.com/datasets/abdallamahgoub/diabetes
D7	[104]	Heart Attack Possibilities	13	N=303	https://www.kaggle.com/datasets/nareshbhat/health-care-data-set-on-heart-attack-possibility
D8	[197]	Chronic Kidney Disease Preprocessed	24	N=400	https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease
D9	[220]	Pima Indians Diabetes	8	N=767	https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

Continued on next page

Table 6.5 continued from previous page

S. No.	Reference	Dataset Name	Number of features used	Dataset Size	Link
D10	[269]	Breast Cancer	5	N=569	https://www.kaggle.com/datasets/merishnasuwal/breast-cancer-prediction-dataset
D11	[254]	Appendicitis	7	N=106	https://www.kaggle.com/datasets/timrie/appendicitis
D12	[66]	Mammographic	5	N=802	https://www.kaggle.com/datasets/overratedgman/mammographic-mass-data-set
D13	[4]	Wine	13	N=178	https://archive.ics.uci.edu/dataset/109/wine

Table 6.6: Anonymized Illustrative Snapshot of the Primary Depression Dataset D1a

Age	Sex	Student	Family Type	D1	D2	D3	D4	D5	D_Score	D_Class
34	Female	1	Joint Family	1	0	0	1	2	20	2
20	Male	1	Nuclear Family	0	0	–	2	3	15	1
21	Male	1	Nuclear Family	0	0	1	0	0	7	0
20	Female	1	Nuclear Family	0	0	0	0	0	0	0
20	Male	1	Nuclear Family	1	0	1	0	2	17	1

Table 6.7: Anonymized Illustrative Snapshot of the Primary Anxiety Dataset D1b

Age	Sex	Student	Family Type	A1	A2	A3	A4	A5	A_Score	A_Class
21	Male	1	Nuclear Family	0	0	2	1	2	26	3
20	Male	1	Nuclear Family	0	1	0	0	0	1	0
20	Male	1	Nuclear Family	0	0	0	0	0	5	0
19	Male	1	Nuclear Family	0	1	1	2	3	20	2
20	Female	1	Nuclear Family	0	0	1	2	1	12	1

6.2.2 Mental health assessment

We used standardised psychological questionnaires from a variety of datasets, including the BDI, BAI, GAD-7, PHQ-9, and SCL-90, to assess participants' mental health conditions as explained in detail in section 2.5. The process of self-reporting using these measures is commonly employed in clinical practice and research to measure symptoms of depression, anxiety, and general psychological distress. The input variables are standardized psychological tests like BDI, BAI, GAD-7, PHQ-9, and SCL-90 subscale scores that quantify symptoms of depression, anxiety, and overall psychological distress. The outcome variable evaluates the seriousness of mental health disorders, classifying people by standardized cutoffs into no mental health disorders, mild, moderate, or severe depression and anxiety. Among these, BDI and BAI were used for primary data collection in this work, while the remaining instruments were considered through benchmark datasets collected by other researchers.

6.2.3 Psychometric Validation

The Python programming language and the statsmodels package (version 0.14.1) were used for the statistical analyses of the exploratory factor analysis and confirmatory factor analysis. All

survey responses were visualized and analyzed using custom Python scripts. The following were done to ensure the reliability and validity of the scale:

- Exploratory Factor Analysis (EFA): Kaiser Meyer Olkin (KMO) measure and Bartlett's Test verified factorability. Principal axis factoring with Promax rotation identified latent factors [39].
- Confirmatory Factor Analysis (CFA): Confirmed EFA factor structure using model fit indices: Normed Fit Index (NFI), Goodness-of-Fit Index (GFI), and Adjusted Goodness-of-Fit Index (AGFI), with acceptable thresholds > 0.90 [175].
- Reliability Analysis: Internal consistency is evaluated using Cronbach's alpha; values > 0.80 are deemed acceptable [96].

6.2.4 Predictive Modeling using KNN and its variants

In this chapter, we compared the performance of ASKNN with multiple KNN variants, including weighted KNN and fuzzy KNN, to evaluate their effectiveness in mental health prediction. The following subsections provide an overview of each variant and its key modifications.

6.2.4.1 KNN

KNN labels a point according to the majority vote of its closest neighbors based on a distance measure, like Euclidean distance [253] as explained in detail in section 2.8.1.3. The ease of KNN is one of its primary strengths, as it has no training phase and can be used directly on the data. KNN has its challenges, particularly in high-dimensional data, because of the curse of dimensionality, where the distance between points becomes less significant. Furthermore, KNN is not very effective when working with huge datasets because it requires computing the distance between the query point and every other data point in the dataset for each prediction. KNN also has difficulties with noisy data and unbalanced classes, where the distribution of classes is not even.

6.2.4.2 AKNN

AKNN is a modified version of the basic KNN algorithm, where the impact of every neighbor is dynamically modified according to some properties of the data. This makes the model flexible to adapt to different data distributions and hence more flexible than the basic KNN [230]. In AKNN, the weights of the neighbors are usually modified depending on their proximity or

some other factors, for example, the data point density in the surrounding areas. The closer neighboring points are assigned higher weights, which will emphasize the most salient data points in regression or classification tasks. This modification helps the algorithm concentrate on the most relevant neighbours, increasing classification accuracy and robustness, and is especially helpful when the dataset has complicated, non-uniform distributions. AKNN does very well on data with varying data density across environments or when the data patterns are changing from region to region.

6.2.4.3 FKNN

FKNN is an extension of the regular KNN algorithm implemented with a system of fuzzy logic for classification purposes. While the traditional KNN assigns a hard class label based on the majority of nearest neighbours, FKNN provides membership values that show how much a data point belongs to any class [112]. This probabilistic approach proves useful when data points are not separable in a linear fashion and boundaries of classes are indistinct. Taking the uncertainty into consideration and allowing for a data point to be part of multiple classes with different levels of membership, FKNN can make more nuanced predictions, particularly where classes intersect or when data is noisy. This method increases the model's stability, especially under real-world settings where hard class boundaries are non-existent, say in medical diagnostics or natural language processing.

6.2.4.4 KMKNN

KMKNN is a hybrid method that employs K-means clustering to integrate with the KNN algorithm. With this method, KNN is applied to the centroids of clusters found by the K-means algorithm rather than raw data points [45]. This minimizes the computational expense of the KNN algorithm by minimizing the number of calculations involving the distance because the K-means clustering minimizes the data to a series of centroids to represent the major features of the data. KMKNN performs best when the data is naturally grouped since it takes advantage of the structure determined by the clustering algorithm. It can make the classification process more efficient and scalable, especially with large sets of data, by using centroids to represent multiple points that possess similar characteristics as opposed to treating every point in isolation.

6.2.4.5 WKNN

WKNN is one of the standard KNN variations, whereby each one of the closest neighbors contributes through the weighting under certain criteria such as proximity, weights, or belonging to a class [130]. They are given larger weights and so receive increased impact when predicting or classifying a point. By modifying the weights, the algorithm can lessen the impact of noisy or meaningless neighbours, making predictions more reliable and accurate. This method enhances the efficacy of KNN by better representing the value of each neighbour, especially in cases where the data contains noise or exceptions. The algorithm may lessen the influence of noisy or irrelevant neighbours by changing the weights, which improves the accuracy and consistency of predictions. WKNN is most useful in unbalanced datasets, where the number of data points in some of the classes is skewed, either being much higher or much lower, allowing the model to give more focus to the minority class and generally improving the classification accuracy.

6.2.4.6 GMDKNN

GMDKNN is a special form of KNN that changes the default distance function used to measure proximity among points. Unlike relying solely on Euclidean or Manhattan distances, this method employs the generalized mean (or power mean) distance, a more general distance function. The generalised mean can be used with different distance measures, depending on the type of data and sensitivity level. The algorithm can control the impact of various features or dimensions in the dataset by using the generalized mean, which makes it more appropriate for complex or heterogeneous data. This adaptation is more flexible and can improve performance when the usual distance measures cannot capture the patterns in the data [81].

6.2.4.7 MKNN

MKNN is an extension of the simple KNN algorithm that considers not only the proximity between data points but also their mutual similarity. In this approach, two data points are neighbors if they are proximal to each other and have a mutual similarity, which can be measured in terms of similarity measures like cosine similarity or correlation [54]. This similarity strategy for mutual similarity can enhance the quality of nearest neighbors by focusing on those that are not only nearest in distance but also have considerable relationships between themselves, thereby enhancing predictions. MKNN is useful in practice where relationships of data points are based on similarity rather than being geometric, e.g., collaborative filtering for recommendations or

in-text classification.

6.2.4.8 EKNN

Several KNN classifiers (or other ML models) are combined to create EKNN, which produces a final prediction. The method makes use of the ensemble learning idea, which involves training several models separately and combining their output, frequently by averaging or voting [92]. In KNN, an ensemble approach may involve the merging of predictions of multiple KNN models with differing values of K, distance functions, or sets of features. This helps to reduce the variance and bias of individual models to yield better and more stable predictions. EKNN is particularly well-suited for noisy data or imbalanced class cases as it takes advantage of the strengths of different models to exhibit greater generalization and reduce overfitting hazards.

6.2.4.9 ASKNN

Here, we introduce ASKNN, an extension of the basic KNN algorithm that includes an adaptive weighting mechanism depending on the local concentration of points. The ASKNN performs better than the standard KNN because it adjusts the influence of each neighbor based on how densely or thinly the points are distributed in the feature space. In the following, we describe the key aspects of the ASKNN algorithm.

- **Local Density Calculation**

  The principle of the ASKNN algorithm is to compute the local density around each test point. Local density is a measure of the proximity of neighbors to one another in feature space. For each test sample, the algorithm estimates the inverse distance to the k-nearest neighbors, which serves as a measure of how proximal or remote the neighbors are. Local density is then defined as the inverse of the mean distance and impacts the adaptation of the sigma parameter governing neighbor weighting.

  The formula for local density:

$$\text{density} = \frac{1}{d_1 + \varepsilon} + \frac{1}{d_2 + \varepsilon} + \cdots \tag{6.2}$$

  Where d_i are the distances to the nearest neighbors, and ε is a small constant to avoid division by zero.

  This step creates a dynamic measurement of how tightly clustered the neighboring points are around the query point. If the neighbors are close together, the calculated density

will be great; otherwise, if the neighbors are more spread out, the density will be low. This local density information serves as an approximation for calculating the ”spread” of data around the test point, which is used later to scale the influence of neighbors in the classification step.

- **Adaptive Sigma Adjustment**

  Once the local density has been computed, it is used to adjust the value of sigma. Sigma controls the spatial extent of the Gaussian function that is used to compute the weights of a neighbor. A larger local density (i.e., nearer neighbors) results in a smaller value for sigma, so each neighbor's weight is more localized. A smaller local density (i.e., a sparse neighborhood) results in a large sigma, so any individual neighbor contributes very little.

  The formula for adaptive sigma:

$$\sigma_{\text{adaptive}} = \frac{\sigma_0}{\text{density} + \varepsilon} \tag{6.3}$$

  Where σ_0 is a base sigma parameter, and ε is a small constant to avoid division by zero.

  The motivation for adaptive sigma is to make the smoothness of the Gaussian kernel vary according to the local data distribution. In high-density areas, we want sharp kernels (small sigma) to highlight near neighbors, while in low-density areas, flat kernels (large sigma) are used to prevent overfitting to individual points. This adaptation makes ASKNN perform similarly to a locally sensitive classifier.

  The use of a dynamically adaptive σ is motivated by the varying density of observations across the feature space. A static global σ applies the same smoothing scale to all regions, although different neighborhoods may have different inter-point distances. ASKNN therefore adapts σ according to the local neighborhood of each query point. Dense regions result in a smaller σ, providing more localized weighting, whereas sparse regions result in a larger σ, providing smoother weighting. Computationally, σ is recalculated for each query point from its local neighborhood density, enabling the Gaussian weighting mechanism to adapt to local data characteristics rather than using a single fixed scale across the feature space.

- **Weight Calculation**

  For every test sample, the neighbor weights are calculated with a Gaussian kernel. The adaptive sigma is used to transform the distance between a test sample and its neighbors,

assigning higher weights to closer neighbors and lower weights to farther ones. The weight of every neighbor is calculated as:

$$\text{weight} = \exp\left(-\frac{d^2}{2 \cdot \sigma^2_{\text{adaptive}}}\right) \tag{6.4}$$

This weight captures the closeness of a neighbor, with closer neighbors having greater weights, and distant ones having lower influence.

This process ensures that neighboring contributions are not equal, as in standard KNN, but weighted by distance and local data density. Using the Gaussian function, ASKNN smoothly diminishes the influence of farther neighbors, making the classification result more insensitive to outliers or sparse samples.

- **Classification**

  Based on the weighted voting process, the method generates a projected class label for the test sample after calculating the weights for the KNN. Each class label is given a score, which is the total of the weights of the neighbours who cast votes for that specific class. The class with the greatest cumulative score chooses the projected label for the test sample.

  Predicted class:

$$\hat{y} = \arg\max_{\omega_j}\left(\sum_{i \in \text{neighbors}} w_i\right) \tag{6.5}$$

  where w_i is the weight assigned to each neighbor, and ω_j is the class label.

  Unlike unweighted KNN, where all neighbors have the same impact, ASKNN enables strong neighbors to dictate the outcome through increased weights as shown in Fig. 6.5. This leads to improved class separation, particularly in imbalanced or heterogeneous datasets.

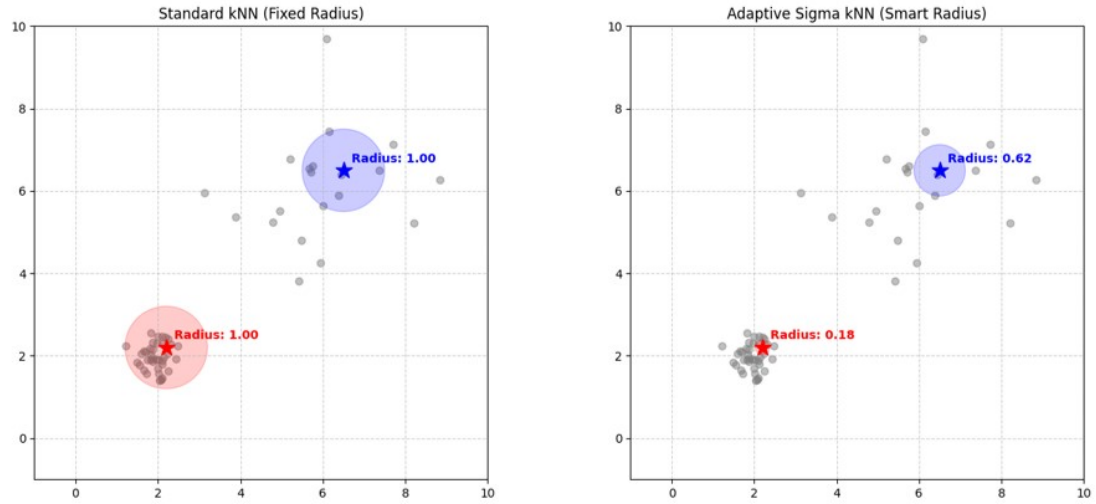


Figure 6.5: Adaptive neighborhood in ASKNN.

6.2.4.10 Advantages of Adaptive Sigma KNN

Adaptive to Local Density: Unlike classical k-NN techniques, which employ a constant weight (most often uniform) for all neighbors, Adaptive Sigma KNN tunes the weight according to the local concentration of the data points. This makes the neighbors' influence proportional to the density of the neighborhood surrounding the query point.

Improved Performance in Sparse Data: Where the data is sparse, the model addresses the absence of proximate neighbors by diminishing the influence of far-away neighbors and making the classification more resilient.

Better Flexibility: The adaptive sigma mechanism makes the model more versatile and performs better on datasets with different densities, particularly where there are imbalances in classes or diverse feature spaces.

Table 6.8 presents a comparative overview of KNN and its variants, focusing on their important features and advantages. The summary facilitates the understanding of how the various modifications of the basic KNN method contribute to the improvements in different contexts.

Table 6.8: Summary of KNN Variants and Their Key Characteristics

KNN Variant	**Ref**	**Description**	**Dataset Used**	**Accuracy**	**Key Features**	**Benefits**
KNN	[246]	Majority voting of nearest neighbors for classification.	Benchmark disease datasets	76.35 (Heart Attack)	No training phase; simple; distance-based	Easy to implement; good for small datasets
AKNN	[246]	Adaptive KNN with dynamic neighbor weighting.	Benchmark disease datasets	73.64 (Heart Attack)	Weighted neighbors; adapts to distribution	Flexible; works for non-uniform data
FKNN	[112]	Fuzzy logic-based classification with membership values.	IRIS, TWOCLASS	91–96%	Probabilistic membership; handles ambiguity	Handles overlapping classes; robust to noise
KMKNN	[246]	K-means clustering combined with KNN using centroids.	Benchmark disease datasets	39.86 (Heart Attack)	Uses centroids; reduces dataset size	Scalable; efficient for large datasets

KNN Variant	Ref	Description	Dataset Used	Accuracy	Key Features	Benefits
WKNN	[23]	Weighted KNN based on distance or importance.	Haberman dataset	0.70	Weight adjustment; reduces noise impact	Better for imbalanced datasets
GMDKNN	[81]	Generalized mean-based distance metric.	UCI datasets	87.76 ± 6.52	Flexible distance; feature tuning	Adapts to diverse data; improves accuracy
MKNN	[54]	Uses mutual similarity along with proximity.	ILPD, Glass	85.92% (Glass)	Similarity metrics; mutual relationships	Improves prediction reliability
EKNN	[23]	Ensemble of multiple KNN classifiers.	Haberman dataset	77%	Multiple models; voting/averaging	Reduces variance; improves stability
ASKNN	Proposed	Adaptive KNN using sigma parameter.	13 datasets	Up to 98.6%	Dynamic weighting using sigma	Robust to noise and complex patterns

6.2.5 Data Preprocessing and Predictive Modelling

For prepping the dataset for predictive modeling, some important data preprocessing steps were undertaken. Missing value imputation was done first to manage incomplete data and ensure dataset integrity. Categorical variables were one-hot encoded as required.

For predictive modeling, several KNN-based classifiers such as ASKNN, AKNN, FKNN, KMKNN, WKNN, GMDKNN, MKNN, and EKNN were used. The models were executed with the k parameter varying from 1 to 10, and the performance was determined based on chosen metrics. Using Grid Search Cross-Validation (Grid Search CV), hyperparameter tweaking was done to find the best ASKNN configuration to get the best model performance.

6.2.6 Evaluation Metrics

The KNN variations are tested for performance using common measures, including accuracy and classification reports. The model is applied to a dataset with labeled ground truth, and predictions are measured against the ground truth. While the classification report includes F1-score, precision, and recall for each class, accuracy is determined by the percentage of properly categorized samples. Additionally, 5-fold cross-validation is employed to further examine the results. These metrics were defined in detail in Section 2.10.

6.2.7 Statistical Analysis of Performance of ASKNN

The statistical analysis was conducted using the Python programming language and the `statsmodels` package (version 0.14.1). Python programs were used to analyze and visualize the survey results. To determine the statistical significance of the performance differences, we used the Wilcoxon signed-rank test to compare ASKNN with other KNN-based models on four crucial assessment metrics: recall, accuracy, precision, and F1-score. The statistical significance was determined using the significance threshold of $\alpha = 0.05$. Additionally, Cohen's d and Cliff's Delta were calculated to show the size of performance disparities.

6.3 Results

This section reports the experimental results, covering statistical analysis, psychometric validation, and model performance evaluation.

6.3.1 Experimental Findings of the Dataset

This section presents the experimental findings, including descriptive statistical analysis, psychometric validation, and the performance evaluation of the proposed ASKNN model.

6.3.1.1 Descriptive and Statistical Analysis of the Dataset

Among the 1552 participants, 46.9% of respondents suffered from a minimal level of depression, 28.4% of respondents suffered from mild depression, 17.4% of respondents suffered from moderate depression, and 7.3% of students suffered from severe depression, as represented by Fig. 6.6a. Out of 1355 participants, 44.2% of students had minimal anxiety, 23.4% of students suffered from mild anxiety, 18.7% of students were diagnosed with moderate anxiety, and 13.7% of students suffered from severe anxiety, as represented by Fig. 6.6b. Table 6.9 and Table 6.10 present the descriptive and factor analytic findings for the BDI and the BAI items, respectively. The BDI items generally yielded moderate means and adequate variability, with some items having positive skewness. In a similar vein, the BAI items, though predominantly marked by low to moderate mean values, evidenced high skewness and kurtosis for some items, suggesting that very few participants reported severe symptoms of anxiety.

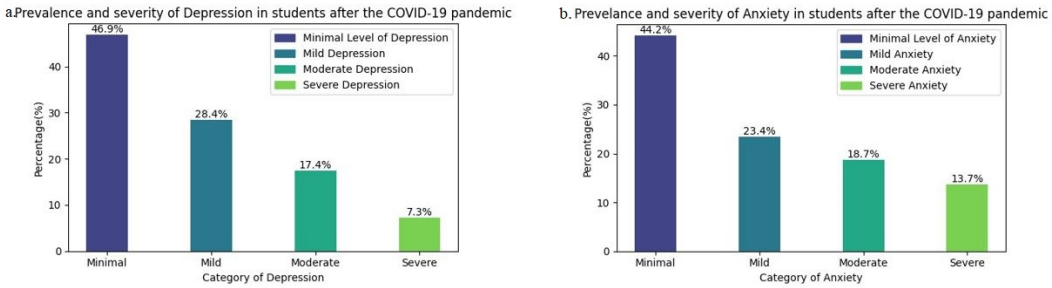


Figure 6.6: a Prevalence and severity of depression in students after the COVID-19 pandemic b Prevalence and severity of anxiety in students after the COVID-19 pandemic

Table 6.9: Descriptive Statistics and Factor Analytic Results for BDI Items

Item	M	SD	95% CI	h^2	FL	Skewness	Kurtosis
1	0.52	0.764	0.48–0.56	0.427	0.653	1.403	1.305
2	0.71	0.91	0.67–0.76	0.278	0.528	1.095	0.204
3	0.47	0.772	0.43–0.51	0.455	0.674	1.548	1.519
4	0.9	0.954	0.85–0.95	0.469	0.685	0.724	-0.551
5	1.1	1.029	1.04–1.15	0.447	0.669	0.442	-1.028
6	0.63	0.825	0.59–0.67	0.499	0.706	1.121	0.373
7	0.7	0.865	0.66–0.75	0.522	0.723	0.998	0.031
8	0.58	0.824	0.53–0.62	0.561	0.749	1.333	0.952
9	0.69	0.87	0.64–0.73	0.556	0.746	1.105	0.342
10	1.02	0.957	0.97–1.07	0.485	0.696	0.563	-0.703
11	0.36	0.717	0.32–0.40	0.478	0.691	2.035	3.468
12	0.44	0.758	0.40–0.48	0.513	0.716	1.727	2.299
13	0.43	0.746	0.39–0.47	0.597	0.773	1.748	2.416
14	0.62	0.884	0.57–0.67	0.547	0.739	1.293	0.675
15	0.42	0.758	0.38–0.46	0.452	0.672	1.839	2.663
16	0.38	0.764	0.34–0.42	0.359	0.599	2.114	3.721
17	0.56	0.816	0.52–0.61	0.49	0.7	1.404	1.241
18	0.58	0.824	0.54–0.63	0.396	0.629	1.3	0.843
19	0.39	0.707	0.35–0.42	0.519	0.721	1.962	3.472
20	0.38	0.707	0.34–0.42	0.507	0.712	1.929	3.215
21	0.55	0.819	0.50–0.59	0.417	0.646	1.453	1.319

CI confidence interval, FL factorial loads, h^2 communalities, M mean, SD standard deviation

Table 6.10: Descriptive Statistics and Factor Analytic Results for BAI Items

Item	M	SD	95% CI	h^2	FL	Skewness	Kurtosis
1	0.65	0.812	0.61–0.69	0.424	0.651	1.326	1.43
2	0.62	0.862	0.58–0.66	0.44	0.663	1.29	0.775
3	0.67	0.885	0.63–0.72	0.394	0.628	1.088	0.12
4	0.72	0.845	0.68–0.76	0.423	0.65	1.038	0.369
5	0.58	0.763	0.54–0.62	0.443	0.666	1.329	1.463
6	0.45	0.803	0.41–0.49	0.339	0.582	1.923	3.039
7	0.52	0.848	0.48–0.57	0.526	0.725	1.559	1.464
8	0.68	0.87	0.63–0.72	0.487	0.698	1.165	0.52
9	0.24	0.624	0.20–0.27	0.387	0.622	3.056	9.35
10	0.63	1.052	0.58–0.68	0.315	0.561	1.434	0.523
11	0.58	0.845	0.54–0.63	0.46	0.678	1.459	1.409
12	0.76	0.893	0.72–0.80	0.491	0.701	1.015	0.179
13	0.59	0.86	0.55–0.63	0.481	0.694	1.451	1.314
14	0.44	0.786	0.40–0.48	0.576	0.759	1.807	2.444
15	0.67	0.824	0.63–0.71	0.572	0.756	1.197	0.886
16	0.91	0.873	0.87–0.96	0.358	0.599	0.712	-0.204
17	0.52	0.779	0.48–0.56	0.484	0.695	1.574	2.058
18	0.72	0.876	0.68–0.76	0.391	0.625	1.132	0.535
19	0.88	0.908	0.83–0.92	0.468	0.684	0.741	-0.382
20	0.66	0.839	0.62–0.70	0.557	0.746	1.24	0.916
21	0.3	0.737	0.27–0.34	0.166	0.408	2.578	5.811

CI confidence interval, FL factorial loads, h^2 communalities, M mean, SD standard deviation

6.3.1.2 Correlation and Reliability Analysis

Pearson correlation analysis was used to determine the most relevant features that correlate with depression and anxiety scores. On the BDI, the highest positive correlations $(r > 0.6)$ were seen with sadness, pessimism, crying, indecisiveness, and self-dislike. There were moderate correlations (r = 0.5 to 0.6) with loss of pleasure, self-criticalness, and agitation. Also, for the BAI, symptoms like inability to relax, fear of worst happening, heart thumping, nervousness, and feeling frightened were strongly correlated $(r > 0.6)$, while numbness, wobbliness, dizziness, and fear of losing control were moderately correlated. The symptom 'fear of dying' was the least correlated. Fig. 6.7 and Fig. 6.8 highlight significant anxiety and depression factors.



Figure 6.7: Significant Depression Factors

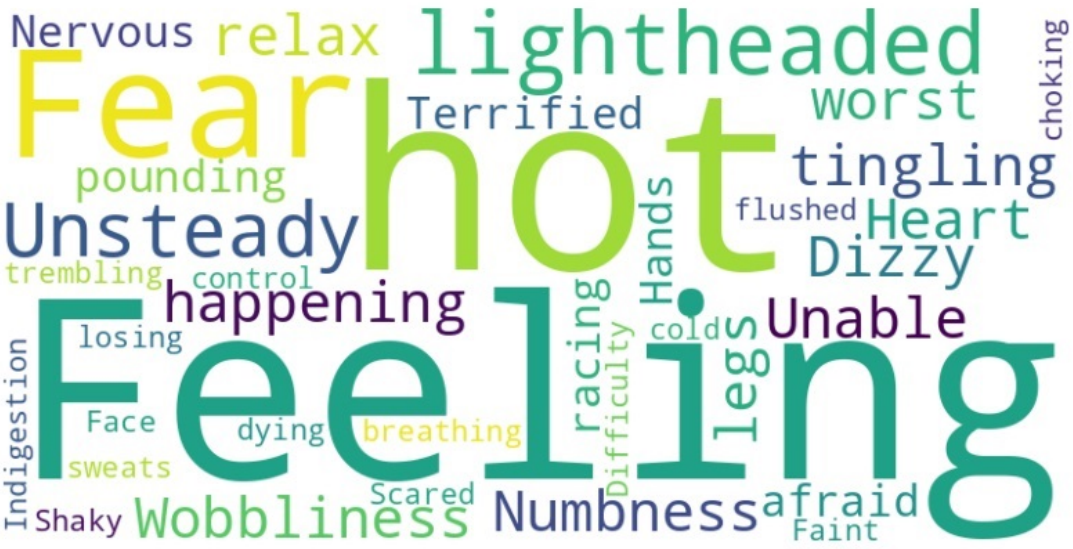


Figure 6.8: Significant Anxiety Factors

Cronbach's alpha was computed for both the total scores and each of the identified subscales in order to evaluate the scales' internal consistency. Table 6.11 shows Cronbach's alpha values, reflecting very good overall reliability for both BDI ($\alpha = 0.93$) and BAI ($\alpha = 0.95$). The subscales further reflected high internal consistency, with alpha values between 0.83 and 0.89, indicating the construct reliability of the extracted factors from both inventories.

Table 6.11: Cronbach's Alpha for Total and Subscales of BDI and BAI

Measure	Cronbach's Alpha (Total)	Subscale 1 α	Subscale 2 α	Subscale 3 α
BDI	0.93	0.83	0.87	0.84
BAI	0.95	0.86	0.89	0.86

6.3.1.3 Factor Analysis (EFA and CFA)

Table 6.12 presents the findings of factor analyses. For BDI, two factors accounted for 56.72% of the variance, whereas three factors for BAI accounted for 53.67%. Both measures showed superb sampling adequacy as evidenced by KMO values greater than 0.9 and Bartlett's test results that were significant at $p < 0.001$. Fit statistics from CFA, such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), NFI, GFI, and AGFI, were all higher than the generally applied standards (> 0.9), and Root Mean Square Error of Approximation (RMSEA) was less than 0.08, establishing good model fit.

Table 6.12: EFA and CFA Summary for BDI and BAI

Measure	No. of Factors	Total Variance Explained (%)	KMO	Bartlett's Test (χ^2, p)	CFI	TLI	RMSEA	NFI	GFI	AGFI
BDI	2	56.72	0.92	730.41, $p < 0.001$	0.91	0.89	0.078	0.91	0.92	0.9
BAI	3	53.67	0.97	608.17, $p < 0.001$	0.91	0.89	0.078	0.94	0.94	0.92

The factor loading structures and latent-observed variable relationships are visually represented in Fig. 6.9 for BDI and Fig. 6.10 for BAI. These results confirm the adequacy and validity of the factor structures for both BDI and BAI.

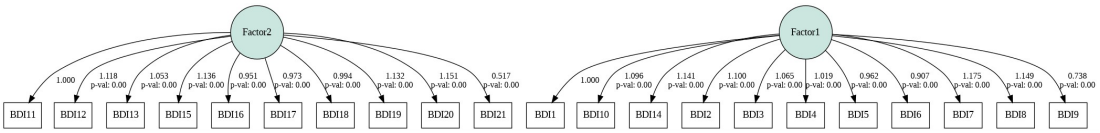


Figure 6.9: CFA path diagram for BDI that displays factor loadings and correlations between latent variables and observable items.

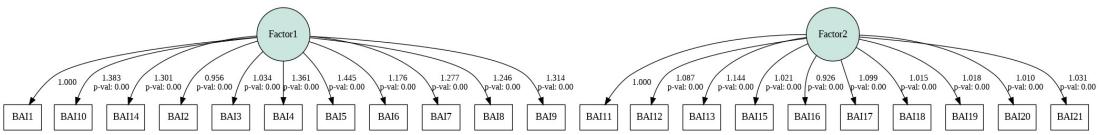


Figure 6.10: CFA path diagram for BAI that displays factor loadings and correlations between latent variables and observable items.

6.3.2 Model Performance Evaluation Across All Datasets

The performance of the suggested ASKNN model is next assessed once factor analysis has confirmed the dataset's validity and reliability. The performance comparison between ASKNN and conventional KNN variations on important metrics, including Accuracy, Precision, Recall, F1-score, and AUC–ROC is shown in this subsection. To determine the model's generalisability, it was tested using BAI and BDI datasets as well as additional medical data.

ASKNN performed better on all the measures with enhanced classification accuracy and stability. Cross-validation was used to confirm consistency, and the Friedman and Wilcoxon signed-rank tests were used to assess statistical significance. Additionally, impact size metrics (Cohen's d and Cliff's Delta) were used to gauge the improvements. The findings, presented in tables and figures, emphasize ASKNN's efficiency in mental health classification, making it a promising predictive diagnostics tool.

6.3.2.1 Accuracy

The accuracy scores depict the performances of various KNN variants on classification tasks. Table 6.13 displays the accuracy scores of multiple KNN-based models on several datasets, including our ASKNN model.

Our ASKNN demonstrates higher accuracy than regular KNN and other alternatives on the majority of the datasets evaluated. Particularly, on D1a and D1b (our main datasets for depression and anxiety detection), ASKNN performs with 91% and 85% accuracy, respectively, compared to KNN (87% and 73%), as depicted in Fig. 6.11. Similar patterns are evident in benchmark datasets like D3a and D5b, where ASKNN achieves 97% and 99% accuracy, respectively.

Table 6.13: Accuracy of various KNN implementations on multiple datasets, such as depression, anxiety, and general medical datasets.

Dataset	**KNN**	**A KNN**	**FKNN**	**KMKNN**	**WKNN**	**GMDKNN**	**MKNN**	**EKNN**	**ASKNN**
D1a	0.87	0.88	0.90	0.48	0.89	0.22	0.84	0.62	**0.91**
D1b	0.73	0.84	0.84	0.23	0.85	0.59	0.80	0.63	**0.85**
D2a	0.80	0.56	**0.91**	0.82	0.45	0.92	0.89	0.88	0.90
D2b	0.91	0.61	0.42	0.62	0.57	0.90	0.87	0.89	**0.91**
D3a	0.90	**0.97**	0.96	0.30	0.93	0.94	0.96	0.80	**0.97**
D3b	0.90	**0.96**	**0.96**	0.30	0.87	0.90	0.94	0.81	0.95
D4a	0.75	**0.87**	0.83	0.53	0.59	0.61	0.83	0.70	**0.87**
D4b	0.71	0.79	0.80	0.44	0.56	0.65	0.77	0.76	**0.82**
D5a	0.95	**0.98**	**0.98**	0.78	0.86	0.90	0.96	0.82	**0.98**
D5b	0.98	0.97	0.98	0.75	0.83	0.90	0.96	0.88	**0.99**
D6	0.68	**0.74**	0.66	0.65	0.70	0.69	**0.74**	0.65	0.71
D7	0.82	0.69	0.67	0.55	0.55	0.69	0.70	0.66	**0.84**
D8	0.97	0.66	0.71	0.63	0.71	**0.99**	0.62	0.59	**1.00**
D9	0.68	**0.75**	0.66	0.63	0.65	0.69	0.74	0.65	0.72
D10	0.92	0.84	0.89	0.84	0.77	0.89	**0.93**	0.92	**0.94**
D11	0.86	0.77	0.77	0.44	0.84	0.81	0.77	0.77	**0.86**
D12	0.75	**0.81**	**0.81**	0.51	**0.81**	0.80	0.80	0.77	**0.81**
D13	0.72	0.69	0.81	0.26	0.80	0.80	0.78	0.81	**0.94**

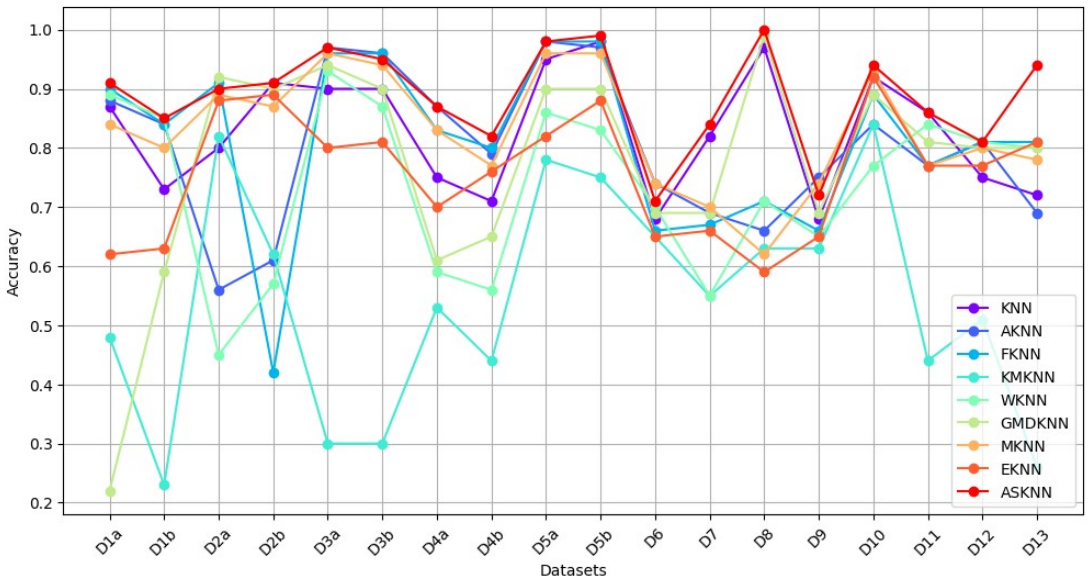


Figure 6.11: Accuracy comparison of different KNN variants across multiple datasets.

Among other alternatives, FKNN, WKNN, and GMDKNN have competitive performance in

certain situations but vary significantly across datasets. KMKNN has overall lower performance, particularly on mental health datasets, suggesting its inability to work effectively on high-dimensional tabular data. EKNN techniques exhibit modest gains but do not consistently surpass ASKNN. Overall, these results demonstrate that ASKNN provides a more stable and high-performing approach for depression and anxiety detection, making it a promising technique for mental health classification tasks.

6.3.2.2 Precision

Higher precision scores indicate a model's capacity to avoid false positives, which is critical for the diagnosis of anxiety and depression. Precision is a critical metric in the classification of mental health conditions since it is the ratio of correctly diagnosed positive instances to all projected positive cases. Higher precision scores indicate a model's capacity to avoid false positives, which is essential for the diagnosis of anxiety and depression. Precision values for several KNN-based models in various datasets, along with our ASKNN approach, are displayed in Table 6.14.

The findings illustrate that ASKNN is consistently outperforming in terms of precision in the majority of datasets. In D1a and D1b (our main datasets for depression and anxiety detection), ASKNN performs with 92% and 86% precision, respectively, outperforming conventional KNN (89% and 77%) as illustrated in Fig. 6.12.

Table 6.14: Precision scores for different KNN variants over different datasets.

Dataset	KNN	A KNN	FKNN	KMKNN	WKNN	GMDKNN	MKNN	EKNN	ASKNN
D1a	0.89	0.9	**0.92**	0.12	0.89	0.24	0.88	0.57	**0.92**
D1b	0.77	**0.86**	**0.86**	0.06	0.85	0.56	0.82	0.62	**0.86**
D2a	0.8	0.48	0.9	0.83	0.31	0.93	0.89	**0.94**	0.89
D2b	0.92	0.46	0.26	0.21	0.4	0.89	0.84	**0.95**	0.89
D3a	0.91	**0.97**	0.96	0.1	0.89	0.91	0.96	0.51	**0.97**
D3b	0.84	**0.92**	0.88	0.1	0.68	0.58	0.95	0.57	0.9
D4a	0.78	**0.9**	0.84	0.18	0.54	0.62	0.84	0.5	**0.9**
D4b	0.72	0.81	0.81	0.15	0.49	0.66	0.81	0.51	**0.82**
D5a	0.96	0.97	**0.98**	0.26	0.78	0.86	0.96	0.74	0.97
D5b	**0.98**	0.97	0.96	0.25	0.79	0.89	0.96	0.91	**0.98**
D6	0.66	0.72	0.64	0.33	0.67	0.66	0.72	**0.82**	0.68
D7	0.82	0.69	0.67	0.27	0.27	0.69	0.7	0.7	**0.84**
D8	0.97	0.65	0.72	0.57	0.7	0.99	0.61	0.54	**1**
D9	0.66	0.74	0.64	0.57	0.33	0.66	0.72	**0.82**	0.7
D10	0.92	0.86	0.89	0.9	0.81	0.89	0.93	0.94	**0.95**
D11	0.82	0.73	0.73	0.57	0.8	0.75	0.73	0.73	**0.92**
D12	0.76	**0.82**	0.81	0.76	0.81	0.8	0.8	0.77	0.81
D13	0.69	0.68	0.78	0.09	0.8	0.8	0.74	0.79	**0.94**

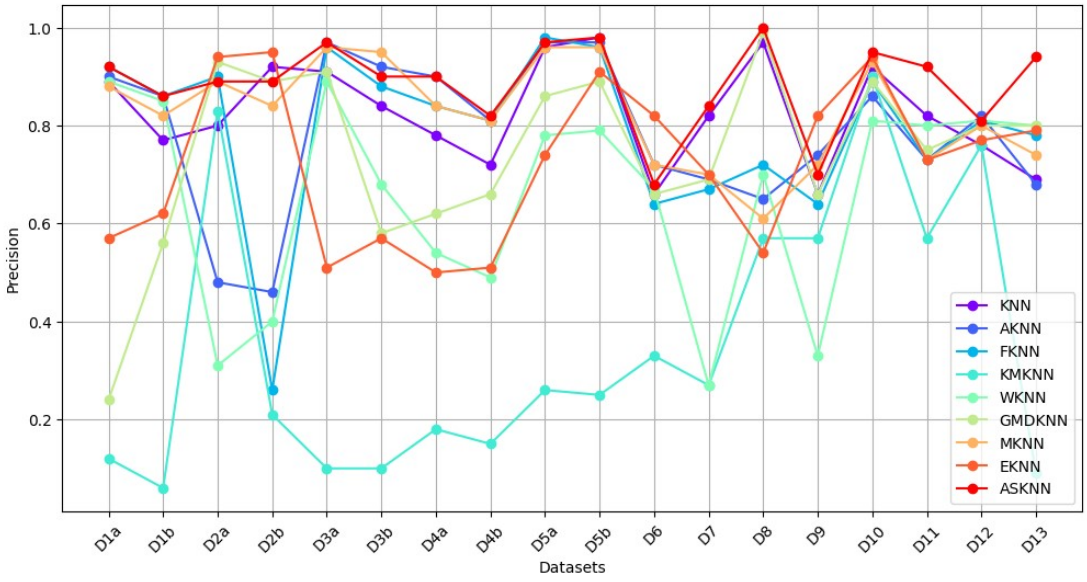


Figure 6.12: Precision comparison of different KNN variants across multiple datasets.

Comparable improvements are seen in benchmark datasets like D3a and D5b, where ASKNN performs with 97% and 98% precision, respectively. Among other KNN variants,

FKNN, GMDKNN, and MKNN perform well in certain datasets but are not as uniformly high-performing as ASKNN. KMKNN, however, performs poorly in most datasets, with much lower precision values, especially on mental health datasets, reflecting its poor performance in these cases. EKNN methods provide modest gains but still fail to surpass ASKNN. Overall, these results further support the effectiveness of ASKNN on mental health classification problems, specifically depression and anxiety detection. The model persistently outperforms traditional KNN and its variants, as it is a viable option for precision-oriented classification tasks.

6.3.2.3 Recall

Recall measures a model's ability to correctly identify positive cases, making it a crucial metric in the classification of mental health. Larger recall values are indicative of improved capacity to capture truly positive cases, which is especially important in depression and anxiety detection, where failing to capture positive cases has severe implications. The findings in Table 6.15 demonstrate a detailed comparison of recall scores for different KNN-based models, including the proposed ASKNN, over different datasets.

The findings show that, on all but two datasets, ASKNN consistently outperforms the other KNN variations in terms of recall values. For example, in the D1a and D1b datasets (anxiety and depression detection), ASKNN obtains a recall of 0.86 and 0.79, respectively, surpassing regular KNN, which registers 0.82 and 0.69. The same trend is found in other datasets, including D3a, where ASKNN has 0.92 recall, and D5b, where it has 0.96 recall. In D8, ASKNN also has a perfect recall of 1.00, reflecting its high performance in classifying all positive cases as depicted in Fig. 6.13.

Table 6.15: Recall scores for different KNN variants over different datasets, establishing the enhanced effectiveness of ASKNN at accurately identifying positive cases.

Dataset	**KNN**	**A KNN**	**FKNN**	**KMKNN**	**WKNN**	**GMDKNN**	**MKNN**	**EKNN**	**ASKNN**
D1a	0.82	0.8	0.86	0.25	0.84	0.21	0.78	0.39	**0.86**
D1b	0.69	0.77	0.79	0.29	0.79	0.48	0.73	0.49	**0.79**
D2a	0.8	0.39	0.91	0.84	0.32	0.9	0.88	0.81	**0.89**
D2b	0.91	0.35	0.27	0.33	0.36	0.88	0.85	0.81	**0.90**
D3a	0.74	0.94	0.93	0.33	0.84	0.84	0.89	0.54	**0.92**
D3b	0.7	0.89	0.88	0.33	0.65	0.6	0.71	0.46	**0.84**
D4a	0.71	0.85	0.81	0.33	0.53	0.58	0.83	0.57	**0.85**
D4b	0.62	0.71	0.81	0.33	0.44	0.56	0.72	0.58	**0.82**
D5a	0.88	0.94	0.95	0.33	0.61	0.72	0.9	0.42	**0.93**
D5b	0.9	0.91	0.93	0.33	0.48	0.67	0.89	0.51	**0.96**
D6	0.61	0.72	0.65	0.5	0.67	0.66	0.69	0.51	**0.68**
D7	0.83	0.69	0.67	0.5	0.5	0.69	0.71	0.64	**0.84**
D8	0.98	0.67	0.74	0.54	0.72	0.99	0.62	0.54	**1.00**
D9	0.61	0.72	0.65	0.54	0.5	0.66	0.69	0.51	**0.67**
D10	0.9	0.81	0.88	0.79	0.7	0.88	0.92	0.9	**0.92**
D11	0.77	0.64	0.64	0.59	0.69	0.62	0.64	0.64	**0.75**
D12	0.76	0.81	0.82	0.5	0.81	0.79	0.8	0.77	**0.81**
D13	0.67	0.67	0.78	0.33	0.8	0.8	0.74	0.8	**0.95**

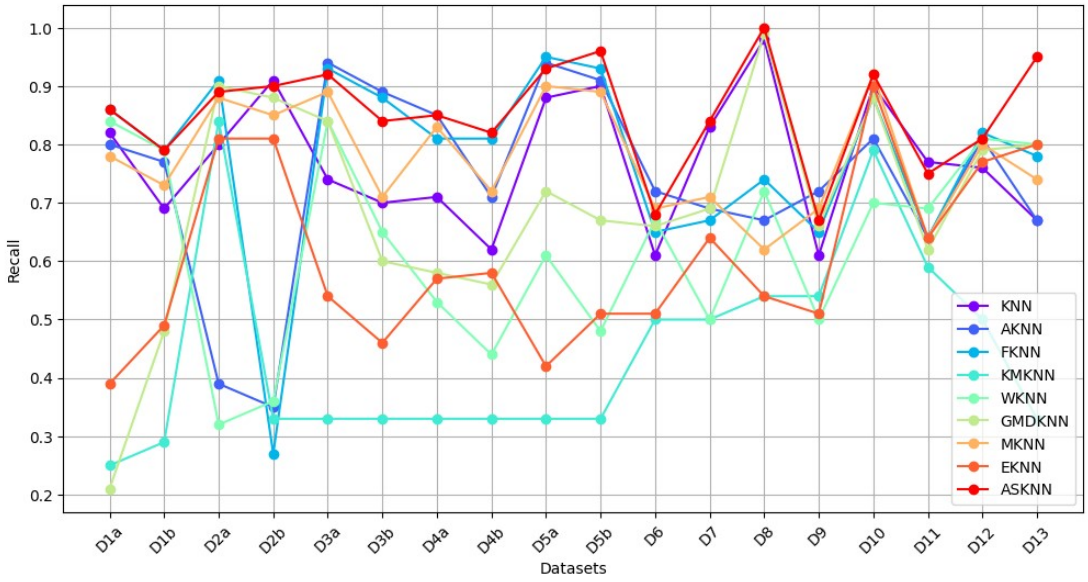


Figure 6.13: Recall comparison of different KNN variants across multiple datasets.

6.3.2.4 F1-score

The F1 score, a crucial performance metric that balances recall and precision, offers a more accurate evaluation of a model's capacity to classify both favourable and unfavourable circumstances. A high F1 score indicates that the model is performing well in terms of correctly identifying positive events and preventing false positives. The F1 scores of many KNN-based models, including ASKNN, for identifying anxiety and depression across several datasets, are displayed in Table 6.16.

As seen from Fig. 6.14, ASKNN exhibits competitive performance with F1 scores usually being at par with or even higher than other variants of KNN. For instance, in dataset D1a (depression detection), ASKNN is at an F1 score of 0.88, narrowly behind FKNN (0.88), but way ahead of KMKNN (0.16), which fails to accurately classify the data. The same pattern is observed in D1b, where the F1 score of ASKNN is 0.82, which is greater than that of KNN (0.71) and KMKNN (0.09).

Table 6.16: F1 scores of different KNN-based models on different datasets.

Dataset	KNN	A KNN	FKNN	KMKNN	WKNN	GMDKNN	MKNN	EKNN	ASKNN
D1a	0.85	0.84	0.88	0.16	0.86	0.16	0.82	0.38	**0.88**
D1b	0.71	0.8	0.81	0.09	0.81	0.46	0.76	0.51	**0.82**
D2a	0.8	0.39	0.9	0.81	0.3	0.91	0.88	0.86	**0.89**
D2b	0.91	0.32	0.26	0.26	0.34	0.88	0.85	0.86	**0.89**
D3a	0.78	0.95	0.95	0.16	0.86	0.88	0.92	0.52	**0.94**
D3b	0.75	0.9	0.88	0.15	0.66	0.59	0.75	0.48	**0.87**
D4a	0.72	0.87	0.82	0.23	0.53	0.6	0.83	0.53	**0.87**
D4b	0.64	0.74	0.81	0.2	0.44	0.59	0.74	0.54	**0.82**
D5a	0.91	0.95	0.93	0.29	0.67	0.77	0.93	0.43	**0.95**
D5b	0.93	0.94	0.94	0.29	0.53	0.73	0.92	0.55	**0.97**
D6	0.61	0.72	0.64	0.4	0.67	0.66	0.7	0.41	**0.68**
D7	0.82	0.69	0.67	0.35	0.35	0.69	0.7	0.62	**0.84**
D8	0.97	0.65	0.71	0.53	0.7	0.99	0.61	0.54	**1.00**
D9	0.61	0.73	0.64	0.53	0.4	0.66	0.7	0.41	**0.68**
D10	0.91	0.83	0.89	0.81	0.72	0.89	0.92	0.91	**0.93**
D11	0.79	0.65	0.65	0.44	0.73	0.64	0.65	0.65	**0.79**
D12	0.75	0.81	0.81	0.35	0.81	0.79	0.8	0.77	**0.81**
D13	0.67	0.67	0.78	0.14	0.79	0.79	0.73	0.78	**0.94**

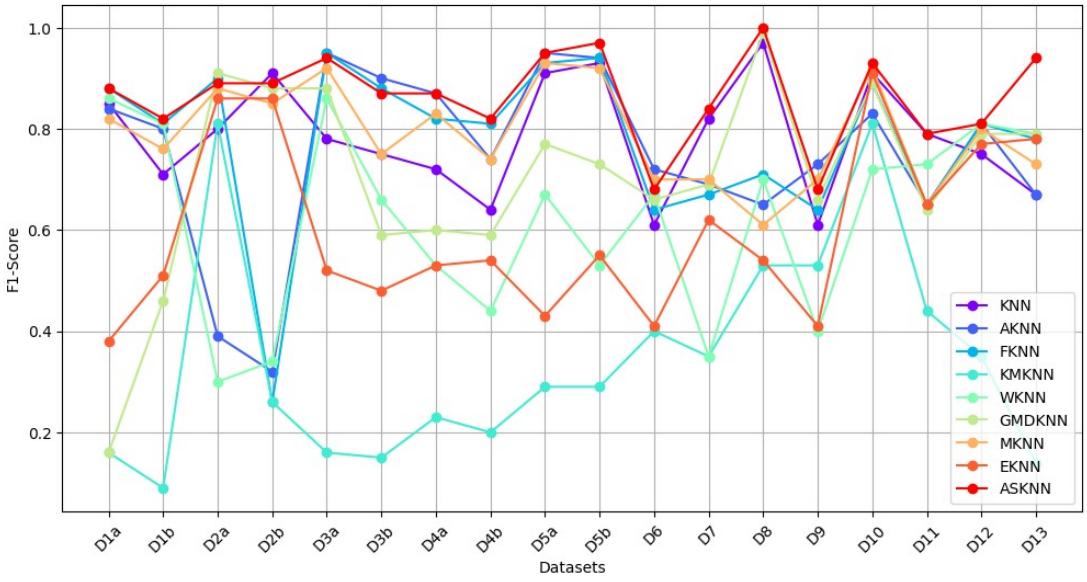


Figure 6.14: F1-Score comparison of different KNN variants across multiple datasets.

ASKNN has the highest F1 of 0.94 in the D3a dataset for anxiety detection, clearly showing how it performs strongly within this dataset. Other algorithms like KNN (0.78) and FKNN

(0.95) are also very strong, yet the performance of ASKNN is superior. Similarly, within the D5a dataset, ASKNN obtains a 0.95 score, which beats regular KNN (0.91) as well as models like WKNN (0.67).

In general, the findings show that ASKNN is a robust and high-performing model in terms of F1 score on many datasets. It shows that ASKNN can effectively trade off precision with recall, rendering it a good candidate for detecting depression and anxiety.

6.3.2.5 5-Fold Cross Validation

It can be seen from Fig. 6.15; Table 6.17 that the ASKNN is constantly attaining competitive accuracy values for the majority of datasets. For example, in D5a (0.9699 ± 0.0054), D5b (0.9639 ± 0.0035), and D3a (0.9649 ± 0.0046) datasets, ASKNN is exceptionally high in terms of accuracy when compared to other forms of KNN models. Specifically, D8 exhibits an exceptionally high value of 0.9969 ± 0.0065 for ASKNN, indicating its exceptional performance on specific types of data.

Contrary to this, KMKNN traditional models reflect lower accuracy values, with a majority of the datasets registering figures in the range of 20–40% (e.g., D1a, D2a, D3b), pointing out the inadequacies of using this method with the data under consideration. On similar lines, WKNN, though better than KMKNN in certain cases, still lags behind ASKNN in datasets such as D2a, D5a, and D5b. The usefulness of ASKNN in identifying patterns in the data is demonstrated by the fact that FKNN and MKNN produce moderate accuracy in select datasets but fall short of ASKNN in the majority of datasets. The EKNN also shows good performance in certain datasets (D5b, D6), but still lags behind ASKNN in other datasets.

In conclusion, the findings indicate that ASKNN not only yields greater accuracy than KNN in its traditional and advanced forms, but also offers strong performance in a variety of datasets, and thus, it is a promising method for detecting depression and anxiety in the context of hand.

Table 6.17: 5-fold cross-validation mean accuracy scores and standard deviation for various KNN-based models.

Dataset	**AKNN**	**FKNN**	**KMKNN**	**WKNN**	**GKNN**	**MKNN**	**EKNN**	**ASKNN**
D1a	0.8531±0.0243	0.8647±0.0265	0.2042±0.0572	0.8490±0.0143	0.2200±0.0212	0.8118±0.0173	0.5934±0.0124	0.8602±0.0286
D1b	0.8192±0.0252	0.8317±0.0190	0.1926±0.1078	0.8421±0.0193	0.5900±0.0212	0.7572±0.0174	0.6037±0.0215	0.8325±0.0092
D2a	0.5640±0.0546	0.9177±0.0081	0.8065±0.0867	0.4500±0.0274	0.6029±0.0302	0.8908±0.0386	0.8621±0.0222	0.8976±0.0151
D2b	0.6108±0.0075	0.4610±0.0251	0.6235±0.0046	0.5700±0.0265	0.6260±0.0205	0.9404±0.0269	0.8828±0.0212	0.9362±0.0116
D3a	0.9649±0.0046	0.9649±0.0045	0.5460±0.1316	0.2102±0.0792	0.9366±0.0090	0.9513±0.0070	0.7982±0.0093	0.9586±0.0071
D3b	0.9524±0.0054	0.9527±0.0071	0.6202±0.1649	0.2180±0.2550	0.9025±0.0082	0.9431±0.0028	0.8198±0.0142	0.9535±0.0077
D4a	0.7954±0.0520	0.7885±0.0291	0.2483±0.1571	0.2600±0.0228	0.6109±0.0646	0.7379±0.0493	0.6782±0.0335	0.8027±0.0366
D4b	0.7655±0.0336	0.7747±0.0203	0.3822±0.1414	0.2758±0.0953	0.6644±0.0662	0.7632±0.0481	0.7241±0.0304	0.7728±0.0583
D5a	0.9699±0.0054	0.9710±0.0036	0.7838±0.0092	0.3055±0.2636	0.9064±0.0080	0.9617±0.0025	0.8201±0.0090	0.9679±0.0058
D5b	0.9639±0.0035	0.9759±0.0041	0.7570±0.0146	0.3178±0.2562	0.8982±0.0091	0.9571±0.0072	0.7982±0.0093	0.9726±0.0071
D6	0.7292±0.0134	0.6836±0.0182	0.3913±0.1176	0.7266±0.0206	0.7164±0.0219	0.7226±0.0364	0.6575±0.0095	0.7521±0.0353
D7	0.6500±0.0624	0.6568±0.0570	0.5445±0.0444	0.6532±0.0421	0.6170±0.0356	0.6631±0.0562	0.6567±0.0344	0.7721±0.0371
D8	0.6725±0.0379	0.7450±0.0386	0.6125±0.0307	0.8940±0.1223	0.9925±0.0050	0.6575±0.0236	0.6400±0.0313	0.9969±0.0065

Continued on next page

Table 6.17 continued from previous page								
Dataset	**AKNN**	**FKNN**	**KMKNN**	**WKNN**	**GKNN**	**MKNN**	**EKNN**	**ASKNN**
D9	0.7292±0.0134	0.6836±0.0182	0.3913±0.1176	0.6944±0.0302	0.7164±0.0219	0.7226±0.0364	0.6575±0.0095	0.7317±0.0175
D10	0.8927±0.0359	0.8769±0.0237	0.8646±0.0221	0.8692±0.0355	0.8806±0.0232	0.8734±0.0143	0.8788±0.0340	0.8857±0.0451
D11	0.8688±0.0671	0.8498±0.0527	0.2061±0.1224	0.8585±0.0638	0.8771±0.0664	0.8675±0.0597	0.8675±0.0576	0.8331±0.0441
D12	0.7952±0.0110	0.7988±0.0131	0.4843±0.0406	0.7657±0.0219	0.7962±0.0437	0.7867±0.0389	0.7350±0.0458	0.7902±0.0122
D13	0.7025±0.0592	0.7364±0.0621	0.3314±0.0411	0.6636±0.0827	0.7237±0.0836	0.6802±0.0641	0.6462±0.0361	0.7902±0.0122

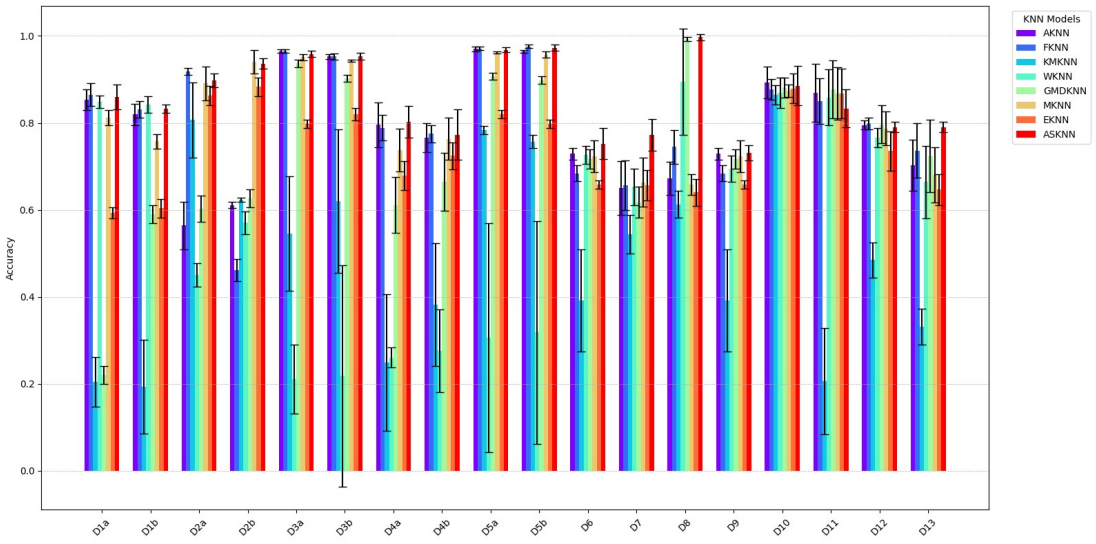


Figure 6.15: CV score comparison of different KNN variants across multiple datasets.

6.3.3 Performance of the ASKNN model on Primary Dataset

The ASKNN model's performance results for classifying mental health on the BDI and BAI datasets are shown in this section. The distribution of successfully and mistakenly categorised cases is displayed in confusion matrices for both datasets, providing insight into the model's classification performance. Additionally, AUC–ROC curves are provided to assess the model's discriminative power for various classes. These results show how effective ASKNN is at accurately identifying and differentiating mental diseases. These results show how effective ASKNN is in accurately recognising and differentiating mental diseases.

6.3.3.1 Depression

The classification performance of the ASKNN model on different degrees of depression severity was evaluated by comparing its accuracy with the BDI dataset using a confusion matrix and AUC–ROC metrics. The confusion matrix indicates that ASKNN had high accuracy for Class 0 (No Depression), identifying 153 cases correctly, with only a single misclassification. Nonetheless, 14 of Class 1 (Mild Depression) were wrongly classified as Class 0, reflecting a bit of overlap between the two groups. The model was equally good for Class 2 (Moderate Depression) and Class 3 (Severe Depression), but there were 6 misclassifications of Class 2 cases to Class 1 and 4 misclassifications of Class 3 cases into Class 2, as illustrated in Fig. 6.16 The above misclassifications indicate that while the model excellently distinguishes between clearly disparate depression levels, some cases around the borders still prove tricky.

To better evaluate the performance of ASKNN, AUC–ROC values were calculated for every class. The model performed an AUC–ROC of 0.95 for Class 0, 0.89 for Class 1, 0.92 for Class 2, and 0.91 for Class 3, as can be seen in Fig. 6.17. All these values indicate that ASKNN performs excellently in separating Class 0 (No Depression) and Class 2 (Moderate Depression) from others. The large AUC scores greater than 0.90 for most classes show excellent model performance, especially in identifying cases of moderate and severe depression. Class 1 (Mild Depression) reported a lower AUC–ROC (0.89), implying some challenges in distinguishing between mild depression and other classes.

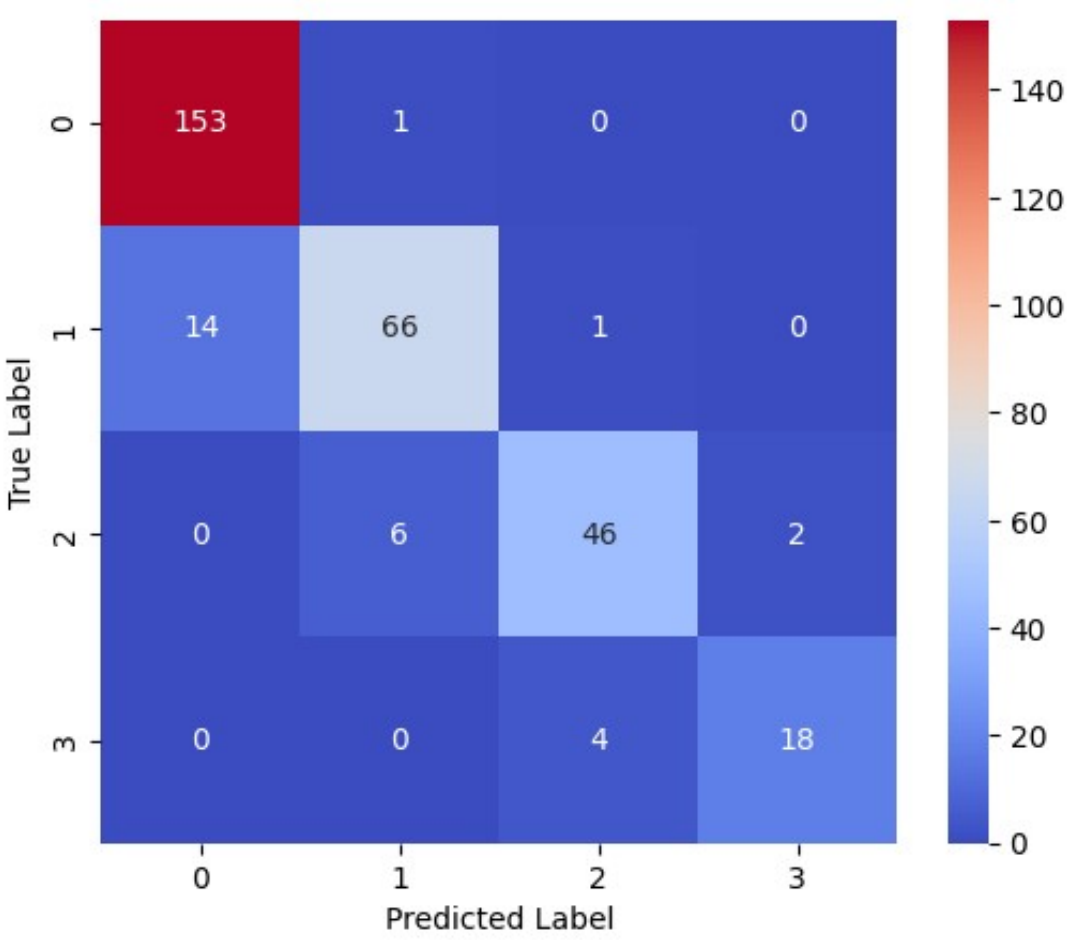


Figure 6.16: Confusion matrix for depression detection employing ASKNN

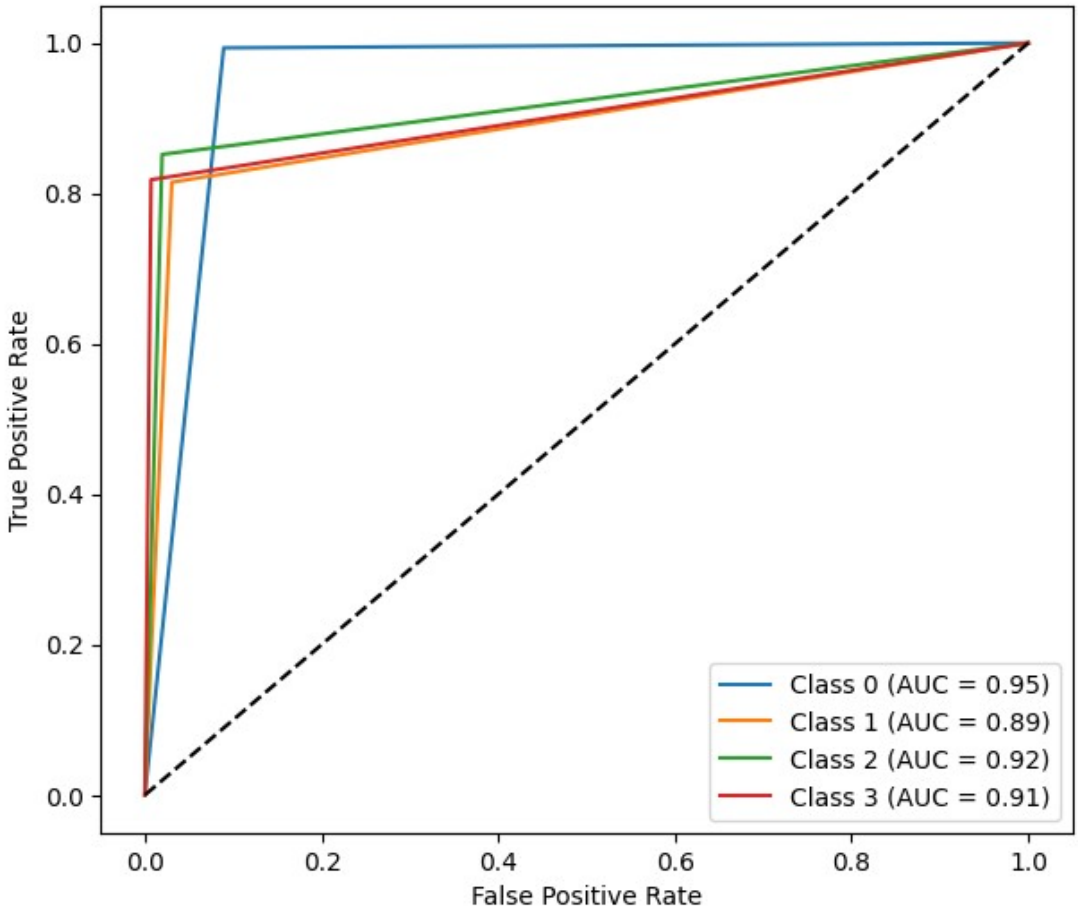


Figure 6.17: AUC-ROC curve for depression detection employing ASKNN

6.3.3.2 Anxiety

The BAI dataset was used to test the ASKNN model's performance, and the confusion matrix and AUC–ROC scores were used to gauge the model's classification accuracy. We can determine how well the model is identifying different anxiety intensity levels by looking at the confusion matrix. ASKNN showed high accuracy for Class 0 (No Anxiety), classifying 127 cases correctly with only one misclassification. But Class 1 (Mild Anxiety) showed some overlap with Class 0 since 20 cases were misclassified. Likewise, Class 2 (Moderate Anxiety) and Class 3 (Severe Anxiety) showed good classification performance, although 9 cases from Class 3 were misclassified as Class 2, as indicated in Fig. 6.18. This shows that although the model can distinguish between moderate and severe cases, some borderline cases can still be challenging.

To further assess ASKNN's performance in classification, AUC–ROC values were calculated for every class. The model obtained an AUC–ROC of 0.93 for Class 0, 0.78 for Class 1, 0.90 for Class 2, and 0.86 for Class 3. The results show that ASKNN works amazingly well to separate Class 0 (No Anxiety) and Class 2 (Moderate Anxiety) from the rest of the

categories, as evidenced in Fig. 6.19 Class 1 (Mild Anxiety) had the lowest AUC-ROC of 0.78, which points to some confusion in separating cases of mild anxiety from non-anxiety individuals. Despite this, the overall performance is satisfactory, especially for classifying moderate and severe cases of anxiety.

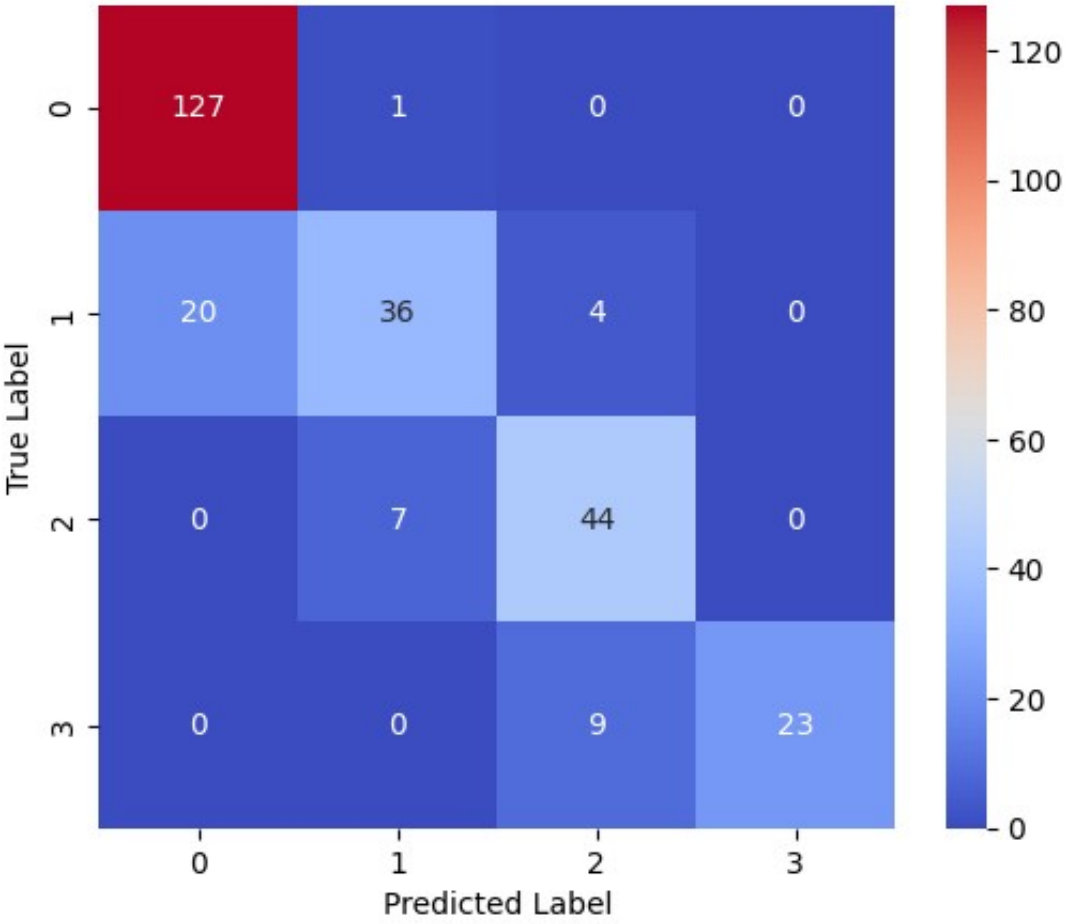


Figure 6.18: Confusion matrix for anxiety detection employing ASKNN

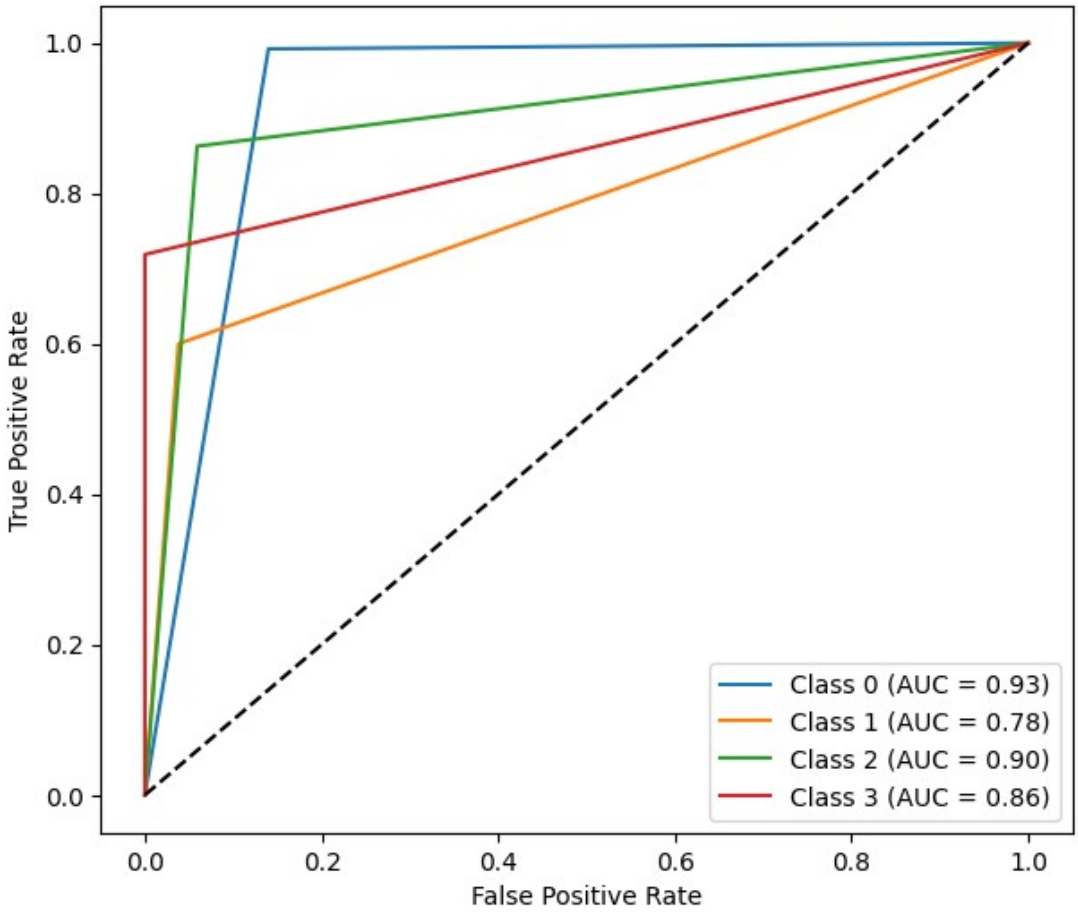


Figure 6.19: AUC-ROC curve for anxiety detection employing ASKNN

These findings stress that ASKNN is an effective model for both depression and anxiety detection, especially for cases of moderate to severe intensity, and offers substantial improvement over the traditional distance-based classifier.

Although the ASKNN model shows robust classification performance, especially for severe and moderate classes, misclassifying borderline cases—e.g., mild anxiety (Class 1) as no anxiety (Class 0), or moderate anxiety (Class 2) as severe (Class 3)—has significant practical consequences. Under/over-estimation of symptom severity in a clinical or screening setting might be the consequence of these errors. For example, false negatives at early-stage (mild) cases can postpone treatment, and false positives at more severe cases can result in over-referral or treatment load. This serves to emphasize the importance of being wary when classifying near class boundaries and supports the usefulness of employing ASKNN as a pre-screening device rather than a sole diagnostic tool. Future applications could gain from combining clinical judgement or threshold-based uncertainty indicators to better control such edge cases.

6.3.4 Statistical Analysis and Significance Testing

To examine the statistical significance of performance variation among models, we ran a suite of non-parametric tests. Overall differences among models were searched using the Friedman test, while Wilcoxon signed-rank tests were run for pairwise tests between ASKNN and all other KNN versions on four main evaluation criteria. Cohen's d and Cliff's Delta were used to measure the effect size in performance variation.

We formulated the following hypotheses:

- **Null Hypothesis (H_0):** There is not much difference in the performance of ASKNN from the other KNN-based models in terms of different evaluation metrics (accuracy, precision, recall, and F1-score). ASKNN performs as well as the other models.
- **Alternative Hypothesis (H_1):** ASKNN demonstrates statistically significant improvement in performance over other KNN-based models across evaluation metrics such as accuracy, precision, recall, and F1-score.

6.3.4.1 Friedman Test Analysis

To identify if there are differences in performance between models, we performed the Friedman test for Recall, Accuracy, Precision, and F1-score. The results of the test, presented in Table 6.18, demonstrate statistically significant variations between models ($p < 0.05$) for all four measures. This implies that there is at least one model that significantly outperforms or underperforms the others.

Table 6.18: Friedman Test Results for Model Comparison

Metric	**Friedman Test Statistic (χ^2)**	**p-value**	**Significance**
Accuracy	67.78	< 0.0001	Significant
Precision	58	< 0.0001	Significant
Recall	68.88	< 0.0001	Significant
F1-score	73.06	< 0.0001	Significant

These findings verify that model performance variation is statistically significant, necessitating further pairwise comparison to determine which models vary most.

6.3.4.2 Wilcoxon Signed-Rank Test Results

We used Wilcoxon signed-rank tests to verify the pairwise significance of ASKNN in contrast to other models. The p-values obtained are presented in Table 6.19.

Table 6.19: Wilcoxon Signed-Rank Test (p-values) Comparing ASKNN with Other Models

Model Comparison	**Accuracy**	**Precision**	**Recall**	**F1-score**
ASKNN vs KNN	0.00043	0.00083	0.0001	0.00041
ASKNN vs A_KNN	0.01421	0.03537	0.01207	0.01346
ASKNN vs FKNN	0.0015	0.00213	0.02439	0.00373
ASKNN vs KMKNN	0.00001	< 0.0001	< 0.0001	< 0.0001
ASKNN vs WKNN	0.00043	0.00029	0.00043	0.00029
ASKNN vs GMDKNN	0.00007	0.00071	0.00002	0.00007
ASKNN vs MKNN	0.00104	0.01568	0.0007	0.00041
ASKNN vs EKNN	< 0.0001	0.0028	< 0.0001	< 0.0001

The Wilcoxon signed-rank test results support that ASKNN outperforms all the other KNN-based models in Recall, Accuracy, Precision, and F1-score, with p-values all being less than 0.05. In particular, the comparison with KMKNN and EKNN produced very low p-values (< 0.0001), which suggest a significant difference in performance as illustrated in Fig. 6.20. These results, combined with those of the According to the Friedman test, ASKNN performs better in classification within the parameters of the examined datasets. We reject the null hypothesis (H_0) and conclude that ASKNN considerably outperforms the comparative KNN-based models in this work because the majority of the pairwise comparisons show statistically significant differences. However, these findings are limited to the datasets and evaluation settings used, and performance may vary across other data domains or tasks.

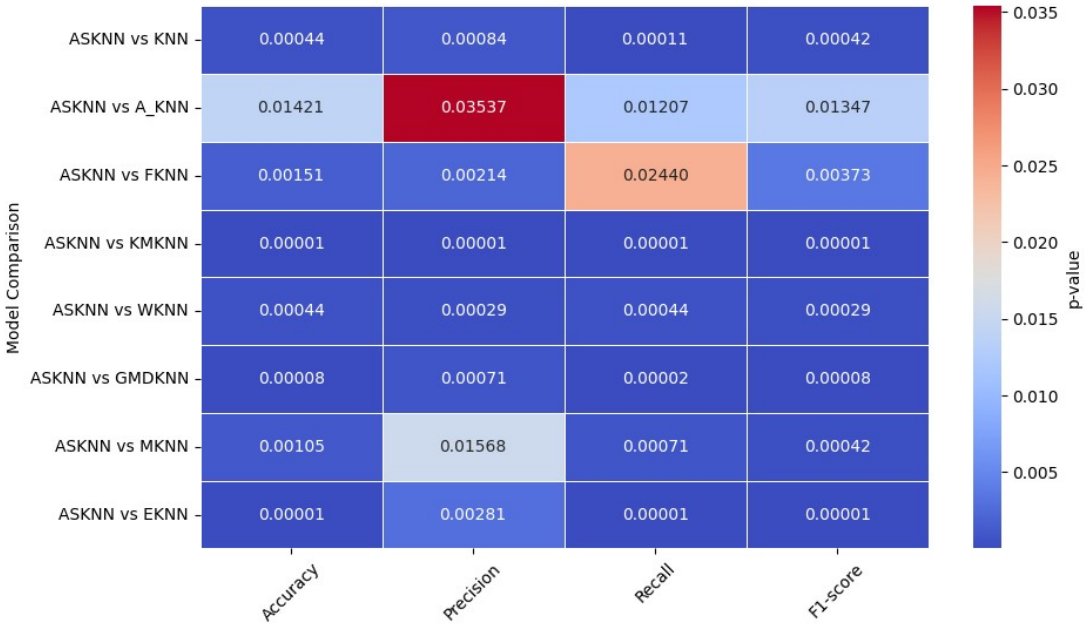


Figure 6.20: Wilcoxon Signed-Rank Test (p-values) Comparing ASKNN with Other Models

6.3.4.3 Effect Size Analysis

Although statistical significance (p-values) establishes whether differences do or do not exist, measures of effect size like Cohen's d and Cliff's Delta (Δ) estimate the size of these differences. High effect sizes show that not only are the improvements in ASKNN's performance statistically significant, but also that they are of practical significance. Table 6.20 shows the effect size values for Accuracy, Precision, Recall, and F1-score in comparing ASKNN with other KNN-based algorithms as shown in Fig. 6.21.

Table 6.20: Effect Size Analysis

Model	**Accuracy (Cohen's d, Cliff's Δ)**	**Precision (Cohen's d, Cliff's Δ)**	**Recall (Cohen's d, Cliff's Δ)**	**F1-score (Cohen's d, Cliff's Δ)**
KNN	0.651, 0.327	0.630, 0.340	0.829, 0.432	0.816, 0.429
AKNN	0.855, 0.423	0.848, 0.392	0.940, 0.469	0.918, 0.432
FKNN	0.700, 0.361	0.742, 0.392	0.635, 0.324	0.716, 0.361
KMKNN	2.571, 0.926	3.016, 0.907	3.236, 0.935	3.544, 0.957
WKNN	1.376, 0.667	1.575, 0.790	1.841, 0.799	1.814, 0.818
GMDKNN	0.881, 0.457	1.080, 0.552	1.187, 0.583	1.166, 0.568
MKNN	0.663, 0.364	0.633, 0.340	0.850, 0.438	0.823, 0.426
EKNN	1.423, 0.682	1.377, 0.599	2.112, 0.830	2.130, 0.815

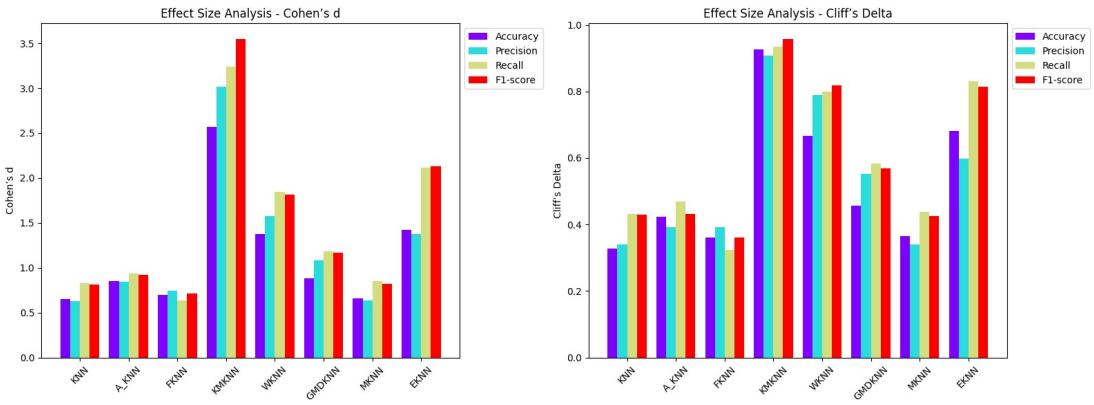


Figure 6.21: An effect size analysis depicting Cohen's d for KNN variants with ASKNN, and b An effect size analysis depicting Cliff's delta for KNN variants with ASKNN

6.4 Discussion

The findings of this work illustrate the efficacy of the proposed ASKNN in its ability to identify depression and anxiety. On several measures of evaluation, compared to other traditional KNN-based variations taken into consideration in this work, ASKNN fared better, as illustrated in Fig. 6.22. The Wilcoxon signed-rank test findings indeed validated that the improvements observed were statistically significant with persistently low p-values ($p < 0.05$), reflecting a real difference in performance under the current experimental conditions. Furthermore, effect size estimation via Cohen's d and Cliff's Delta also supported ASKNN, showing practically significant improvements relative to the compared models. However, these results should be interpreted in the context of the specific datasets and models used, and generalization to other domains or data types should be approached with caution.

The analysis of effect size indicates that ASKNN performs better than the standard KNN-based models evaluated in this work. Interestingly, it is seen that KMKNN and EKNN also perform well but were consistently outperformed by ASKNN on the selected datasets. WKNN and GMDKNN show moderate performance, whereas standard KNN, AKNN, and FKNN show relatively lower effect sizes. These results imply that, in the current experimental setting, ASKNN may improve classification accuracy for mental health detection.

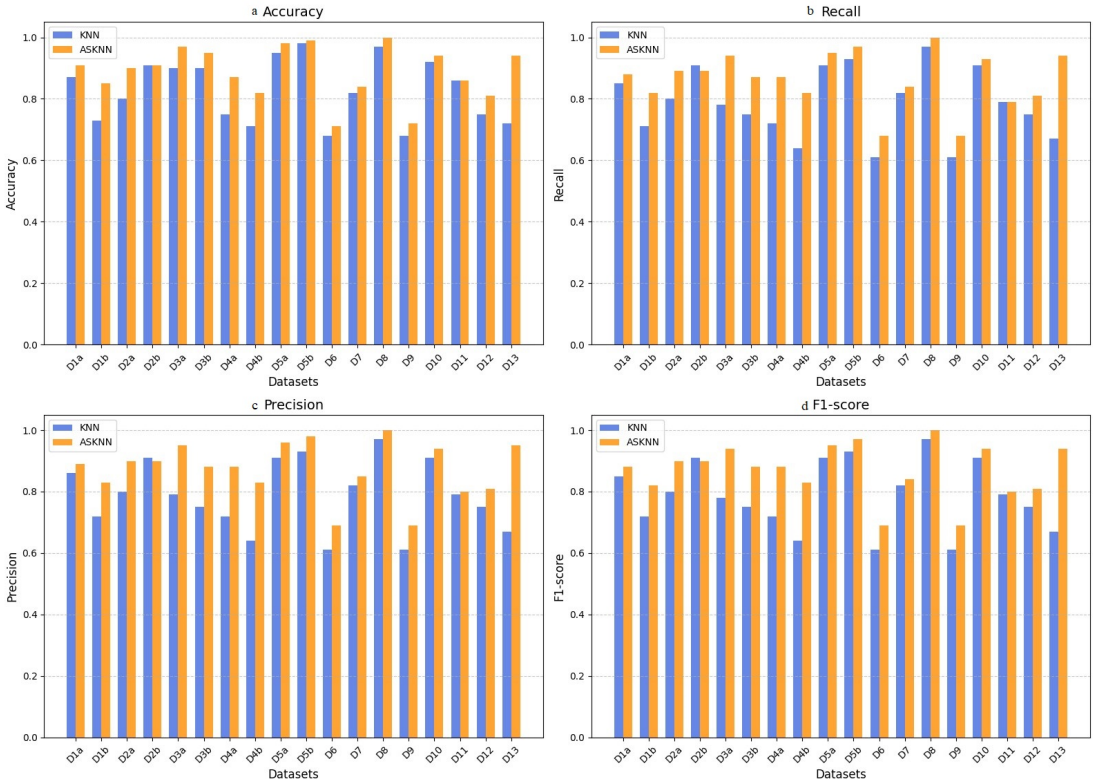


Figure 6.22: Comparison between the performance of KNN and ASKNN for various datasets a Accuracy b Recall c Precision d F1 score

Although ASKNN showed good predictive performance, one should note its computational intensity. Benchmarking outcomes indicated that ASKNN expended significantly greater inference time and memory than less complex models such as basic KNN, FKNN, and MKNN, according to Table 6.21. The increase in the computational burden comes from the adaptive weighting strategy in the model according to local density and estimation of sigma, which adds extra computation per query. In resource-limited or real-time usage, e.g., in mobile mental health screening applications or large-scale clinical deployments, these costs could be practical barriers. Conversely, models such as MKNN and KMKNN achieved a superior speed-accuracy trade-off in some datasets. Thus, although ASKNN delivers high-performance improvement, its deployment needs to be cautiously considered in terms of existing computational resources and application scale. The performance across datasets also varies, indicating that no model is optimal in all environments, and that model choice should be based on both performance and resource considerations.

Table 6.21: Comparative theoretical overview of computational details for different KNN-based models

Model	Training Time	Inference Time	Memory Requirement	Scalability	Notes
KNN	Low	Medium	Moderate to High	Moderate	Baseline; simple majority voting
AKNN	Low	Low to Medium	Moderate	Moderate	Adaptively selects k per instance
FKNN	Low	High	High	Low	Fuzzy membership increases computation
KMKNN	Medium	Medium	High	Low	Involves clustering per class
WAKNN	Low	Very High	High	Very Low	Inverse log weighting adds significant inference time
GMDKNN	Low	High	Moderate to High	Low	The custom distance function is computationally heavy
MKNN	Medium	Medium	Moderate	Moderate	Filters training points using mutual neighbors
EKNN	High	High	Very High	Very Low	Multiple KNNs with different parameters aggregated
ASKNN	Low	Very High	High	Low	Adaptive local density & Gaussian weights impact runtime

6.5 Summary

A novel ASKNN model for the multi-class classification of anxiety and depression in the post-COVID-19 setting was provided in this chapter. The proposed approach extends traditional KNN by dynamically adjusting neighborhood influence through an adaptive sigma parameter, enabling improved handling of data heterogeneity and class variability commonly observed in mental health datasets.

Extensive experimental evaluation across multiple datasets demonstrated that ASKNN consistently outperformed conventional KNN and its variants in terms of accuracy, F1-score, and AUC–ROC. The observed performance gains were statistically validated using the Wilcoxon signed-rank test, and effect size analysis further demonstrated the usefulness of the suggested approach. The results indicate that adaptive neighborhood selection plays a critical role in enhancing classification reliability for psychological assessment tasks.

ASKNN's enhanced predictive performance illustrates a useful trade-off between accuracy and economy, especially in applications where diagnostic reliability is prioritised, even though its adaptive weighting technique results in greater processing costs.

Even while ASKNN's adaptive weighting technique raises processing costs, its improved predictive performance shows a helpful trade-off between accuracy and economy, particularly in applications where diagnostic reliability is valued.

Chapter 7

Feature-Weighted Adaptive Kernel KNN: A Novel Approach for Post-COVID-19 Depression and Anxiety Detection

7.1 Introduction

Earlier chapters of this thesis have covered the COVID-19 pandemic's effects on mental health worldwide in great detail, emphasising a notable increase in anxiety and depression in a variety of demographics, especially among young adults and students. Building upon this foundation, the focus now shifts from problem understanding to the development of more effective and interpretable computational solutions for early detection.

As established in section 2.5 and subsequent discussions, psychometric instruments such as the BDI and BAI remain standard tools for assessing mental health conditions. Furthermore, Chapters 4 through 6 explored the application of various ML techniques, including SVM, RF, and hybrid models for improving predictive performance in mental health classification tasks. While these approaches demonstrated promising results, limitations related to interpretability, adaptability, and feature sensitivity were consistently observed.

Among the existing methods, the KNN algorithm has been recognized for its simplicity, transparency, and suitability for deployment in low-resource and clinical settings. However, as critically analyzed in the previous chapter, traditional KNN suffers from two fundamental

limitations: (i) equal weighting of all features during similarity computation, and (ii) reliance on fixed or simplistic neighbor voting strategies. These assumptions are particularly restrictive when applied to psychometric datasets, where questionnaire items differ significantly in their relevance and may introduce noise or redundancy.

Motivated by these limitations, this chapter proposes an enhanced variant of KNN, termed **Feature-Weighted Adaptive Kernel KNN (FW-AK-KNN)**. The proposed model introduces two key modifications:

- Incorporation of feature importance into the distance computation process
- Utilization of locally adaptive kernel-based weighting for neighbor contributions

These improvements seek to preserve the KNN framework's intrinsic simplicity while simultaneously increasing the model's predicted accuracy and interpretability.

In addition to contributing to the larger goal of creating scalable, data-driven screening systems, which is in line with global healthcare priorities like SDG 3 (Good Health and Well-Being), this work represents a step toward more resilient and flexible machine learning solutions for mental health prediction. The suggested method also advances the more general goal of creating scalable, data-driven screening systems, which is in line with international healthcare targets like SDG 3 (Good Health and Well-Being).

7.2 Proposed Methodology

By altering two crucial elements, the suggested FW-AK-KNN model is a hybrid classification framework intended to improve the conventional KNN method:

- Distance computation
- Neighbor voting mechanism

We compared the proposed FW-AK-KNN with the standard KNN algorithm to assess the impact of:

- Introducing feature-based weighting in distance computation
- Using locally adaptive kernel functions instead of uniform or fixed-distance voting

7.2.1 Datasets used

To compare the performance of the introduced FW-AK-KNN model, experiments were carried out on several benchmark datasets with a specific interest in the mental health screening task using depression and anxiety screening scales. Other benchmark biomedical datasets, such as heart disease, diabetes, and mammographic, were also utilized to assess generalizability as indicated in Table 7.1. All feature sets were normalized before training models.

Tables 6.6 and 6.7 present anonymized illustrative snapshots of the primary depression and anxiety datasets (D1a and D1b), respectively, showing representative demographic attributes, questionnaire responses, derived scores, and corresponding classes used in ML models in this work. Detailed characteristics and anonymized illustrative snapshots of other datasets are provided in Chapter 3 and 5 (Table 3.2, Table 3.3, Table 3.4 and Table 5.3).

Table 7.1: Summary of datasets used in the evaluation of FW-AK-KNN

S. No.	Reference	Dataset Name	#Features	Dataset Size
D1a	[12]	Depression	21	N=1552
D1b	[12]	Anxiety	21	N=1356
D2a	[255]	Depression	13	N=5108
D2b	[255]	Anxiety	13	N=5108
D3a	[188]	Depression	9	N=3569
D3b	[188]	Anxiety	7	N=3569
D4a	[38]	Depression	14	N=435
D4b	[38]	Anxiety	14	N=435
D5	[104]	Heart Attack Possibilities	13	N=303
D6	[66]	Mammographic	5	N=802
D7	[148]	Diabetes	8	N=768
D8	[4]	Wine	13	N=178

7.2.2 Mental health assessment

To determine the mental health condition of participants, we made use of standardized psychological questionnaires from a range of datasets, namely the BDI, BAI, GAD-7, PHQ-9, and SCL-90 as explained in detail in section 2.5.

7.2.3 ML Algorithms

In this chapter, KNN was employed as part of the proposed framework for depression and anxiety prediction. The theoretical foundations, working principles, advantages, and limitations of this model have already been discussed in detail in Section 2.8.1.3.

7.2.4 Overview of the Proposed Model

FW-AK-KNN integrates feature weighting and adaptive kernel functions into the KNN framework. The model operates in three main stages:

- Feature importance extraction
- Weighted distance computation
- Kernel-based adaptive voting

The model can more accurately represent the underlying structure of biological and psychometric datasets thanks to this architecture.

7.2.4.1 Feature Weighting Using RF

In conventional KNN, all features are treated equally, as depicted in Figure 7.1a. However, in real-world datasets, especially psychometric scales, certain features have higher predictive significance.

To address this, an RF model is trained on the dataset to compute feature importance scores, as depicted in Figure 7.1b. These scores are normalized and incorporated into the distance metric.

The feature-weighted Euclidean distance is defined as:

$$d(x,x') = \sum_i w_i (x_i - x'_i)^2 \tag{7.1}$$

where w_i represents the normalized importance weight of the i^{th} feature.

This modification ensures that more relevant features contribute more significantly to similarity calculations.

7.2.4.2 Locally Adaptive Kernel Function

Instead of assigning equal or inverse-distance weights to neighbors, the proposed model uses kernel functions (e.g., Gaussian, Epanechnikov, Triangular) to compute soft weights.

A key innovation is the adaptive computation of kernel bandwidth γ, defined as:

$$\gamma(x) = \frac{(\text{mean distance to k-nearest neighbors})^2 + \varepsilon}{1} \tag{7.2}$$

This enables:

- Smoother decision boundaries in sparse regions
- Sharper boundaries in dense clusters

Thus, the model dynamically adapts to local data distributions.

7.2.4.3 Kernel-Weighted Voting Mechanism

Each neighbor contributes to the final classification based on its kernel weight. The predicted class is determined by aggregating weighted votes across all k-nearest neighbors.

This approach improves:

- Sensitivity to local patterns
- Interpretability of predictions
- Robustness to noise

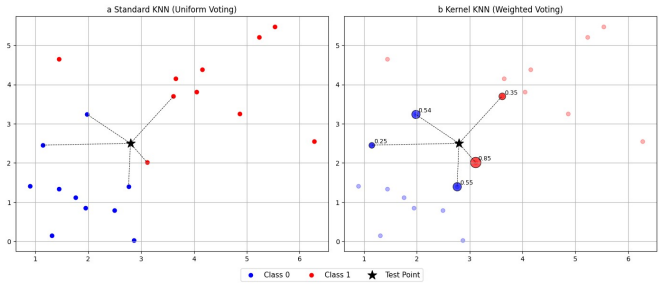


Figure 7.1: Working of KNN with FW-AK-KNN a KNN assigns equal votes b FW-AK-KNN weighs neighbors by distance

7.2.5 Evaluation metrics

Performance was measured with Precision, Recall, Accuracy, and F1-score, computed with stratified train-test splits on 12 datasets. These metrics were defined in detail in Section 2.10. The model was coded in Python with custom code for adaptive kernel weighting and feature-weighted distance. To better understand the impact of these elements on classification performance, ablation experiments were conducted to test against several kernel functions (Gaussian, Epanechnikov, Triangular) and gamma configurations (fixed and adjustable).

7.3 Results

This section presents a comprehensive evaluation of the proposed FW-AK-KNN model in comparison with the traditional KNN algorithm across multiple psychometric and biomedical datasets. The results, summarized in Table 7.2, clearly demonstrate that the proposed model consistently outperforms the baseline KNN in terms of Accuracy, Precision, Recall, and F1-score, thereby validating the effectiveness of integrating feature weighting and adaptive kernel-based neighbor voting into the classification framework.

Table 7.2: Performance comparison of the proposed FW-AK-KNN model

Dataset	Accuracy (KNN)	Accuracy (FW-AK-KNN)	Precision (KNN)	Precision (FW-AK-KNN)	Recall (KNN)	Recall (FW-AK-KNN)	F1-Score (KNN)	F1-Score (FW-AK-KNN)
D1a	0.84	**0.88**	0.86	**0.9**	0.79	**0.82**	0.82	**0.85**
D1b	0.73	**0.86**	0.75	**0.88**	0.68	**0.81**	0.71	**0.85**
D2a	0.95	**0.98**	0.96	**0.98**	0.89	**0.95**	0.92	**0.96**
D2b	0.96	**0.97**	0.94	**0.95**	0.85	**0.9**	0.89	**0.93**
D3a	0.94	**0.98**	0.93	**0.97**	0.85	**0.96**	0.88	**0.97**
D3b	0.93	**0.99**	**0.95**	**0.95**	0.72	**0.99**	0.79	**0.97**
D4a	0.71	**0.97**	0.76	**0.97**	0.72	**0.96**	0.72	**0.96**
D4b	0.72	**0.98**	0.75	**0.98**	0.66	**0.98**	0.68	**0.98**
D5	0.66	**0.69**	0.65	**0.69**	0.64	**0.69**	0.64	**0.69**
D6	0.77	**0.78**	0.78	**0.78**	0.77	**0.78**	0.77	**0.78**
D7	0.67	**0.69**	0.66	**0.67**	0.6	**0.67**	0.6	**0.67**
D8	0.61	**0.81**	0.53	**0.78**	0.56	**0.78**	0.52	**0.78**

Significant performance improvements are observed in the primary psychometric datasets, which form the core focus of this work. For the depression dataset (D1a), the proposed model improves accuracy from 84% to 88%, along with notable gains in precision (0.86 to 0.90), recall (0.79 to 0.82), and F1-score (0.82 to 0.85) as shown in Fig. 7.2. A more pronounced improvement is evident in the anxiety dataset (D1b), where accuracy increases substantially from 73% to 86%, and F1-score improves from 0.71 to 0.85. These results indicate that FW-AK-KNN is particularly effective in handling more challenging datasets characterized by higher variability and noise. Recall improvements in both datasets point to improved ability to detect genuine positive cases, which is particularly important in mental health applications where missed detections can have dire repercussions. Furthermore, the consistent increase in F1-

score reflects a better balance between precision and recall, indicating improved classification of subtle and moderate symptom categories that are often difficult to detect using conventional methods.

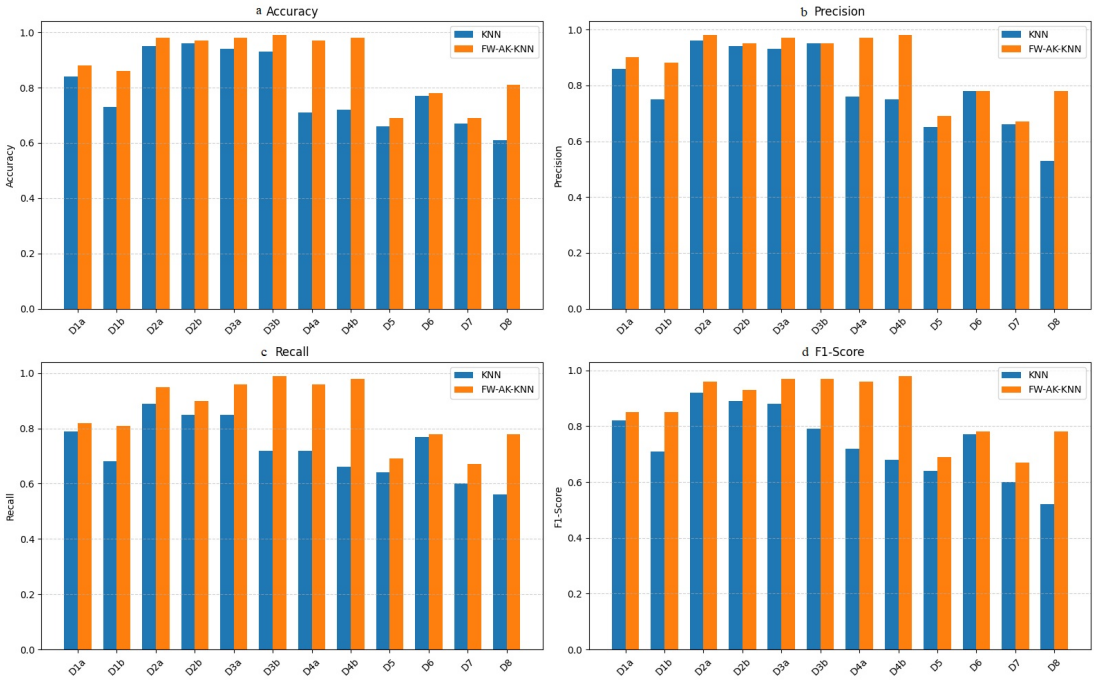


Figure 7.2: Working of KNN with FW-AK-KNN a KNN assigns equal votes b FW-AK-KNN weighs neighbors by distance

The superiority of the proposed model is further reinforced through its performance on additional psychometric datasets (D2a–D4b), where it achieves near-perfect classification in several cases. Accuracy values reach up to 98–99%, with corresponding improvements in F1-scores, demonstrating the robustness of the model across varying dataset sizes and feature distributions. Particularly notable are the substantial improvements in datasets D4a and D4b, where accuracy increases from 0.71 to 0.97 and from 0.72 to 0.98, respectively. These findings demonstrate how feature weighting can effectively emphasise essential characteristics while reducing the impact of noisy or redundant features, improving the model's capacity for discrimination.

In order to assess generalizability, the proposed FW-AK-KNN model was also evaluated on standard biomedical datasets, including heart disease, mammographic, diabetes, and wine datasets. The results indicate moderate improvements in datasets such as heart disease (0.66 to 0.69) and diabetes (0.67 to 0.69), while only marginal gains are observed in the mammographic dataset (0.77 to 0.78). However, a significant improvement is recorded in the wine dataset, where accuracy increases from 0.61 to 0.81 and F1-score from 0.52 to 0.78. This variation

in performance suggests that the proposed model provides the greatest benefit in datasets with complex feature interactions and non-linear class boundaries, where adaptive kernel weighting and feature importance play a more critical role. In contrast, datasets with relatively simpler structures or well-separated classes exhibit smaller performance gains.

Overall, the experimental results reveal several important trends. First, FW-AK-KNN consistently beats the conventional KNN on all datasets, demonstrating the efficacy of the suggested improvements. Second, it demonstrates greater improvements in datasets characterized by feature heterogeneity, noise, and class overlap, which are typical in psychometric data. Third, the increase in recall and F1-score indicates improved detection of minority and subtle classes, which is particularly important for early-stage identification of depression and anxiety. Finally, the model exhibits strong generalizability across different domains, indicating its applicability beyond mental health datasets.

From a practical perspective, even modest improvements in performance metrics are highly significant in healthcare applications. Enhanced recall and F1-score directly translate to better identification of at-risk individuals, reduced misclassification, and improved reliability of automated screening systems. Consequently, the suggested FW-AK-KNN model improves classification performance and helps create more efficient and scalable mental health prediction systems.

7.3.1 Feature Contribution Analysis for Depression Detection using FW-AK-KNN

Using the feature weighting mechanism described in the previous subsection, the proposed FW-AK-KNN model was applied to the BDI dataset to analyze the contribution of individual depressive symptoms. The outcomes are displayed in Table 7.3

The findings suggest that some BDI characteristics are more important in predicting depression. In particular, Features 20, 11, and 8 exhibit the highest importance scores, suggesting that these depressive symptoms play a dominant role as shown in Fig. 7.3. Conversely, Feature 21 and Feature 9 show relatively lower contributions.

This variation highlights that depressive symptoms do not equally influence prediction, and the proposed FW-AK-KNN model effectively captures these differences through adaptive feature weighting.

Furthermore, the ability of the model to identify key contributing BDI items enhances interpretability, making it particularly useful in mental health analysis where understanding

Table 7.3: Feature Importance Scores for BDI-based Depression Detection using FW-AK-KNN

BDI Feature	Importance
20	0.082085
11	0.080046
8	0.076451
14	0.067574
19	0.056134
13	0.056117
15	0.053161
3	0.051810
7	0.050653
17	0.049071
1	0.047382
12	0.044884
10	0.044333
18	0.038972
2	0.036595
4	0.035682
16	0.031676
5	0.028238
6	0.026832
9	0.023002
21	0.019300

symptom-level influence is essential.

This illustrates how the suggested FW-AK-KNN model addresses the interpretability issues with conventional KNN techniques while simultaneously enhancing classification performance and offering significant insights into symptom-level contributions for depression detection.

7.3.2 Feature Contribution Analysis for Anxiety Detection using FW-AK-KNN

To assess each anxiety-related characteristic's contribution, the BAI dataset was subjected to the same feature weighting method. The outcomes are displayed in Table 7.4

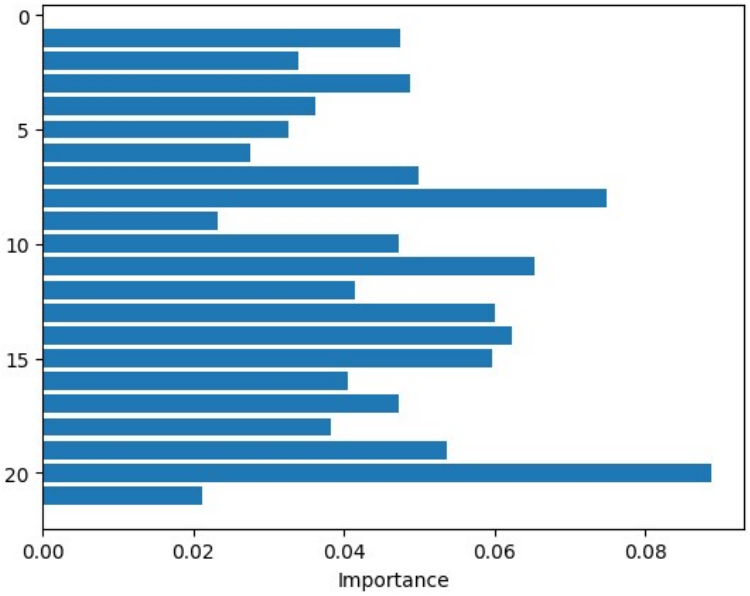


Figure 7.3: Feature Importance Analysis for Depression Detection using FW-AK-KNN

Table 7.4: Feature Importance Scores for BAI-based Anxiety Detection using FW-AK-KNN

BAI Feature	**Importance**
10	0.079580
8	0.073244
7	0.069770
9	0.063659
4	0.061047
5	0.061027
6	0.057505
14	0.051738
13	0.051137
17	0.048978
3	0.047646
2	0.045006
18	0.037465
12	0.037013
20	0.036097
15	0.034414
19	0.032637
21	0.031829
1	0.030939
11	0.028424
16	0.020843

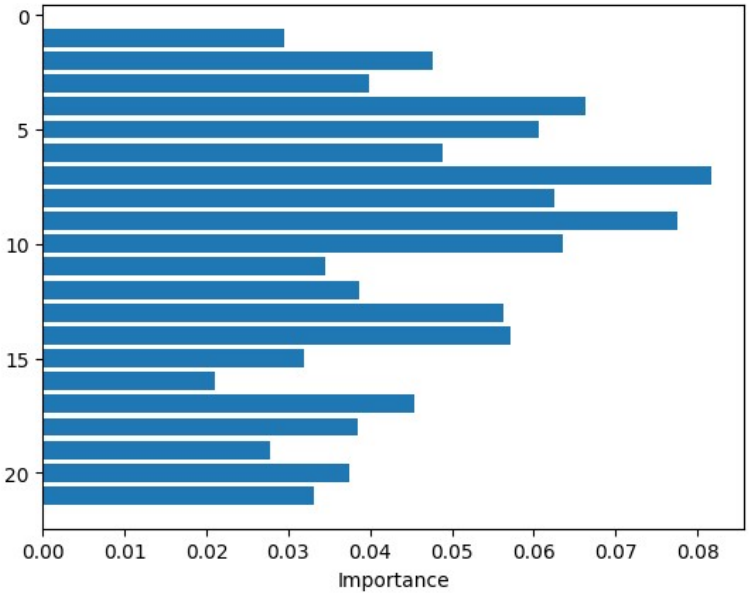


Figure 7.4: Feature Importance Analysis for Anxiety Detection using FW-AK-KNN

The results indicate that certain BAI features contribute more significantly to anxiety prediction. In particular, Features 10, 8, and 7 exhibit the highest importance scores, suggesting that these symptoms play a dominant role in determining anxiety levels as shown in Fig. 7.4. Conversely, features such as Feature 16 and Feature 11 show relatively lower contributions.

This variation highlights that not all anxiety symptoms contribute equally to prediction, and the proposed FW-AK-KNN model effectively captures these differences through adaptive weighting.

Furthermore, the ability of the model to identify key contributing BAI items enhances interpretability, making it especially valuable in mental health applications where understanding symptom-level influence is critical.

This shows that the suggested FW-AK-KNN model addresses the interpretability issues with conventional KNN approaches while simultaneously improving classification performance and offering insightful information about symptom-level contributions for anxiety detection.

7.4 Discussion

By examining the effects of various kernel functions and gamma configurations, the ablation study offers a greater understanding of the behaviour and adaptability of the suggested FW-AK-KNN model. Despite the introduction of the adaptive gamma formulation as a crucial advance to dynamically adapt to local data distributions, the experimental results indicate that fixed

gamma values—particularly $\gamma = 0.1$—often yield superior performance in practice. This trend is especially evident when using Epanechnikov and Triangular kernel functions, as demonstrated in Table 7.5.

Table 7.5: Ablation study results on the BDI and BAI datasets using various kernel functions and gamma values. The highest accuracy and F1-score were obtained using the Epanechnikov kernel with a gamma value of 0.1 on both datasets.

Kernel	γ	D1-D (Acc / F1 / Δ)	D1-A (Acc / F1 / Δ)
Gaussian	0.1	0.9068 / 0.9050 / +0.07	0.7860 / 0.7770 / +0.05
Epanechnikov	0.1	**0.9068 / 0.9049 / +0.07**	**0.8007 / 0.7923 / +0.07**
Triangular	1	0.9003 / 0.8990 / +0.06	0.7860 / 0.7779 / +0.05

Note: Δ = improvement over KNN (BDI baseline: 0.84 / BAI baseline: 0.73). Bold = highest performance.

This observation does not contradict the proposed methodology; rather, it highlights the adaptability and modular nature of the FW-AK-KNN framework. The model is designed such that its key components, including feature-weighted distance computation, kernel-based neighbor weighting, and gamma scaling, can be independently configured.

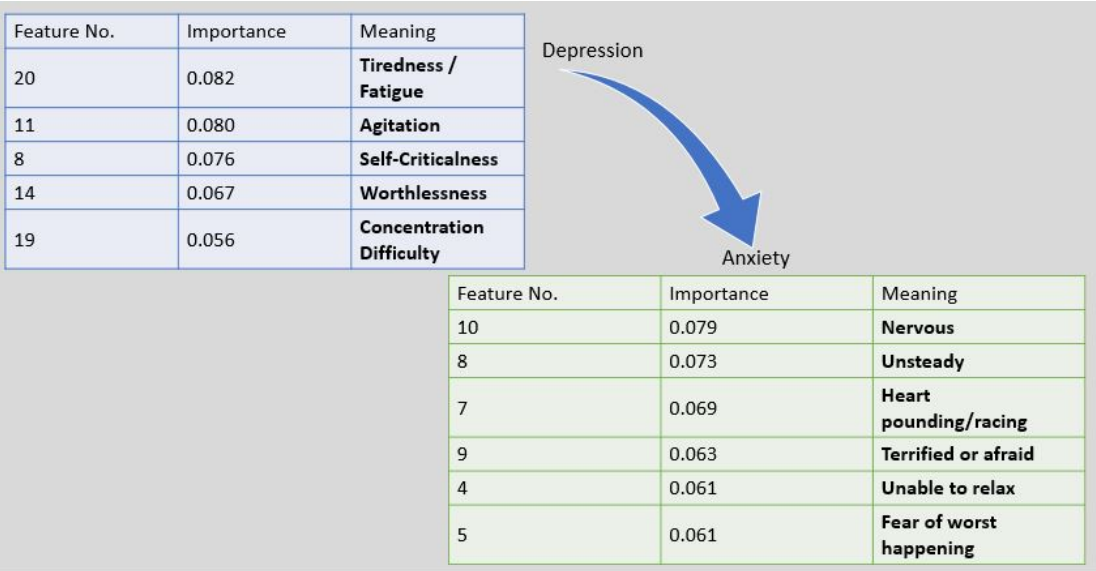
The ablation results further reveal that the Epanechnikov kernel with $\gamma = 0.1$ achieves the best overall performance on both depression (BDI) and anxiety (BAI) datasets, producing the highest accuracy and F1-score improvements over baseline KNN. This suggests that, in certain cases, a well-chosen fixed parameter may provide a more stable and effective decision boundary compared to dynamically varying parameters, particularly when the dataset exhibits relatively consistent local structures.

In addition to performance improvements, the proposed FW-AK-KNN model provides valuable interpretability through its feature weighting mechanism. The analysis of feature importance on both BDI and BAI datasets reveals that not all features contribute equally to classification. Instead, a subset of features consistently exhibits higher importance scores, indicating their stronger influence in determining depression and anxiety outcomes.

For the BDI dataset, features such as 20, 11, and 8 were identified as the most influential, whereas in the BAI dataset, features 10, 8, and 7 contributed more prominently as depicted by Fig. 7.5. This distinction highlights that depression and anxiety are characterized by different dominant symptom patterns, despite their clinical overlap.

Because the model can automatically capture and highlight condition-specific feature contributions, these results further verify the efficacy of the suggested feature-weighted distance formulation. This improves classification accuracy and offers insightful information

on symptom-level relevance, which is crucial for applications related to mental health.



Feature No.	Importance	Meaning
20	0.082	**Tiredness / Fatigue**
11	0.080	**Agitation**
8	0.076	**Self-Criticalness**
14	0.067	**Worthlessness**
19	0.056	**Concentration Difficulty**

Feature No.	Importance	Meaning
10	0.079	**Nervous**
8	0.073	**Unsteady**
7	0.069	**Heart pounding/racing**
9	0.063	**Terrified or afraid**
4	0.061	**Unable to relax**
5	0.061	**Fear of worst happening**

Figure 7.5: Significant Depression and Anxiety Features

Overall, the findings reinforce the strength of FW-AK-KNN as a flexible and interpretable classification framework. Its ability to adapt through configurable components ensures consistent performance improvements across datasets with varying characteristics. This makes the model not only effective for mental health prediction but also robust across broader biomedical domains, supporting its applicability as a domain-agnostic solution.

7.5 Summary

The FW-AK-KNN model for identifying anxiety and depression in post-COVID-19 environments was introduced in this chapter. The proposed approach enhanced the traditional KNN algorithm by incorporating feature importance into the distance computation and introducing kernel-based adaptive weighting for neighbor contributions. The shortcomings of traditional KNN, specifically its incapacity to manage feature heterogeneity and fluctuating data distributions in psychometric datasets, were intended to be addressed by these improvements.

Standard performance indicators, such as accuracy, precision, recall, and F1-score, were used to assess the model on several psychometric and biological datasets. All datasets showed consistent improvements over baseline KNN in the experimental findings, with primary depression and anxiety datasets showing especially notable advances. The suggested model was more successful in identifying mild to moderate cases because it demonstrated enhanced

ability to recognise subtle symptom patterns and achieved a better balance between precision and memory.

Further analysis through ablation studies highlighted the flexibility of the FW-AK-KNN framework, showing that both adaptive and fixed parameter configurations can yield strong performance depending on dataset characteristics. This modular design allows the model to be tuned according to specific application requirements, enhancing its robustness and applicability across domains.

All things considered, the FW-AK-KNN model offers a practical, comprehensible, and computationally efficient method for predicting mental health. It is appropriate for real-world deployment because of its capacity to provide enhanced performance while preserving simplicity, especially in settings with limited resources. Thus, the suggested method has possibilities for future adaptations to real-time and multimodal applications and aids in the creation of scalable, data-driven mental health screening systems.

Chapter 8

Conclusion, Future Work & Social Impact

In this last chapter, we summarise the main conclusions, highlight the research's major contributions, and talk about possible future prospects for this field of study.

8.1 Conclusion

This thesis examined automated diagnosis of anxiety and depression using machine learning, deep learning, and hybrid intelligence models in relation to the COVID-19 pandemic. Globally, the COVID-19 pandemic had a significant influence on psychological health, which increased the need for trustworthy and precise methods for preliminary mental health screening. The main aim of this research was to develop and test intelligent models to determine different levels of depression and anxiety using standardized questionnaires like BDI and BAI.

The thesis addressed this objective through three core research objectives:

1. To conduct a systematic literature review and analysis of existing research on anxiety and depression detection employing ML and DL models.
2. To study, implement, and analyze the performance of DL/ML classification models for anxiety and depression detection induced during COVID-19.
3. To study and analyze the impact of COVID-19 on ongoing college students employing an improved DL/ML model for depression and anxiety detection.

To begin with, a comprehensive SLR was performed in order to detect key trends, methodological approaches, and research gaps in COVID-19 impact studies on mental health

analysis. The main difficulties identified were related to limited variability in data samples, binary prediction models, and unexplored data on sustained symptoms in people with never-affected mental health. These results served as the impetus for the development of even more generalized models that are statistically validated and can handle multiclass classification.

Later on, the thesis went on to empirically study the impact of dataset size on traditional ML classifiers. The findings indicated an improvement in model robustness with an increase in dataset size until it reached saturation. While larger datasets benefited the most from ensemble and margin-based classifiers, simpler models demonstrated acceptable performance at moderate data sizes. The present work makes several practical insights into the data sufficiency requirements of mental health prediction systems.

Numerous DL model designs, including ANN, RNN, LSTM, CNN, and the TabNet architecture, were also examined in the thesis. Testing on many datasets revealed that DL models, such as LSTM, CNN, and TabNet, are superior to classical ML models for the job of multi-class classification and are effective at capturing the non-linear interactions between psychological variables.

The concept of DT+KNN, RF+NN, and DA-TabSVM models gained popularity as a solution to the shortcomings of standalone models. Among the mentioned models, the DA-TabSVM model performed well compared to other models because the DA-TabSVM model was able to perform near perfectly in classifications.

In conclusion, the thesis proposed a novel model called ASKNN, a variation of traditional KNN, which adaptively controls neighborhood influence. The proposed model ASKNN always achieved better performance compared to different baseline versions of traditional KNN on different evaluation metrics. The Friedman test, Wilcoxon signed-rank test, and effect size test all supported the significance of these gains.

In addition to improving classification performance, the proposed FW-AK-KNN provides enhanced interpretability through its feature weighting mechanism. By learning and assigning weights to individual features, the model is capable of quantifying the contribution of each psychometric attribute in predicting mental health conditions.

The feature contribution analysis performed on both BDI and BAI datasets revealed that only a subset of features significantly influences prediction outcomes, with distinct dominant features observed for depression and anxiety. This illustrates how the suggested approach can identify symptom patterns unique to a disorder rather than treating all features equally.

Because it offers deeper insights into symptom-level relevance and promotes more educated screening and intervention tactics, this interpretability is especially crucial in mental health

applications. As a result, the suggested model enhances predictive accuracy while also making mental health assessments more comprehensible and clinically significant.

8.1.1 Lessons Learnt

Based on the literature review, experimental analysis, and findings, the following key lessons were derived:

1. **Why is early mental health detection crucial?** Depression and anxiety might often go undiagnosed because of stigma and lack of access to psychiatric professionals. Automated screening tools utilizing standardized questionnaires make it easier to detect the conditions and help intervene in a timely manner, especially in the scenario of a global outbreak.

2. **How does dataset size influence mental health prediction models?** A larger data set is associated with low variability and stable models. But there is a threshold beyond which the rate of growth of model performances slows down. It is thus important to determine the most appropriate size of the data set.

3. **Why are DL models effective for psychological data?** Psychological inventories are characterized by complex, non-linear interactions of symptoms. Neural networks, specifically LSTMs, CNNs, and TabNet, capture such interactions well and outperform traditional ML algorithms in multiclass classification tasks.

4. **Why are hybrid models necessary?** Hybrid models exploit the complementarity of strengths in different algorithms. Mixing models that are interpretable with strong classifiers enhances robustness, accuracy, and flexibility over data sets, illustrated in the DA-TabSVM framework. Hybrid models exploit the complementarity of strengths in different algorithms. Mixing models that are interpretable with strong classifiers enhances robustness, accuracy, and flexibility over data sets, illustrated in the DA-TabSVM framework.

5. **How does adaptive learning improve KNN performance?** The traditional KNN suffers badly from sensitivity to distance metrics and neighborhood size. The proposed ASKNN model, on the other hand, adjusts the neighbor's influence dynamically for better discrimination and stability in classification.

6. **Why is statistical validation essential?** Performance improvements must be statistically validated by appropriate tests of significance and measures of effect size. That means any observed gains are due to reasons other than random chance, adding strength to the reliability of conclusions drawn from experiments.

8.1.2 Ethical Implications and Responsible Use

This is where automated mental health assessment systems raise an important ethical consideration, considering the sensitive nature of psychological data.

8.1.2.1 Privacy Concerns

Mental health information is very private and confidential in nature. In collecting information through questionnaires, it is important to consider all ethical requirements, such as informed consent, use of information with clarity, volunteering to participate, and the right to withdraw upon choice without penalty and risk. Information storage and anonymization are very important to avoid abuse.

8.1.2.2 Potential Biases

There can be bias because of demographic representation, self-selection bias in online surveys, and assumptions made by the algorithm. It is important to address these potential biases while handling the data.

8.1.2.3 Responsible Use

Mental models should aid, and must never replace, professional decision-making. It is necessary to employ the models in a manner that ensures that results are interpreted by a professional.

8.1.2.4 Limitations

Despite all empirical performances reported in multiple models using ML, DL, and their hybrids in this thesis, there are a number of limitations that need to be discussed. First, most of the data in this work were collected via online self-report questionnaires like PHQ-9, GAD-7, BDI, and BAI using online platforms and snowball sampling. This approach was certainly appropriate and practical during the post-COVID-19 era; however, it has inherently brought certain self-selection and sampling biases. There may be an overrepresentation of individuals who are

either more technologically adept or concerned about their mental health, which could inflate symptom prevalence and limit generalizability to the broader population. Although this work used cross-validation techniques to improve the stability of models, these cannot fully remove underlying biases in the sampling.

Second, reliance on self-reported psychological inventories adds other layers of bias, including social desirability bias and recall bias, which may affect the accuracy of symptom reporting and, by extension, lower predictive model performance. In addition, because all the datasets in this work are cross-sectional, one is limited to causal inference due to the fact that observed associations between features and mental health outcomes are correlational, not longitudinal. This limits the ability to model symptom progression or detect early warning signals over time.

Another limitation involves the loss of contextual richness inherent in questionnaire-based measures. This work did not include direct clinical interviews, behavioral observations, or real-time physiological data; thus, it might be limited in the depth of mental health characterization. Finally, while multiple models achieved near-ceiling performance, especially with increasing dataset sizes, such plateau effects point to lower returns beyond thresholds in data and further highlight careful model and feature selection. Lastly, although highly advanced hybrid and adaptive learning models like DA-TabSVM and ASKNN performed well in generalization capabilities on given sets, further testing is required in different populations and environments before usage on a massive scale.

An additional limitation concerns the demographic and geographical scope of the study populations. Since the datasets primarily comprise college and university students and were collected predominantly from an Indian population, the findings may not directly generalize to non-student populations or individuals from different cultural and geographical contexts. Differences in cultural norms, social environments, educational backgrounds, and perceptions of mental health may influence responses to psychological assessment instruments and consequently affect model predictions. Moreover, the post-pandemic context in which the data were collected represents a particular period of changing mental health experiences and behavioural patterns. Therefore, the long-term applicability of the proposed models to populations whose experiences differ from the post-pandemic context cannot be assumed without further validation. Future research should evaluate these models using larger, longitudinal, cross-cultural, and demographically diverse datasets, including non-student populations, to establish their robustness and generalizability across different populations and time periods.

8.2 Future Work

The limitations identified above offer some valuable pointers to future studies. Firstly, there is scope for adopting a better sampling method, like stratified random sampling, to improve the representativeness of the findings. Moreover, participants from underrepresented groups should be encouraged to take part in studies. A combination of clinician-conducted diagnostic interviews and self-report inventories may help improve the validity of ground truth and end the need to rely on subjective data. Moreover, longitudinal research should be explored to monitor the development of depression and anxiety over time so that causal relationships and the development of symptoms can be investigated.

Future work can also focus on conceptualizing the fusion of multimodal sources of information, such as wearable sensor, speech, text, behavior, and physiological modalities, which may further improve the contextual and diagnostic information. The multimodal approach may complement the methodology of questionnaires to capture nuances of dynamic patterns of mental health. From a modeling point of view, research work on ensemble/hybrid models, hyperparameter optimization, and adaptive learning methods may further improve robustness/scaling capabilities on diverse datasets. From a modeling point of view, research work on ensemble/hybrid models, hyperparameter optimization, and adaptive learning methods may further improve robustness/scaling capabilities on diverse datasets.

Interpretability of the models is still an important priority. Future work should consider the application of other AI interpretability methods like SHAP or LIME. Transparencies of the models will help improve acceptability among the practitioners of mental health. There are other ethical considerations that should be explored by the models as they are on the brink of being implemented. Finally, extending such proposed frameworks toward real-time or automated screening systems and validating them in practical healthcare and community settings could bring a significant improvement in early diagnosis, reducing delays in interventions and contributing to global mental health aligned with the United Nations Sustainable Development Goal 3: Good Health and Well-being.

8.3 Social Impact and Applications

The research work offered in this thesis has immense social implications, most importantly, within the backdrop of the rising need for mental health due to the emergence and spread of the COVID-19 pandemic. Depression and Anxiety became a major concern among all sections

of people, cutting across age, profession, and class. This work, by the application of ML and DL techniques for early detection and classification of anxiety and depression, contributes to the development of scalable, data-driven mental health frameworks that can support timely identification, intervention, and resource allocation at both the individual and societal levels.

One of the major social impacts that this research may have is that it could potentially enable healthcare professionals and mental health practitioners to offer early screening and decision support. Models trained on standardized psychological instruments, such as PHQ-9, GAD-7, BDI, and BAI, could support health professionals by automatically flagging patients at high risk, thus minimizing diagnostic delay and alleviating the pressure on mental health services. These tools are even more valuable in resource-scarce environments where access to mental health professionals may be a challenge during a mass disaster. These tools help in screening candidates who require critical attention in a timely and accurate manner.

This research also holds importance from a public health policy and a huge population surveillance perspective. The relevant authority can use the findings from the mental health analytics of the ML approach for understanding the trend of psychological issues during a pandemic or any other crisis. The effectiveness of analyzing large datasets makes it possible for a policymaker to develop a strategic intervention for mental health issues and implement a preventive program and resource allocation efficiently. In addition, insights into the connection between the size of a given set of data and the strength of predictions made by a mental health predictive model can be extremely useful for future survey development.

In learning institutions such as universities and colleges that were disrupted for an extended period through online learning during the COVID-19 period, the proposed models can be used to aid in initiatives related to monitoring student mental health. Early identification of anxiety and depression among university and college students allows institutions to provide preventive and intervention programs. Similarly, in work settings, most notably in a corporate setting where there are high levels of stress, the screening process can also be implemented in employee wellness initiatives to promote mental wellness, build a healthier corporate culture, and all that while strictly observing confidentiality.

In terms of technology, the hybrid and adaptive models proposed in the form of DA-TabSVM and ASKNN in this thesis show the applicability of a combination of classical ML techniques and advanced feature learning and adaptive strategies for the development of accurate mental health assessment tools. These models are suited for implementation into digital health platforms, mobile applications, and web portals for mental health assessment and would require no specialized hardware and intrusive methods of data collection for scalability to be achieved.

Significantly, this work can be associated with the international interest in improving mental wellness, as it adheres to SDG 3: Good Health and Well-being, advocated by the United Nations. The goal of enhancing the accuracy and validity rates of detecting depression and anxiety can be addressed by the proposed methodologies in reducing undiagnosed mental illness and the escalation thereof. Most importantly, aside from the development in technology, this work focuses on the necessary uses of artificial intelligence in improving mental health by promoting fairness in mental health strategies utilizing data analytics.

In summary, this work shows that the accuracy, dependability, and usefulness of automated mental health assessment systems can be greatly improved by combining interpretable feature-weighted techniques, adaptive machine learning, and deep learning.

References

[1] United nations sustainable development goals report. Online. Accessed: Jul. 03, 2024.

[2] Number of undergraduate degree holders across delhi in india in 2022, by gender, 2024. Accessed: Aug. 05, 2024.

[3] Mrayam Biglari Abhari, John W Fisher, Azita Kheiltash, and Marzieh Nojomi. Validation of the persian version of spiritual well-being questionnaires. *Iranian journal of medical sciences*, 43(3):276, 2018.

[4] Stefan Aeberhard and M. Forina. Wine dataset. `https://archive.ics.uci.edu/dataset/109/wine`, n.d. UCI Machine Learning Repository.

[5] Adrian Aguilera, Erin Bruehlman-Senecal, Orianna Demasi, Paloma Avila, Maya Breton, Pavel Cernasov, Luis Chavez, Daniel Garcia, Guillermo Hernandez, Jessica Martinez, Alejandra Sanchez, Catalina Velez, Camila Figueroa, Enrique Rodriguez-Villa, C. Rojas, and Stephen M. Schueller. A text messaging intervention (StayWell at home) to counteract depression and anxiety during COVID-19 social distancing: pre-post study. *JMIR Mental Health*, 8(11):e25298, 2021.

[6] Essam Al Daoud. Comparison between xgboost, lightgbm and catboost using a home credit dataset. *International Journal of Computer and Information Engineering*, 13(1):6–10, 2019.

[7] Emmanuel Oluwadunsin Alabi, Oluwashola David Adeniji, Tolulope Moyo Awoyelu, and Oluwakemi Dunsin Fasae. Hybridization of machine learning techniques in predicting mental disorder. *International Journal of Human Computing Studies*, 3(6):22–30, 2021.

[8] Faisal Mashel Albagmi, Aisha Alansari, Deema Saad Al Shawan, Heba Yaagoub AlNujaidi, and Sunday O Olatunji. Prediction of generalized anxiety levels during the covid-19 pandemic: A machine learning-based modeling approach. *Informatics in Medicine Unlocked*, 28:100854, 2022.

[9] Douglas G Altman and J Martin Bland. Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, 308(6943):1552, 1994.

[10] Sercan Ö Arik and Tomas Pfister. Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6679–6687, 2021.

[11] Priyanka Arora and Sonika Dahiya. From virus to vulnerability: A systematic literature review of psychological aftershocks following the covid-19 pandemic. In *Proceedings of ICDAM 2025*. Springer, 2025.

[12] Priyanka Arora and Sonika Dahiya. A novel adaptive sigma knn model for depression and anxiety detection following the covid 19 pandemic. *Scientific Reports*, 15(1):40553, 2025.

[13] Mariette Awad and Rahul Khanna. Support vector regression. In *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, pages 67–80. Springer, 2015.

[14] Nan Zou Bakkeli. Predicting psychological distress during the covid-19 pandemic: do socioeconomic factors matter? *Social Science Computer Review*, 41(4):1227–1251, 2023.

[15] Célyne H Bastien, Annie Vallières, and Charles M Morin. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep medicine*, 2(4):297–307, 2001.

[16] Ben Beaglehole, Roger T Mulder, Chris M Frampton, Joseph M Boden, Giles Newton-Howes, and Caroline J Bell. Psychological distress and psychiatric disorder after natural disasters: systematic review and meta-analysis. *The British Journal of Psychiatry*, 213(6):716–722, 2018.

[17] Aaron Temkin Beck, C Ward, M Mendelson, J Mock, and J Erbauch. Beck depression inventory (bdi), 1961.

[18] Arie Ben-David. A lot of randomness is hiding in accuracy. *Engineering Applications of Artificial Intelligence*, 20(7):875–885, 2007.

[19] Shajulin Benedict. Shared mobility intelligence using permissioned blockchains for smart cities. *New generation computing*, 40(4):1009–1027, 2022.

[20] Laure Bera, Mathilde Souchon, Audrey Ladsous, Vincent Colin, and Jorge Lopez-Castroman. Emotional and behavioral impact of the covid-19 epidemic in adolescents. *Current psychiatry reports*, 24(1):37–46, 2022.

[21] José Eduardo Rondón Bernard. Depression: A review of its definition. *MOJ Addiction Medicine & Therapy*, 5(1):6–7, 2018.

[22] Hind A Beydoun, May A Beydoun, Jordan Weiss, Rana S Gautam, Sharmin Hossain, Brook T Alemu, and Alan B Zonderman. Predictors of covid-19 level of concern among older adults from the health and retirement study. *Scientific Reports*, 12(1):4396, 2022.

[23] S. Bharathi, D. Poongodi, and L. Krithika. K-nn version metrics for predicting breast cancer survival. *African Journal of Biological Sciences*, pages 429–435, January 2024.

[24] Shaurya Bhatnagar, Jyoti Agarwal, and Ojasvi Rajeev Sharma. Detection and classification of anxiety in university students through the application of machine learning. *Procedia Computer Science*, 218:1542–1550, 2023.

[25] BH Bhavani and NC Naveen. An approach to determine and categorize mental health condition using machine learning and deep learning models. *Engineering, Technology & Applied Science Research*, 14(2):13780–13786, 2024.

[26] Josephine Bierbaum, Melissa Lynn, and Louis Yu. Utilizing pattern mining and classification algorithms to identify risk for anxiety and depression in the lgbtq+ community during the covid-19 pandemic. In *Companion Proceedings of the Web Conference 2022*, pages 663–672, 2022.

[27] Brit A Blaauw, Grete Dyb, Knut Hagen, Turid L Holmen, Mattias Linde, Tore Wentzel-Larsen, and John-Anker Zwart. Anxiety, depression and behavioral problems among adolescents with recurrent headache: the young-hunt study. *The journal of headache and pain*, 15(1):38, 2014.

[28] Christy A Blevins, Frank W Weathers, Margaret T Davis, Tracy K Witte, and Jessica L Domino. The posttraumatic stress disorder checklist for dsm-5 (pcl-5): Development and initial psychometric evaluation. *Journal of traumatic stress*, 28(6):489–498, 2015.

[29] Adam Bodner, Lauren Ruhl, Emily Barr, Akshaya Shridhar, Stephanie Skakoon-Sparling, and Kiffer George Card. The impact of working from home on mental health: A cross-sectional study of canadian workers' mental health during the third wave of the COVID-

19 pandemic. *International Journal of Environmental Research and Public Health*, 19(18):11588, September 2022.

[30] Athman Bouguettaya, Qi Yu, Xumin Liu, Xiangmin Zhou, and Andy Song. Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, 42(5):2785–2797, 2015.

[31] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[32] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[33] Carol S Burckhardt and Kathryn L Anderson. The quality of life scale (qols): reliability, validity, and utilization. *Health and quality of life outcomes*, 1(1):60, 2003.

[34] David D Burns. *The feeling good handbook: The groundbreaking program with powerful new techniques and step-by-step exercises to overcome depression, conquer anxiety, and enjoy greater intimacy*. Penguin, 1999.

[35] Daniel J Buysse, Charles F Reynolds III, Timothy H Monk, Susan R Berman, and David J Kupfer. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2):193–213, 1989.

[36] Hanshu Cai, Jiashuo Han, Yunfei Chen, Xiaocong Sha, Ziyang Wang, Bin Hu, Jing Yang, Lei Feng, Zhijie Ding, Yiqiang Chen, et al. A pervasive approach to eeg-based depression detection. *complexity*, 2018(1):5238028, 2018.

[37] Daniela Caldirola, Silvia Dacco, Francesco Cuniberti, Massimiliano Grassi, Alessandra Alciati, Tatiana Torti, and Giampaolo Perna. First-onset major depression during the covid-19 pandemic: A predictive machine learning model. *Journal of Affective Disorders*, 310:75–86, 2022.

[38] Alessandro Carollo, Andrea Bizzego, Giulio Gabrieli, Keri Ka-Yee Wong, Adrian Raine, and Gianluca Esposito. Self-perceived loneliness and depression during the covid-19 pandemic: a two-wave replication study. *UCL Open Environment*, 4:e051, 2022.

[39] Francisco Roberto Sanchez Cavalheiro and Marina Greghi Sticca. Adaptation and validation of the brazilian version of the fear of covid-19 scale. *International Journal of Mental Health and Addiction*, 20(2):921–929, 2022.

[40] Danton S Char, Nigam H Shah, and David Magnus. Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, 378(11):981, 2018.

[41] Sankhadeep Chatterjee, Soumyajit Maity, Mayukh Bhattacharjee, Soumen Banerjee, Asit Kumar Das, and Weiping Ding. Variational autoencoder based imbalanced covid-19 detection using chest x-ray images. *New Generation Computing*, 41(1):25–60, 2023.

[42] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[43] Shuquan Chen, Kaiwen Bi, Pei Sun, and George A Bonanno. Psychopathology and resilience following strict covid-19 lockdowns in hubei, china: Examining person-and context-level predictors for longitudinal trajectories. *American Psychologist*, 77(2):262, 2022.

[44] Tianqi Chen. Xgboost: A scalable tree boosting system. *Cornell University*, 2016.

[45] Walid Cherif. Optimization of k-nn algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis. *Procedia Computer Science*, 127:293–299, 2018.

[46] Junggu Choi and Sanghoon Han. Investigation of factors associated with mental health during the early part of the covid-19 pandemic in south korea based on machine learning algorithms: A cohort study. *Digital Health*, 9:20552076231207573, 2023.

[47] Junggu Choi, Kion Kim, Soo Hyun Park, Juyoen Hur, Hyunjung Yang, Young-Hoon Kim, Hakbae Lee, and Sanghoon Han. Investigation of factors regarding the effects of covid-19 pandemic on college students' depression by quantum annealer. *Scientific reports*, 14(1):4684, 2024.

[48] Arman Hossain Chowdhury, Dana Rad, and Md Siddikur Rahman. Predicting anxiety, depression, and insomnia among bangladeshi university students using tree-based machine learning models. *Health Science Reports*, 7(4):e2037, 2024.

[49] Sheldon Cohen, Tom Kamarck, Robin Mermelstein, et al. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 10(2):1–2, 1994.

[50] GBD 2019 Mental Disorders Collaborators et al. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet Psychiatry*, 9(2):137–150, 2022.

[51] Kathryn M Connor, Jonathan R Davidson, et al. Connor–davidson resilience scale. *Confirmatory Factor Analysis of the Connor-Davidson Resilience Scale*, 2023.

[52] John L Cox, Jeni M Holden, and Ruth Sagovsky. Detection of postnatal depression: development of the 10-item edinburgh postnatal depression scale. *The British journal of psychiatry*, 150(6):782–786, 1987.

[53] D. Foundation. Covid impact survey. `https://www.covid-impact.org/`, June 2020. Accessed: 2023-09-11.

[54] Joydip Dhar, Ashaya Shukla, Mukul Kumar, and Prashant Gupta. A weighted mutual k-nearest neighbour for classification mining. *arXiv preprint arXiv:2005.08640*, 2020.

[55] Christoffer Dharma, Daniel Grace, Carmen Logie, Alex Abramovich, Nicholas Mitsitakis, Bruce Baskersville, and Michael Chaiton. Identifying factors affecting depressive symptoms and incidence of mental health diagnosis within 1 year among 2slgbtq+ youth during covid-19 using machine learning methods. 2022.

[56] Mariagrazia Di Giuseppe, John Christopher Perry, Matilde Lucchesi, Monica Michelini, Sara Vitiello, Aurora Piantanida, Matilde Fabiani, Sara Maffei, and Ciro Conversano. Preliminary reliability and validity of the dmrs-sr-30, a novel self-report measure based on the defense mechanisms rating scales. *Frontiers in psychiatry*, 11:870, 2020.

[57] ED Diener, Robert A Emmons, Randy J Larsen, and Sharon Griffin. The satisfaction with life scale. *Journal of personality assessment*, 49(1):71–75, 1985.

[58] YanHong Dong, Mei Chun Yeo, Xiang Cong Tham, Rivan Danuaji, Thang H Nguyen, Arvind K Sharma, Komalkumar Rn, Meenakshi Pv, Mei-Ling Sharon Tai, Aftab Ahmad, et al. Investigating psychological differences between nurses and other health care workers from the asia-pacific region during the early phase of covid-19: Machine learning approach. *JMIR nursing*, 5(1):e32647, 2022.

[59] GOLDBERG DP. User's guide to the general health questionnaire. *Windsor*, 1988.

[60] M Darby Dyar and Cai R Ytsma. Effect of data set size on geochemical quantification accuracy with laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 177:106073, 2021.

[61] Giordano D'Urso, Alfonso Magliacano, Sayna Rotbei, Felice Iasevoli, Andrea de Bartolomeis, and Alessio Botta. Predicting the severity of lockdown-induced psychiatric symptoms with machine learning. *Diagnostics*, 12(4):957, 2022.

[62] Jane L Eisen, Katharine A Phillips, Lee Baer, Douglas A Beer, Katherine D Atala, and Steven A Rasmussen. The brown assessment of beliefs scale: reliability and validity. *American Journal of Psychiatry*, 155(1):102–108, 1998.

[63] Christo El Morr, Manar Jammal, Imad Bou-Hamad, Sahar Hijazi, Dinah Ayna, Maya Romani, and Reem Hoteit. Predictive machine learning models for assessing lebanese university students' depression, anxiety, and stress during covid-19. *Journal of Primary Care & Community Health*, 15:21501319241235588, 2024.

[64] Jon D Elhai, Haibo Yang, Dean McKay, Gordon JG Asmundson, and Christian Montag. Modeling anxiety and fear of covid-19 using machine learning in a sample of chinese adults: associations with psychopathology, sociodemographic, and exposure variables. *Anxiety, Stress, & Coping*, 34(2):130–144, 2021.

[65] Hakkun Elmunsyah, Risalatul Mu'awanah, Triyanna Widiyaningtyas, Ilham AE Zaeni, and Felix Andika Dwiyanto. Classification of employee mental health disorder treatment with k-nearest neighbor algorithm. In *2019 international conference on electrical, electronics and information engineering (ICEEIE)*, volume 6, pages 211–215. IEEE, 2019.

[66] Matthias Elter, Rüdiger Schulz-Wendtland, and Thomas Wittenberg. The prediction of breast cancer biopsy outcomes using two cad approaches that both emphasize an intelligible decision process. *Medical physics*, 34(11):4164–4172, 2007.

[67] Gunther Eysenbach. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (cherries), 2004.

[68] Rajib Ahmed Faisal, Mary C Jobe, Oli Ahmed, and Tanima Sharker. Mental health status, anxiety, and depression levels of bangladeshi university students during the covid-19 pandemic. *International journal of mental health and addiction*, 20(3):1500–1515, 2022.

[69] Juliana Fernandes Filgueiras Meireles, Clara Mockdece Neves, Fabiane Frota da Rocha Morgado, and Maria Elisa Caputo Ferreira. Zika virus and pregnant women: a psychological approach. *Psychology & Health*, 32(7):798–809, 2017.

[70] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[71] Luca Flesia, Merylin Monaro, Cristina Mazza, Valentina Fietta, Elena Colicino, Barbara Segatto, and Paolo Roma. Predicting perceived stress related to the covid-19 outbreak through stable psychological traits and machine learning models. *Journal of clinical medicine*, 9(10):3350, 2020.

[72] Chiara Foà, Annalisa Tonarelli, Luca Caricati, and Laura Fruggeri. Cope-nvi-25: validazione italiana della versione ridotta della coping orientation to the problems experienced (cope-nvi). *Psicologia della salute: quadrimestrale di psicologia e scienze della salute: 2, 2015*, pages 123–140, 2015.

[73] David Fong, Tianshu Chu, Matthew Heflin, Xiaosi Gu, and Oshani Seneviratne. Predicting depression and anxiety: a multi-layer perceptron for analyzing the mental health impact of covid-19. In *Proceedings of the 2024 8th International Conference on Medical and Health Informatics*, pages 325–331, 2024.

[74] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[75] Toshi A Furukawa, Norito Kawakami, Mari Saitoh, Yutaka Ono, Yoshibumi Nakane, Yosikazu Nakamura, Hisateru Tachimori, Noboru Iwata, Hidenori Uda, Hideyuki Nakane, et al. The performance of the japanese version of the k6 and k10 in the world mental health survey japan. *International journal of methods in psychiatric research*, 17(3):152–158, 2008.

[76] Sander L Gilman. Moral panic and pandemics. *The Lancet*, 375(9729):1866–1867, 2010.

[77] Lucas Giner, Constanza Vera-Varela, Diego De La Vega, Giovani M Zelada, and Julio A Guija. Suicidal behavior in the first wave of the covid-19 pandemic. *Current psychiatry reports*, 24(1):1–10, 2022.

[78] S Gnanambal, M Thangaraj, VT Meenatchi, and V Gayathri. Classification algorithms with attribute selection: an evaluation study using weka. *International Journal of Advanced Networking and Applications*, 9(6):3640–3644, 2018.

[79] Wayne K Goodman. The yale-brown obsessive compulsive scale. *Archives of General Psychiatry*, 46(11):1012, 1989.

[80] Robin Goodwin, Shamsul Haque, Felix Neto, and Lynn B Myers. Initial psychological responses to influenza a, h1n1 (" swine flu"). *BMC infectious diseases*, 9(1):166, 2009.

[81] Jianping Gou, Hongxing Ma, Weihua Ou, Shaoning Zeng, Yunbo Rao, and Hebiao Yang. A generalized mean distance-based k-nearest neighbor classifier. *Expert Systems with Applications*, 115:356–372, 2019.

[82] Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359. Springer, 2005.

[83] Michael M Grant. Beck anxiety inventory. In *Encyclopedia of child behavior and development*, pages 215–217. Springer, 2011.

[84] Gianluigi Guido, Alessandro M Peluso, Mauro Capestro, and Mariafrancesca Miglietta. An italian version of the 10-item big five inventory: An application to hedonic and utilitarian shopping values. *Personality and Individual Differences*, 76:135–140, 2015.

[85] Nevin Güler Dinçer, Serdar Demir, and Muhammet Oğuzhan Yalçın. Forecasting covid19 reliability of the countries by using non-homogeneous poisson process models. 2022.

[86] Charu Gupta, Dev Gaur, Prateek Agrawal, and Deepali Virmani. Huda_covid human disposition analysis during covid-19 using machine learning. *International Journal of E-Health and Medical Communications (IJEHMC)*, 13(2):1–15, 2021.

[87] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422, 2002.

[88] Bijay Halder, Jatisankar Bandyopadhyay, and Papiya Banik. Covid-19 pandemic: a health challenge for commoners during the first unlock phase in india. *Journal of Public Health*, 31(3):427–433, 2023.

[89] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. *ACM sigmod record*, 29(2):1–12, 2000.

[90] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[91] Md Emran Hasan, Mohammad Arif, SM Rakibul Hasan, Moses Muwanguzi, Joan Abaatyo, Mark Mohan Kaggwa, Moneerah Mohammad ALmerab, Pawel A Atroszko, Mohammad Muhit, Firoj Al-Mamun, et al. Prevalence, associated factors, and machine learning-based prediction of depression, anxiety, and stress among university students: a cross-sectional study from bangladesh. *Journal of Health, Population and Nutrition*, 44(1):361, 2025.

[92] Ahmad Basheer Hassanat, Mohammad Ali Abbadi, Ghada Awad Altarawneh, and Ahmad Ali Alhasanat. Solving the problem of the k parameter in the knn classifier using an ensemble learning approach. *arXiv preprint arXiv:1409.0919*, 2014.

[93] Cornelia Herbert, Amira El Bolock, and Slim Abdennadher. How do you feel during the COVID-19 pandemic? a survey using psychological and linguistic self-report measures, and machine learning to investigate mental health, subjective experience, personality, and behaviour during the COVID-19 pandemic among university students. *BMC Psychology*, 9(1):1–23, 2021.

[94] Timothy O Hodson. Root mean square error (rmse) or mean absolute error (mae): When to use them or not. *Geoscientific Model Development Discussions*, 2022:1–10, 2022.

[95] Torsten Hothorn, Kurt Hornik, and Achim Zeileis. ctree: Conditional inference trees. *The comprehensive R archive network*, 8:1–34, 2015.

[96] Jeff Huarcaya-Victoria, David Villarreal-Zegarra, Angela Podestà, and María Alejandra Luna-Cuadros. Psychometric properties of a spanish version of the fear of covid-19 scale in general population of lima, peru. *International Journal of Mental Health and Addiction*, 20(1):249–262, 2022.

[97] Katrina Hueniken, Nibene Habib Somé, Mohamed Abdelhack, Graham Taylor, Tara Elton Marshall, Christine M Wickens, Hayley A Hamilton, Samantha Wells, Daniel Felsky, et al. Machine learning–based predictive modeling of anxiety and depressive

symptoms during 8 months of the covid-19 global pandemic: Repeated cross-sectional survey study. *JMIR mental health*, 8(11):e32876, 2021.

[98] Jehad Imlawi and Mohammad Alsharo. Evaluating classification accuracy: the impact of resampling and dataset size. *International Journal of Business Information Systems*, 24(1):91–101, 2017.

[99] Muhammad Irfan, Ahmad Shaf, Tariq Ali, Maryam Zafar, Saifur Rahman, Meiaad Ali I. Hendi, Shatha Abduh M. Baeshen, Maryam Mohammed Mastoor Maghfouri, Hailah Saeed Mohammed Alahmari, Ftimah Ahmed Ibrahim Shahhar, et al. An intelligent framework to measure the effects of covid-19 on the mental health of medical staff. *Plos one*, 18(6):e0286155, 2023.

[100] Md Akhtarul Islam, Sutapa Dey Barna, Hasin Raihan, Md Nafiul Alam Khan, and Md Tanvir Hossain. Depression and anxiety among university students during the covid-19 pandemic in bangladesh: A web-based cross-sectional survey. *PloS one*, 15(8):e0238162, 2020.

[101] Tahir Islam, Abdul Hameed Pitafi, Vikas Arya, Ying Wang, Naeem Akhtar, Shujaat Mubarik, and Liang Xiaobei. Panic buying in the covid-19 pandemic: A multi-country examination. *Journal of Retailing and Consumer Services*, 59:102357, 2021.

[102] Vikramaditya Jakkula. Tutorial on support vector machine (svm). *School of EECS, Washington State University*, 37(2.5):3, 2006.

[103] Mohamed F Jalloh, Wenshu Li, Rebecca E Bunnell, Kathleen A Ethier, Ann O'Leary, Kathy M Hageman, Paul Sengeh, Mohammad B Jalloh, Oliver Morgan, Sara Hersey, et al. Impact of ebola experiences and risk perceptions on mental health in sierra leone, july 2015. *BMJ global health*, 3(2), 2018.

[104] Andras Janosi, William Steinbrunn, Matthias Pfisterer, and Robert Detrano. Heart disease. uci machine learning repository. *UCI Machine Learning Repository*, 1988.

[105] Indra Prakash Jha, Raghav Awasthi, Ajit Kumar, Vibhor Kumar, and Tavpritesh Sethi. Learning the mental health impact of covid-19 in the united states with explainable artificial intelligence: observational study. *JMIR mental health*, 8(4):e25097, 2021.

[106] Mario Jojoa, Esther Lazaro, Begonya Garcia-Zapirain, Marino J Gonzalez, and Elena Urizar. The impact of covid 19 on university staff and students from iberoamerica: Online

learning and teaching experience. *International journal of environmental research and public health*, 18(11):5820, 2021.

[107] Bahram Jowkar, Oddgeir Friborg, and Odin Hjemdal. Cross-cultural validation of the resilience scale for adults (rsa) in iran. *Scandinavian journal of psychology*, 51(5):418–425, 2010.

[108] K Kalaiselvi and M Deepika. Machine learning for healthcare diagnostics. In *Machine learning with health care perspective: machine learning and healthcare*, pages 91–105. Springer, 2020.

[109] Sweta Kaman and Ishani Vajpai. Predicting depression among it sector employees in india during covid-19 pandemic using deep neural network and supervised machine learning classifiers: A survey approach. *Editorial Board*, 9(8):23, 2020.

[110] Robert Karasek, Chantal Brisson, Norito Kawakami, Irene Houtman, Paulien Bongers, and Benjamin Amick. The job content questionnaire (jcq): an instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of occupational health psychology*, 3(4):322, 1998.

[111] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[112] James M Keller, Michael R Gray, and James A Givens. A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics*, (4):580–585, 1985.

[113] Corey LM Keyes, Marié Wissing, Johan P Potgieter, Michael Temane, Annamarie Kruger, and Sinette Van Rooy. Evaluation of the mental health continuum–short form (mhc–sf) in setswana-speaking south africans. *Clinical psychology & psychotherapy*, 15(3):181–192, 2008.

[114] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 53(8):5455–5516, 2020.

[115] Shakir Khan and Salihah Alqahtani. Hybrid machine learning models to detect signs of depression. *Multimedia Tools and Applications*, 83(13):38819–38837, 2024.

[116] Nida Khateeb and Muhammad Usman. Efficient heart disease prediction system using k-nearest neighbor classification technique. In *Proceedings of the international conference on big data and internet of thing*, pages 21–26, 2017.

[117] Siamak Khodarahimi. Hope and flourishing in an iranian adults sample: Their contributions to the positive and negative emotions. *Applied Research in Quality of Life*, 8(3):361–372, 2013.

[118] Kenji Kira and Larry A Rendell. A practical approach to feature selection. In *Machine learning proceedings 1992*, pages 249–256. Elsevier, 1992.

[119] Daphne Koller and Mehran Sahami. Toward optimal feature selection. Technical report, Stanford InfoLab, 1996.

[120] Vipul Kothari, J Anuradha, Shreyak Shah, and Prerit Mittal. A survey on particle swarm optimization in feature selection. In *International Conference on Computing and Communication Systems*, pages 192–201. Springer, 2011.

[121] Reshmy Krishnan, Shantha Kumari, Ali Al Badi, Shermina Jeba, and Menila James. Predictive machine learning model for mental health issues in higher education students due to covid-19 using hads assessment. *Arab Gulf Journal of Scientific Research*, 42(4):1472–1490, 2024.

[122] Kurt Kroenke, Robert L. Spitzer, and Janet B. W. Williams. The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical Care*, 41(11):1284–1292, 2003.

[123] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9):606–613, 2001.

[124] Kurt Kroenke, Robert L Spitzer, JB Williams, Mark Linzer, Steven R Hahn, FV deGruy 3rd, and David Brody. Physical symptoms in primary care. predictors of psychiatric disorders and functional impairment. *Archives of family medicine*, 3(9):774–779, 1994.

[125] Kurt Kroenke, Tara W. Strine, Robert L. Spitzer, Janet B. W. Williams, Joyce T. Berry, and Ali H. Mokdad. The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, 114(1–3):163–173, April 2009.

[126] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645, September 2009.

[127] Akshi Kumar. Machine learning for psychological disorder prediction in indians during covid-19 nationwide lockdown. *Intelligent Decision Technologies*, 15(1):161–172, 2021.

[128] Prashant Kumar, Sushil Chandra, et al. Prediction and comparison of psychological health during covid-19 among indian population and rajyoga meditators using machine learning algorithms. *Procedia computer science*, 218:697–705, 2023.

[129] Anasua Kundu, Rui Fu, Daniel Grace, Carmen H Logie, Alex Abramovich, Bruce Baskerville, Christina Yager, Robert Schwartz, Nicholas Mitsakakis, Lynn Planinac, et al. Correlates of wanting to seek help for mental health and substance use concerns by sexual and gender minority young adults during the covid-19 pandemic: A machine learning analysis. *Plos one*, 17(11):e0277438, 2022.

[130] Liza M Kunjachen and R Kavitha. Comparative study on detection of heart disease using knn algorithm. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, pages 14–19. IEEE, 2022.

[131] Theodoros Kyriazos, Michalis Galanakis, Eirini Karakasidou, and Anastassios Stalikas. Early covid-19 quarantine: A machine learning approach to model what differentiated the top 25% well-being scorers. *Personality and Individual Differences*, 181:110980, 2021.

[132] Theodoros A Kyriazos, Anastassios Stalikas, Konstantina Prassa, and Vasiliki Yotsidi. A 3-faced construct validation and a bifactor subjective well-being model using the scale of positive and negative experience, greek version. *Psychology*, 9(5):1143–1175, 2018.

[133] Hannah M Layman, Ingibjorg Eva Thorisdottir, Thorhildur Halldorsdottir, Inga Dora Sigfusdottir, John P Allegrante, and Alfgeir Logi Kristjansson. Substance use among youth during the covid-19 pandemic: a systematic review. *Current psychiatry reports*, 24(6):307–324, 2022.

[134] Scikit Learn. Random forest classifier. *línea]. Disponible en: https://scikit-learn. org/stable/modules/generated/sklearn. ensemble. RandomForestClassifier. html [Accedido 14 Dic. 2021]*, 2016.

[135] Nicholas Leigh-Hunt, David Bagguley, Kristin Bash, Victoria Turner, Stephen Turnbull, Nicole Valtorta, and Woody Caan. An overview of systematic reviews on the public health consequences of social isolation and loneliness. *Public health*, 152:157–171, 2017.

[136] Daphna Levinson and Giora Kaplan. What does self rated mental health represent. *Journal of Public Health Research*, 3(3):jphr–2014, 2014.

[137] Huining Li, Enhao Zheng, Zijian Zhong, Chenhan Xu, Nicole Roma, Steven Lamkin, Tania T Von Visger, Yu-Ping Chang, and Wenyao Xu. Stress prediction using micro-ema and machine learning during covid-19 social isolation. *Smart Health*, 23:100242, 2022.

[138] Umesh Kumar Lilhore, Surjeet Dalal, Neeraj Varshney, Yogesh Kumar Sharma, KBV Brahma Rao, VVR Maheswara Rao, Roobaea Alroobaea, Sarita Simaiya, Martin Margala, and Prasun Chakrabarti. Prevalence and risk factors analysis of postpartum depression at early stage using hybrid deep learning model. *Scientific reports*, 14(1):4533, 2024.

[139] Kevin Linka, Mathias Peirlinck, Amelie Schäfer, Oguz Ziya Tikenogullari, Alain Goriely, and Ellen Kuhl. Effects of b. 1.1. 7 and b. 1.351 on covid-19 dynamics: A campus reopening study. *Archives of computational methods in engineering*, 28(6):4225–4236, 2021.

[140] Yang Liu, Ya-Nan Xie, Wen-Gang Li, Xin He, Hong-Gu He, Long-Biao Chen, and Qu Shen. A machine learning-based risk prediction model for post-traumatic stress disorder during the covid-19 pandemic. *Medicina*, 58(12):1704, 2022.

[141] Edward H Livingston. The mean and standard deviation: what does it all mean? *Journal of Surgical Research*, 119(2):117–123, 2004.

[142] Wei-Yin Loh. Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23, 2011.

[143] Lorena Cecilia López Steinmetz, Margarita Sison, Rustam Zhumagambetov, Juan Carlos Godoy, and Stefan Haufe. Machine learning models predict the emergence of depression in argentinean college students during periods of covid-19 quarantine. *Frontiers in Psychiatry*, 15:1376784, 2024.

[144] Sunny S Lou, Hanyang Liu, Benjamin C Warner, Derek Harford, Chenyang Lu, and Thomas Kannampallil. Predicting physician burnout using clinical activity logs: model performance and lessons learned. *Journal of biomedical informatics*, 127:104015, 2022.

[145] Sydney H Lovibond. Manual for the depression anxiety stress scales. *Sydney psychology foundation*, 1995.

[146] Wisit Lumchanow and Sakol Udomsiri. Image classification of malaria using hybrid algorithms: convolutional neural network and method to find appropriate k for k-nearest neighbor. *Indonesian Journal of Electrical Engineering and Computer Science*, 16(1):382–388, 2019.

[147] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[148] Abdalla Mahgoub. Diabetes predication system with knn algorithm. `https://www.kaggle.com/datasets/abdallamahgoub/diabetes`, n.d. Kaggle Dataset.

[149] Sultan Mahmud, Md Mohsin, Abdul Muyeed, Shaila Nazneen, Md Abu Sayed, Nabil Murshed, Tajrin Tahrin Tonmon, and Ariful Islam. Machine learning approaches for predicting suicidal behaviors among university students in bangladesh during the covid-19 pandemic: A cross-sectional study. *Medicine*, 102(28):e34285, 2023.

[150] Arun Malik, Mohammad Shabaz, Evans Asenso, et al. Machine learning based model for detecting depression during covid-19 crisis. *Scientific African*, 20:e01716, 2023.

[151] Jayawant N Mandrekar. Receiver operating characteristic curve in diagnostic test assessment. *Journal of thoracic oncology*, 5(9):1315–1316, 2010.

[152] Nikhil Marriwala, Deepti Chaudhary, et al. A hybrid model for depression detection using deep learning. *Measurement: Sensors*, 25:100587, 2023.

[153] Aleix M Martinez and Avinash C Kak. Pca versus lda. *IEEE transactions on pattern analysis and machine intelligence*, 23(2):228–233, 2001.

[154] Gonzalo Martinez-Ales, Alejandro Szmulewicz, Teresa Lopez-Cuadrado, Christopher N Morrison, Katherine M Keyes, and Ezra S Susser. Suicide following the covid-19 pandemic outbreak: variation across place, over time, and across sociodemographic groups. a systematic integrative review. *Current psychiatry reports*, 25(7):283–300, 2023.

[155] Christina Maslach and Susan E Jackson. The measurement of experienced burnout. *Journal of organizational behavior*, 2(2):99–113, 1981.

[156] Brian W Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, 1975.

[157] Dastan Maulud and Adnan M Abdulazeez. A review on linear regression comprehensive in machine learning. *Journal of applied science and technology trends*, 1(2):140–147, 2020.

[158] Robert G Maunder, William J Lancee, Kenneth E Balderson, Jocelyn P Bennett, Bjug Borgundvaag, Susan Evans, Christopher MB Fernandes, David S Goldbloom, Mona Gupta, Jonathan J Hunter, et al. Long-term psychological and occupational effects of providing hospital healthcare during sars outbreak. *Emerging infectious diseases*, 12(12):1924, 2006.

[159] MG Mazza, M Palladini, S Poletti, and F Benedetti. P. 0086 clinical and psychopathological predictors of fatigue in covid-19 survivors: a machine learning study. *European Neuropsychopharmacology*, 53:S60, 2021.

[160] G. C. McDonald. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):93–100, 2009.

[161] Zhang Min and Duan Kai-fei. Improved research to k-means initial cluster centers. In *2015 Ninth international conference on frontier of computer science and technology*, pages 349–353. IEEE, 2015.

[162] KM Mitravinda, Devika S Nair, and Gowri Srinivasa. Mental health in tech: Analysis of workplace risk factors and impact of covid-19. *SN computer science*, 4(2):197, 2023.

[163] Amera Mojahed, Judith T Mack, Lina Specht, Vania Sandoz, and Susan Garthus-Niegel. Impact of the covid-19 pandemic on violence against children: A narrative review. *Current psychiatry reports*, 25(10):533–543, 2023.

[164] Marina Mondo, Cristina Sechi, and Cristina Cabras. Psychometric evaluation of three versions of the italian perceived stress scale. *Current Psychology*, 40(4):1884–1892, 2021.

[165] Marco Antonio Montufar-Benítez, Jaime Mora-Vargas, Joselito Medina-Marín, Héctor Rivera-Gómez, and Sergio Blas Ramírez-Reyna. Application of machine learning using unsupervised models on databases of university students under the effect of covid-19. In *Data Analysis Techniques and Applications*, pages 387–406. Springer, 2026.

[166] Emma Motrico, Rena Bina, Sara Domínguez-Salas, Vera Mateus, Yolanda Contreras-García, Mercedes Carrasco-Portiño, Erilda Ajaz, Gisele Apter, Andri Christoforou, Pelin Dikmen-Yildiz, et al. Impact of the covid-19 pandemic on perinatal mental health (riseup-ppd-covid-19): protocol for an international prospective cohort study. *BMC Public Health*, 21(1):368, 2021.

[167] Rajendrani Mukherjee, Srestha Sadhu, and Aurghyadip Kundu. Heart disease detection using feature selection based knn classifier. In *Proceedings of Data Analytics and Management: ICDAM 2021, Volume 1*, pages 577–585. Springer, 2022.

[168] Sreerama K Murthy. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, 2(4):345–389, 1998.

[169] Alexey Natekin and Alois Knoll. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7:21, 2013.

[170] Md Iqbal Hossain Nayan, M Sheikh Giash Uddin, Md Ismail Hossain, Md Mohibul Alam, Maliha Afroj Zinnia, Iqramul Haq, Md Moshiur Rahman, Rejwana Ria, and Md Injamul Haq Methun. Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among university students in bangladesh: A result of the first wave of the covid-19 pandemic. *Asian Journal of Social Health and Behavior*, 5(2):75–84, 2022.

[171] Tasnim Niger, Hasanur Rayhan, Rashidul Islam, Kazi Asif Abdullah Noor, and Kamrul Hasan. Framework for behavioral disorder detection using machine learning and application of virtual cognitive behavioral therapy in covid-19 pandemic. *arXiv preprint arXiv:2204.13900*, 2022.

[172] Giovanna Nigro. Contributo alla taratura italiana della scala di controllo interno-esterno (ie) di rotter. *Giunti Organizzazioni Speciali*, 1983.

[173] Roghieh Nooripour, Simin Hosseinian, Abir Jaafar Hussain, Mohsen Annabestani, Ameer Maadal, Laurel E Radwin, Peyman Hassani-Abharian, Nikzad Ghanbari

Pirkashani, and Abolghasem Khoshkonesh. How resiliency and hope can predict stress of covid-19 by mediating role of spiritual well-being based on machine learning. *Journal of religion and health*, 60(4):2306–2321, 2021.

[174] Charis Ntakolia, Dimitrios Priftis, Mariana Charakopoulou-Travlou, Ioanna Rannou, Konstantina Magklara, Ioanna Giannopoulou, Konstantinos Kotsis, Aspasia Serdari, Emmanouil Tsalamanios, Aliki Grigoriadou, et al. An explainable machine learning approach for covid-19's impact on mood states of children and adolescents during the first lockdown in greece. In *Healthcare*, volume 10, page 149. MDPI, 2022.

[175] Sinan Okur, Begüm Satici, M Engin Deniz, and Seydi Ahmet Satici. Dispositional mindfulness, covid-19 burnout, and life satisfaction: Examining longitudinal mediation. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, 42(2):495–506, 2024.

[176] Itsuki Osawa, Tadahiro Goto, Takahiro Tabuchi, Hayami K Koga, and Yusuke Tsugawa. Machine-learning approaches to identify determining factors of happiness during the covid-19 pandemic: retrospective cohort study. *BMJ open*, 12(12):e054862, 2022.

[177] Keiron O'shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[178] Augustine Osman, Courtney L Bagge, Peter M Gutierrez, Lisa C Konick, Beverly A Kopper, and Francisco X Barrios. The suicidal behaviors questionnaire-revised (sbq-r): validation with clinical and nonclinical samples. *Assessment*, 8(4):443–454, 2001.

[179] Christian Otte, Stefan M Gold, Brenda W Penninx, Carmine M Pariante, Amit Etkin, Maurizio Fava, David C Mohr, and Alan F Schatzberg. Major depressive disorder. *Nature reviews Disease primers*, 2(1):1–20, 2016.

[180] Rutvija Pandya and Jayati Pandya. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications*, 117(16), 2015.

[181] Betty Pfefferbaum, Phebe Tucker, Vijayabharathi Ekambaram, and Richard L Van Horn. Children's sleep and mental health during the covid-19 pandemic. *Current Psychiatry Reports*, 25(12):847–856, 2023.

[182] Betty Pfefferbaum and Richard L Van Horn. Physical activity and sedentary behavior in children during the covid-19 pandemic: Implications for mental health. *Current psychiatry reports*, 24(10):493–501, 2022.

[183] Roland Pongou, Bright Opoku Ahinkorah, Sophie Maltais, Marie Claude Mabeu, Anshuman Agarwal, and Sanni Yaya. Psychological distress during the COVID-19 pandemic in canada. *PLOS ONE*, 17(11):e0277238, November 2022.

[184] Piero Porcelli and Jenny Guidi. The clinical utility of the diagnostic criteria for psychosomatic research: a review of studies. *Psychotherapy and psychosomatics*, 84(5):265–272, 2015.

[185] Liana CL Portugal, Camila Monteiro Fabricio Gama, Raquel Menezes Gonçalves, Mauro Vitor Mendlowicz, Fátima Smith Erthal, Izabela Mocaiber, Konstantinos Tsirlis, Eliane Volchan, Isabel Antunes David, Mirtes Garcia Pereira, et al. Vulnerability and protective factors for ptsd and depression symptoms among healthcare workers during covid-19: A machine learning approach. *Frontiers in Psychiatry*, 12:752870, 2022.

[186] AS Prakaash, K Sivakumar, B Surendiran, S Jagatheswari, and K Kalaiarasi. Design and development of modified ensemble learning with weighted rbm features for enhanced multi-disease prediction model. *New Generation Computing*, 40(4):1241–1279, 2022.

[187] Tracy A Prout, Sigal Zilcha-Mano, Katie Aafjes-van Doorn, Vera Békés, Isabelle Christman-Cohen, Kathryn Whistler, Thomas Kui, and Mariagrazia Di Giuseppe. Identifying predictors of psychological distress during covid-19: a machine learning approach. *Frontiers in psychology*, 11:586202, 2020.

[188] Radwan Qasrawi, Malak Amro, Stephanny VicunaPolo, Diala Abu Al-Halawa, Hazem Agha, Rania Abu Seir, Maha Hoteit, Reem Hoteit, Sabika Allehdan, Nouf Behzad, et al. Machine learning techniques for predicting depression and anxiety in pregnant and postpartum women during the covid-19 pandemic: a cross-sectional regional study. *F1000Research*, 11:390, 2022.

[189] Meghan Quinn, Samuel Dickinson, and Shram Shukla. Training military psychiatrists to adapt and overcome: how covid-19 highlighted the unique flexibility of military psychiatry in training and in the fleet. *Current psychiatry reports*, 24(8):369–374, 2022.

[190] Asmaa H Rabie, Alaa M Mohamed, MA Abo-Elsoud, and Ahmed I Saleh. A new covid-19 diagnosis strategy using a modified knn classifier. *Neural Computing and Applications*, 35(23):17349–17373, 2023.

[191] Robert H Remien, Michael J Stirratt, Nadia Nguyen, Reuben N Robbins, Andrea N Pala, and Claude A Mellins. Mental health and hiv/aids: the need for an integrated response. *Aids*, 33(9):1411–1420, 2019.

[192] Ziyuan Ren, Yaodong Xin, Junpeng Ge, Zheng Zhao, Dexiang Liu, Roger CM Ho, and Cyrus SH Ho. Psychological impact of covid-19 on college students after school reopening: a cross-sectional study based on machine learning. *Frontiers in Psychology*, 12:641806, 2021.

[193] Douglas Reynolds. Gaussian mixture models. In *Encyclopedia of biometrics*, pages 827–832. Springer, 2015.

[194] Mostafa Rezapour and Lucas Hansen. A machine learning analysis of covid-19 mental health data. *Scientific reports*, 12(1):14965, 2022.

[195] Reyhan Achmad Rizal, Nurlela Octavia Purba, Lidya Aprilla Siregar, Kristina Sinaga, and Nur Azizah. Analysis of tuberculosis (tb) on x-ray image using surf feature extraction and the k-nearest neighbor (knn) classification method. *Jaict*, 5(2):9, 2020.

[196] Indrani Roy, Rinita Shai, Arijit Ghosh, Anirban Bej, and Soumen Kumar Pati. Cowarriornet: A novel deep-learning framework for covid-19 detection from chest x-ray images. *New Generation Computing*, 40(4):961–985, 2022.

[197] L. Rubini, S. P., and E. P. Chronic kidney disease [dataset]. `https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease`, n.d. UCI Machine Learning Repository.

[198] Hans-Jürgen Rumpf, Christian Meyer, Ulfert Hapke, and Ulrich John. Screening for mental health: validity of the mhi-5 using dsm-iv axis i psychiatric disorders as gold standard. *Psychiatry research*, 105(3):243–253, 2001.

[199] Seyed-Ali Sadegh-Zadeh, Mahboobe Bahrami, Amirreza Najafi, Meisam Asgari-Ahi, Russell Campion, and Amir M Hajiyavand. Evaluation of covid-19 pandemic on components of social and mental health using machine learning, analysing united states data in 2020. *Frontiers in Psychiatry*, 13:933439, 2022.

[200] S Rasoul Safavian and David Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.

[201] Santosh Kumar Sahoo, Gopinath Palai, Baraa Riyadh Altahan, Sk Hasane Ahammad, P Poorna Priya, Md Amzad Hossain, and Ahmed Nabih Zaki Rashed. An optimized deep learning approach for the prediction of social distance among individuals in public places during pandemic. *New Generation Computing*, 41(1):135–154, 2023.

[202] Afef Saidi, Slim Ben Othman, and Slim Ben Saoud. Hybrid cnn-svm classifier for efficient depression detection system. In *2020 4th international conference on advanced systems and emergent technologies (IC_ASET)*, pages 229–234. IEEE, 2020.

[203] Hanaa Salem, Mahmoud Y Shams, Omar M Elzeki, Mohamed Abd Elfattah, Jehad F. Al-Amri, and Shaima Elnazer. Fine-tuning fuzzy knn classifier based on uncertainty membership for the medical diagnosis of diabetes. *Applied Sciences*, 12(3):950, 2022.

[204] Aarón Salinas-Rodríguez, Betty Manrique-Espinoza, Isaac Acosta-Castillo, Martha Ma Téllez-Rojo, Aurora Franco-Núñez, Luis Miguel Gutiérrez-Robledo, and Ana Luisa Sosa-Ortiz. Validation of a cutoff for the depression scale of the center for epidemiologic studies, brief version (cesd-7). *Salud publica de Mexico*, 55(3):267–274, 2013.

[205] Arthur L Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229, 1959.

[206] Damian F Santomauro, Ana M Mantilla Herrera, Jamileh Shadid, Peng Zheng, Charlie Ashbaugh, David M Pigott, Cristiana Abbafati, Christopher Adolph, Joanne O Amlag, Aleksandr Y Aravkin, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the covid-19 pandemic. *The Lancet*, 398(10312):1700–1712, 2021.

[207] Irwin G Sarason, Henry M Levine, Robert B Basham, and Barbara R Sarason. Assessing social support: the social support questionnaire. *Journal of personality and social psychology*, 44(1):127, 1983.

[208] Sandeep Kumar Satapathy, Shreyaa Saravanan, Shruti Mishra, and Sachi Nandan Mohanty. A comparative analysis of multidimensional covid-19 poverty determinants: An observational machine learning approach. *New generation computing*, 41(1):155–184, 2023.

[209] Wilmar B Schaufeli, Arnold B Bakker, and Marisa Salanova. The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and psychological measurement*, 66(4):701–716, 2006.

[210] Ralf Schwarzer, John Mueller, and Esther Greenglass. Assessment of perceived general self-efficacy on the internet: data collection in cyberspace. *Anxiety, stress and coping*, 12(2):145–161, 1999.

[211] Philip Sedgwick. Pearson's correlation coefficient. *Bmj*, 345, 2012.

[212] Jana Shafi and Md Sahidur Rahman. Application of machine learning models for detecting mental stress during covid-19 pandemic. *Available at SSRN 3778839*, 2021.

[213] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro. Fatigue severity scale (fss). In *STOP, THAT and one hundred other sleep scales*, pages 167–168. Springer, 2011.

[214] Saleh Shahinfar, Paul Meek, and Greg Falzon. "how many images do i need?" understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. *Ecological Informatics*, 57:101085, 2020.

[215] K Shailaja, Banoth Seetharamulu, and MA Jabbar. Machine learning in healthcare: A review. In *2018 Second international conference on electronics, communication and aerospace technology (ICECA)*, pages 910–914. IEEE, 2018.

[216] Raahil Sheikh, Lavleen Sharma, Disha Singh, Pragyan Priyadarshini Rout, Shreya Das, Tejaswini Yadav, et al. From fear to forecast: Machine learning solutions for mass psychosis in crisis scenarios. *Cureus Journals*, 3(1), 2026.

[217] X Shen, L Xu, and H Liu. The application of posttraumatic diagnostic scale-chinese in an earthquake-exposed sample. *Chinese Journal of Health Psychology*, 22(10.13342), 1849.

[218] Mai Shouman, Tim Turner, and Rob Stocker. Applying k-nearest neighbour in diagnosing heart disease patients. *International Journal of Information and Education Technology*, 2(3):220–223, 2012.

[219] Marija Simjanoski et al. Lifestyle predictors of depression and anxiety during COVID-19: A machine learning approach. *Trends in Psychiatry and Psychotherapy*, 44, 2022.

[220] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes. Pima indians diabetes database. `https://www.kaggle.com/datasets/uciml/pima-i`

ndians-diabetes-database, n.d. Kaggle Dataset (originally from UCI Machine Learning Repository).

[221] A'yunin Sofro, Khusnia Nurul Khikmah, and Asri Maharani. Depression, anxiety, and stress among indonesians during covid-19 pandemic: Multiclass prediction machine learning approach. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 12(1):73–87, 2026.

[222] Margarita Sordo and Qing Zeng. On sample size and classification accuracy: A performance comparison. In *International Symposium on Biological and Medical Data Analysis*, pages 193–201. Springer, 2005.

[223] Donald F Specht et al. A general regression neural network. *IEEE transactions on neural networks*, 2(6):568–576, 1991.

[224] Charles D Spielberger. State-trait anxiety inventory for adults. 1983.

[225] Charles Donald Spielberger. Manual for the state-trait anxiety inventory (self-evaluation questionnaire). *(No Title)*, 1970.

[226] Robert L Spitzer, Kurt Kroenke, Janet BW Williams, and Bernd Löwe. A brief measure for assessing generalized anxiety disorder: the gad-7. *Archives of internal medicine*, 166(10):1092–1097, 2006.

[227] David RB Stockwell and A Townsend Peterson. Effects of sample size on accuracy of species distribution models. *Ecological modelling*, 148(1):1–13, 2002.

[228] Jill C Stoltzfus. Logistic regression: a brief primer. *Academic emergency medicine*, 18(10):1099–1104, 2011.

[229] Sridevi Subbiah, Kalaiarasi Sonai Muthu Anbananthen, Saranya Thangaraj, Subarmaniam Kannan, and Deisy Chelliah. Intrusion detection technique in wireless sensor network using grid search random forest with boruta feature selection algorithm. *Journal of Communications and Networks*, 24(2):264–273, 2022.

[230] Shiliang Sun and Rongqing Huang. An adaptive k-nearest neighbor algorithm. In *2010 seventh international conference on fuzzy systems and knowledge discovery*, volume 1, pages 91–94. IEEE, 2010.

[231] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel S Emer. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329, 2017.

[232] Jasmine Tall and Matthew Biel. The effects of social determinants of health on child and family mental health: implications of the covid-19 pandemic and beyond. *Current psychiatry reports*, 25(9):387–394, 2023.

[233] Yigitcanlar Tan, Kankanamge Nayomi, Alexander Preston, Gill Palvinderjit Singh, Rezayee Maqsood, Ostadnia Mahsan, Bo Xia, and Ioppolo Giuseppe. How can social media analytics assist authorities in pandemic-related policy decisions? insights from australian states and territories. *Health Information Science and Systems*, 8(1), 2020.

[234] June P Tangney, Angie Luzio Boone, and Roy F Baumeister. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. In *Self-regulation and self-control*, pages 173–212. Routledge, 2018.

[235] Xiaohui Tao, Oliver Chi, Patrick J Delaney, Lin Li, and Jiajin Huang. Detecting depression using an ensemble classifier based on quality of life scales. *Brain Informatics*, 8(1):2, 2021.

[236] Emanuele Tauro, Alessandra Gorini, Chiara Caglio, Paola Gabanelli, and Enrico Gianluca Caiani. Covid-19 and mental disorders in healthcare personnel: A novel framework to develop personas from an online survey. *Journal of biomedical informatics*, 126:103993, 2022.

[237] M Thomason, A Graham, E Sullivan, and MI van den Heuvel. Covid-19 and perinatal experiences study. *Open Science Framework*, 2020.

[238] Zhanxiao Tian, Wei Qu, Yanli Zhao, Xiaolin Zhu, Zhiren Wang, Yunlong Tan, Ronghuan Jiang, and Shuping Tan. Predicting depression and anxiety of chinese population during covid-19 in psychological evaluation data by xgboost. *Journal of Affective Disorders*, 323:417–425, 2023.

[239] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

[240] Roman Timofeev. Classification and regression trees (cart) theory and applications. *Humboldt University, Berlin*, 54:48, 2004.

[241] Dimple Tiwari and Bharti Nagpal. Keaht: A knowledge-enriched attention-based hybrid transformer model for social sentiment analysis. *New Generation Computing*, 40(4):1165–1202, 2022.

[242] Gul Tokdemir. Software professionals during the covid-19 pandemic in turkey: Factors affecting their mental well-being and work engagement in the home-based work setting. *Journal of Systems and Software*, 188:111286, 2022.

[243] Mickey Trockel, Bryan Bohman, Emi Lesure, Maryam S Hamidi, Dana Welle, Laura Roberts, and Tait Shanafelt. A brief instrument to assess both burnout and professional fulfillment in physicians: reliability and validity, including correlation with self-reported medical errors, in a sample of resident and practicing physicians. *Academic Psychiatry*, 42(1):11–24, 2018.

[244] TA Tuan, T Dai An, NH Nghia, DTT Loan, et al. An analysis of machine learning for detecting depression, anxiety, and stress of recovered covid-19 patients. *Journal of Human Earth and Future*, 5(1):1–18, 2024.

[245] María Tubío-Fungueiriño, Eva Cernadas, Óscar F Gonçalves, Cinto Segalas, Sara Bertolín, Lorea Mar-Barrutia, Eva Real, Manuel Fernández-Delgado, Jose M Menchón, Sandra Carvalho, et al. Viability study of machine learning-based prediction of covid-19 pandemic impact in obsessive-compulsive disorder patients. *Frontiers in Neuroinformatics*, 16:807584, 2022.

[246] Shahadat Uddin, Ibtisham Haque, Haohui Lu, Mohammad Ali Moni, and Ergun Gide. Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction. *Scientific Reports*, 12(1):6256, 2022.

[247] Marcel Van Gerven and Sander Bohte. Artificial neural networks as models of neural information processing, 2017.

[248] Bhanu Verma, Sonam Gupta, and Lipika Goel. A neural network based hybrid model for depression detection in twitter. In *International Conference on Advances in Computing and Data Sciences*, pages 164–175. Springer, 2020.

[249] Sami Virpioja. Birch: Balanced iterative reducing and clustering using hierarchies, 2008.

[250] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[251] T. Vos et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The Lancet*, 388(10053):1545–1602, October 2016.

[252] Chongying Wang, Hong Zhao, and Haoran Zhang. Chinese college students have higher anxiety in new semester of online learning during covid-19: a machine learning approach. *Frontiers in psychology*, 11:587413, 2020.

[253] Lishan Wang. Research and implementation of machine learning classifier based on knn. In *IOP Conference Series: Materials Science and Engineering*, volume 677, page 052038. IOP publishing, 2019.

[254] Min Wang, Ying-Yi Zhang, and Fan Min. Appendicitis dataset. `https://www.kaggle.com/datasets/timrie/appendicitis`, n.d. Kaggle Dataset.

[255] Xiaofeng Wang, Hu Li, Chuanyong Sun, Xiumin Zhang, Tan Wang, Chenyu Dong, and Dongyang Guo. Prediction of mental health in medical workers during covid-19 based on machine learning. *Frontiers in public health*, 9:697850, 2021.

[256] Daniel S Weiss. The impact of event scale: revised. In *Cross-cultural assessment of psychological trauma and PTSD*, pages 219–238. Springer, 2007.

[257] World Health Organization. Ebola virus disease. Online. Accessed: Sep. 04, 2023.

[258] World Health Organization. Hiv. Online. Accessed: Sep. 05, 2023.

[259] World Health Organization. Zika virus. Online. Accessed: Sep. 04, 2023.

[260] World Health Organization. *Depression and other common mental disorders: global health estimates*. World Health Organization, 2017.

[261] Ya-Ning Xie. A preliminary study on the reliability and validity of the simple coping style scale. *Chin J Clin Psychol*, 6(2):114–115, 1998.

[262] Gaurav Yadav and Mohammad Ubaidullah Bokhari. Hybrid classifier for optimizing mental health prediction: feature engineering and fusion technique. *International Journal of Mental Health and Addiction*, pages 1–41, 2024.

[263] Takashi Yoshioka, Ryo Okubo, Takahiro Tabuchi, Satomi Odani, Tomohiro Shinozaki, and Yusuke Tsugawa. Factors associated with serious psychological distress during the

covid-19 pandemic in japan: a nationwide cross-sectional internet-based study. *BMJ open*, 11(7):e051115, 2021.

[264] Muhamad Saiful Bahri Yusoff. Psychometric properties of the depression anxiety stress scale in a sample of medical degree applicants. *International Medical Journal*, 20(3):295–300, 2013.

[265] Mohammed Zakariah, Muna Al-Razgan, and Taha Alfakih. Pathological voice classification using meel features and svm-tabnet model. *Speech Communication*, 162:103100, 2024.

[266] Zhenyu Zhao, Radhika Anand, and Mallory Wang. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In *2019 IEEE international conference on data science and advanced analytics (DSAA)*, pages 442–452. IEEE, 2019.

[267] Yiwei Zhou, Zejie Zhang, Qin Li, Guangyun Mao, and Zumu Zhou. Construction and validation of machine learning algorithm for predicting depression among home-quarantined individuals during the large-scale covid-19 outbreak: based on adaboost model. *BMC psychology*, 12(1):230, 2024.

[268] Zhaohe Zhou, Dan Luo, Bing Xiang Yang, and Zhongchun Liu. Machine learning-based prediction models for depression symptoms among chinese healthcare workers during the early covid-19 outbreak in 2020: a cross-sectional study. *Frontiers in Psychiatry*, 13:876995, 2022.

[269] Jiayu Zhu. Breast cancer prediction based on the machine learning algorithm lightgbm. *Applied and Computational Engineering*, 40:77–84, 2024.

[270] Chuanjun Zhuo, Gongying Li, Xiaodong Lin, Deguo Jiang, Yong Xu, Hongjun Tian, Wenqiang Wang, and Xueqin Song. The rise and fall of mri studies in major depressive disorder. *Translational psychiatry*, 9(1):335, 2019.

[271] Alexander Zien, Nicole Krämer, Sören Sonnenburg, and Gunnar Rätsch. The feature importance ranking measure. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 694–709. Springer, 2009.

[272] Hamad Zogan, Imran Razzak, Xianzhi Wang, Shoaib Jameel, and Guandong Xu. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web*, 25(1):281–304, 2022.

[273] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320, 2005.

[274] Md Sabab Zulfiker, Nasrin Kabir, Al Amin Biswas, Tahmina Nazneen, and Mohammad Shorif Uddin. An in-depth analysis of machine learning approaches to predict depression. *Current research in behavioral sciences*, 2:100044, 2021.

[275] William W Zung. A rating instrument for anxiety disorders. *Psychosomatics: Journal of Consultation and Liaison Psychiatry*, 1971.

[276] William WK Zung. A self-rating depression scale. *Archives of general psychiatry*, 12(1):63–70, 1965.

th

ORIGINALITY REPORT

9%	%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Proceedings of Data Analytics and Management", Springer Science and Business Media LLC, 2026 Publication	<1%
2	Keri Ka-Yee Wong, Kimberly Loke, Kyleigh Melville. "Reflections, Resilience, and Recovery: A qualitative study of the COVID-19 impact on an international general population's mental health and priorities for support", PsyArXiv, 2022 Publication	<1%
3	"Information System Design: Intelligent Healthcare Informatics", Springer Science and Business Media LLC, 2025 Publication	<1%
4	Stein, Kyle Richard. "Integrating Transformers for Cyber Defense Under Unseen and Evolving Threats With Deep Packet Inspection, Zero-Shot Backdoor Detection, and Continual Learning.", The University of West Florida Publication	<1%
5	Gaurav Aggarwal, Rishabh Dev Shukla, Innocent Ewean Davidson, Ramesh Bansal. "Integrated Technologies in Electrical, Electronics and Bio-technology Engineering", CRC Press, 2026 Publication	<1%
6	Mark E. Maruish. "Handbook of Psychological Assessment in Primary Care Settings", Routledge, 2017 Publication	<1%
7	B. M. Parashiva Murthy, Ik Hesti Agustin, Sowjanya Bharathi, Rayappan Lotus, S. Bathrinath, Rishabh Garg. "103Machine Learning Applications in Mental Health: Ensemble-Based Predictive Modeling for Depression and Anxiety detection", Walter de Gruyter GmbH, 2025 Publication	<1%
8	"Proceedings of the 3rd International Conference on Emerging Trends and Innovation (3rd ICETI)", Springer Science and Business Media LLC, 2026 Publication	<1%
9	Maria Grazia Criscuolo, Carla Di Stefano, Debra Rose Wilson, Arvind Kumar Shukla. "The Impact of COVID-19 on Psychological Disorders - Current and Future Interventions", Routledge, 2026 Publication	<1%

Biography – Supervisor



Dr. Sonika Dahiya is an Assistant Professor at Delhi Technological University, serving in the Department of Software Engineering since 2016. She has rich academic experience and has contributed significantly to teaching and research in areas such as fuzzy clustering, data mining, deep learning applications, image processing, and machine learning. She has guided undergraduate and postgraduate students in various research and academic projects and has published research papers in reputed national and international journals and conferences. Her research work focuses on developing intelligent and efficient computational models for solving real-world problems. She is actively involved in academic mentoring and continues to contribute to interdisciplinary research in emerging areas of computer science and artificial intelligence.

Biography – Student



Priyanka Arora is currently pursuing her doctoral degree from Delhi Technological University. She holds an MCA in Computer Science from Rukmini Devi Institute of Advanced Studies, Guru Gobind Singh Indraprastha University (2011), and a Bachelor's degree in Mathematics from Miranda House, Delhi University (2008). She has industry experience as a Software Developer and Team Lead, working in web development and database management using Java, PHP, and Python. She has strong academic and research interests in machine learning, deep learning, and mental health analytics. Her work focuses on developing intelligent and interpretable models for the detection of depression and anxiety using standardized psychological instruments. Alongside her academic pursuits, she has balanced her responsibilities as a mother of three children, viewing her doctoral journey as an example of lifelong learning and perseverance.

Current Psychology
https://doi.org/10.1007/s12144-025-07432-8

Unveiling the impact of dataset size on machine learning models for anxiety and depression prediction amid the COVID-19 pandemic: determining optimal data collection thresholds

Priyanka Arora[1] · **Sonika Dahiya**[1,2]

Accepted: 20 January 2025
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

Abstract
Our emotional, psychological, and social well-being are all parts of our mental health. An individual's routine can be disrupted and their mental is health affected by stress, despair, and anxiety. Mental health preservation and restoration are essential for each person as well as for communities and society as a whole. The COVID-19 pandemic has triggered a strong emotional and psychological reaction in many people, in addition to triggering a global health emergency. The pandemic's uncertainty, disruptions, and social changes have amplified stress, fear, and depression, which are common responses to crises. Data collection for the COVID-19-related depression and anxiety assessment was limited to online methods because of the ongoing pandemic. In the field of mental health evaluation, the application of machine learning techniques has emerged as a promising strategy for identifying and grasping anxiety and depression symptoms. This paper employed K-Nearest Neighbors (kNN), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM) techniques on the prevalence of anxiety and depression among Bangladeshi university students during the COVID-19 pandemic. This paper addresses how the accuracy of predictions made by various machine learning models is affected by the size of the datasets. The findings of this study illuminate the scalability and generalizability of different machine-learning methods. The findings validate that how accuracy of the models has consistently and significantly improved as the dataset size varies. The performance of classification models is further assessed using the F1 score, precision, and recall.

Keywords Machine learning · Depression · Anxiety · Dataset size · Classification · Performance analysis

Introduction

Machine learning is defined as a "Field of study that gives computers the capability to learn without being explicitly programmed" (Janiesch et al., 2021; Linda n.d.). The widespread utilization of machine learning algorithms for analyzing large datasets, discovering patterns, and making predictions have led to advancements and improvements in various fields such as industries, businesses, and almost any organization, including government agencies, e-commerce, telecommunications, banking, and financial services, healthcare, sales and marketing, transportation, social networking, and many others (Sarker, 2021; Sharifani & Amini, 2023; Sharma et al., 2021).

Considerable advancements have been observed in the field of healthcare. Utilizing machine learning with electronic health records can produce useful insights that can be used to streamline hospital operations, forecast the beginning of disease, and improve patient risk score systems (Callahan & Shah, 2017). It is problematic to decide how much data is sufficient to be fed into a machine-learning model for the interpretation of the results. Small dataset size leads to overfitting (Hawkins, 2004). As the training datasets get smaller, the models have fewer examples to learn from, increasing the risk of overfitting whereas a sufficient dataset size results in better classification performance when fed into machine learning models.

✉ Sonika Dahiya
sonika.dahiya@dtu.ac.in

[1] Department of Software Engineering, Delhi Technological University, Delhi, India

[2] Office of Software Engineering Department, AB4-101, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India

Published online: 31 January 2025

New Gener. Comput. (2025) 43:20
https://doi.org/10.1007/s00354-025-00305-5

NEW GENERATION COMPUTING

RESEARCH PAPER

COVID-19-Induced Depression and Anxiety Detection: A Systematic Literature Review of Machine Learning and Deep Learning Techniques

Priyanka Arora[1] · **Sonika Dahiya**[1]

Received: 7 November 2024 / Accepted: 2 October 2025
© The Author(s), under exclusive licence to The Japanese Society for Artificial Intelligence and Springer Nature Japan KK, part of Springer Nature 2025

Abstract

Epidemics or pandemics, like the current COVID-19 outbreak, present special obstacles to society. Leaders in several nations implemented policies that restricted social connections as a proactive attempt to stop the virus's spread. They intended to harness the power of social isolation as a strategy to mitigate the rapid increase in infections, effectively curving the trajectory of contamination. Shifts in societal conventions over this period have indirectly impacted the mental well-being of individuals. The United Nations (UN) prioritizes mental health, aiming to reduce premature mortality from non-communicable diseases by 2030. During the second wave of the COVID-19 epidemic, many people experienced significant life upheavals, including managing the death of family members, dealing with unstable employment, and managing elevated levels of emotional and financial stress. Owing to these circumstances, a sizeable portion of the populace has struggled with mental health conditions like anxiety, depression symptoms, PTSD, and mood disorders. In the past, pandemics or epidemics have had indirect consequences on mental health, leaving a sizable share of depressed people unidentified and untreated. Less is known about how emotion control affects distressed people's feelings during and after the pandemic. By performing a thorough analysis of the existing literature, the main goal of this study is to identify developments, limitations, obstacles, and possible directions for future research in the field of diagnosing depression and anxiety during and after the COVID-19 pandemic. A comprehensive analysis included sixty-six current research that focused on identifying mental health problems related to the epidemic. The selection criteria included studies that used psychological tools, such self-questionnaires, to detect a range of mental health conditions. The groundwork for tackling one key issue—the length of depression and anxiety symptoms—is laid forth by this systematic literature analysis's understanding of the landscape of pandemic-related mental health disorders. This understanding can greatly aid in developing focused prevention initiatives, which will ultimately lessen the burden of these disorders.

Extended author information available on the last page of the article

Published online: 21 October 2025

scientific reports

Check for updates

OPEN

A novel adaptive sigma KNN model for depression and anxiety detection following the COVID 19 pandemic

Priyanka Arora & Sonika Dahiya

Mental health disorders, such as depression and anxiety, are increasing, and thus, there is a necessity for accurate and effective detection. K-Nearest Neighbors (KNN) and extensions have been extensively used in disease detection. In this work, Adaptive Sigma KNN (ASKNN) is proposed, an improved version of KNN that adapts neighbor influence dynamically with an adaptive sigma parameter, enhancing predictive accuracy and stability in mental health classification. ASKNN is tested on our dataset and various public mental health and benchmark medical datasets. Performance has been measured through Accuracy, Precision, Recall, and F1-score, with cross-validation ensuring reliability. Friedman and Wilcoxon signed-rank tests were used to evaluate statistical significance, whereas Cohen's d and Cliff's Delta estimated effect size. ASKNN performed with 91.00% accuracy for depression and 84.50% for anxiety, with precision, recall, and F1-score of 0.91, 0.87, and 0.89 for depression, and 0.86, 0.79, and 0.82 for anxiety. AUC-ROC values additionally supported performance, with values of 0.95–0.91 (depression) and 0.93–0.78 (anxiety) between classes. Our statistical testing validates that ASKNN considerably surpasses baseline KNN models for all four measurements. In addition, the research resonates with the United Nations Sustainable Development Goal (SDG) 3, underlining the impact of artificial intelligence-based methodologies in optimizing mental health diagnoses. By increasing depression and anxiety identification accuracy, ASKNN promotes early intervention plans, ultimately benefiting international efforts to lower the global burden of mental illnesses.

Keywords Mental health assessment, Depression and anxiety detection, K-Nearest neighbors, Statistical analysis, Machine learning

The COVID-19 pandemic, which was announced by the World Health Organization (WHO) in March 2020, has affected world health enormously by infecting millions and bringing huge mortality[1]. Up to February 20, 2022, there were over 421 million laboratory-confirmed cases and 5.85 million related fatalities worldwide[2]. The highly contagious character of the virus, however, has created difficulties in attempts to limit its spread, leading to waves and the development of new strains[3]. Besides its primary medical impact, COVID-19 has had significant effects on mental health, such as increased rates of depression, anxiety, stress, and other psychiatric illnesses[4]. Such mental health issues have been linked to factors like social isolation, fear of infection, financial insecurities, and disruption of day-to-day life. Moreover, healthcare workers, particularly nurses, have suffered immense psychological trauma, including post-traumatic stress disorder (PTSD), emotional exhaustion, and moral distress due to their frontline response to the pandemic[5].

Mental illnesses such as depression and anxiety have emerged as a major public health concern, with millions of individuals being impacted worldwide[6]. Early and accurate identification is extremely crucial for early intervention and treatment. However, traditional diagnosis relies heavily on clinical interviews, which can be time-consuming, subjective, and largely unavailable in resource-constrained environments. To overcome these challenges, machine learning methods have emerged as increasingly crucial in mental illness studies since they can process massive data and identify advanced patterns that are associated with mental disorders[7]. The use of machine learning-based methods in the diagnosis of mental illnesses has the potential for expansion of early

Department of Software Engineering, Delhi Technological University, Office of Software Engineering Department, AB4-101, Shahbad Daulatpur, Main Bawana Road, New Delhi, Delhi 110042, India. email: sonika.dahiya@dtu.ac.in

2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit) | 979-8-3503-7971-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/GLOBALAISUMMIT62156.2024.10947873

Exploring Deep Learning Architectures for Mental Health Assessment: A Comparative Analysis

Priyanka Arora
Software Engineering
Delhi Technological University
New Delhi, India
priyankaarora1019@gmail.com

Sonika Dahiya
Software Engineering
Delhi Technological University
New Delhi, India
sonika.dahiya@dtu.ac.in

***Abstract*— The major consequences of the covid-19 pandemic have influenced the question of the ability of anger - especially depression and anxiety - to dominate public discourse. The aim of this research is to find advanced methods of understanding, illness and situation, using deep learning methods in the research of depression and stress. Five different models of deep learning are defined and classified in a motivational and anxiety framework based on three different data sets: dataset 1 (data set from student surveys that do not focus), dataset 2 (self-isolation and depression through separation) and dataset 3 (depression and anxiety of significant partners and childbirth). These models are called Neural Networks (NN), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), TabNet, and Convolutional Neural Networks (CNN). The interest is growing in exploring how deep learning is used for analyzing more complex data. Many studies have been done on the analysis of depression and anxiety using machine learning models, mainly for the dual classification. For the case of datasets 1 and 3, predictions with the highest accuracies for anxiety and depression were made by LSTM. Such results for dataset 2 were obtained using CNN. The work opens wide avenues in demonstrating deep learning applied to multiclass classification that could improve diagnosis and provide personalized mental health treatment approaches.**

Keywords— Depression; Anxiety; Mental Health; Deep Learning; Performance Analysis

I. INTRODUCTION

AI-based disease detection could change the face of medicine by using machine learning algorithms that process medical data that is huge in quantity to understand the patterns and trends that are difficult for human eye to decipher [1],[2]. In artificial intelligence, machine learning allows computers to acquire information and perform better tasks over time without the need for programming [3],[4]. Deep learning has a capability to draw inspirations from the composition and operation of a human brain [5], [6]. Deep learning has great promise in identifying anxiety and depression by evaluating data from psychological tests [7]. Patient Health Questionnaire (PHQ-9), Beck Depression Inventory, and Generalized Anxiety Disorder 7 (GAD-7) are some examples of typical psychological questionnaires that can be modeled by deep learning models using network technology. These models can handle complicated and high-dimensional data. A self-report tool called GAD-7 is used to assess the GAD symptoms have been persistent throughout the previous two weeks. It gauges seven different things: anxiety, difficulty relaxing, concern, restlessness, feeling tense or agitated, being unable to sit quiet, and feeling frightened, worried, or tense [8]. A self-report tool which is called PHQ-9 is employed to screen for, diagnose, and rank the severity of sadness. There are 9 questions about mood, interest in activities, eating patterns and changes in sleep which are related to depression that have been persistent during the past two weeks [9]. These instruments provide for systematic and measurable assessment of the symptoms of mental health. The data generated from these assessments may be fed into algorithms of deep learning, which could detect patterns and relationships that by conventional techniques of analyses may not be observable. What makes deep learning models very particularly appropriate for mental health diagnostics is that they are capable of learning from large datasets and improving predictive accuracy over time, thus fitting well into an area where symptoms can be nuanced and varied [10]. Only with the automation and enhancement of this diagnostic process will deep learning have a chance to provide with an accurate, timely, and personalized diagnosis, with an attendant improvement in outcomes for persons suffering from such mental health conditions. In these regards, we are going to provide some selected deep neural network architectures for the detection of mental disorders like depression and anxiety. We used five different deep-learning models in this paper—NN, CNN, RNN, LSTM, and Tabnet—to detect depression and anxiety from data from the psychological instruments. These are distinct models with entirely different capabilities that enable us to comprehend the many facets of the data and so enhance prediction accuracy and resilience.

Research Gap: To fully understand how well different multiclass strategies work in differentiating between anxiety and depression, more research is needed, even though the majority of research uses machine learning models for binary classification in mental health analysis. Deep learning models can be utilized to accomplish this.

II. RELATED WORK

Depression has been identified from social media, textual data, and interviews using deep learning [11], [12], [13], [14]. Research on using psychological methods to assess sadness and despair employing deep learning has been lacking. Fong et al. developed a complex neural network to forecast anxiety and depression, which highlights the importance of predicting mental health symptoms in global health emergencies in the difficult times like these [15]. In order to predict depression among Indian IT workers, Kaman & Vajpai used both machine learning models and deep learning models in their study [16].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on April 17,2025 at 06:20:42 UTC from IEEE Xplore. Restrictions apply.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 258 (2025) 1348–1364

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Enhancing Mental Health Diagnosis with DA-TabSVM: A Multi-Class Hybrid Approach for Detecting Depression and Anxiety

Priyanka Arora[a], Sonika Dahiya[b]

[a,b]*Software Engineering, Delhi Techonological University, New Delhi 110042, India*

Abstract

Following the COVID-19 pandemic, as individuals adapted to a new normal, the likelihood of despair and anxiety rose. Depression's effects can be lessened by early identification, suitable treatment, and support. The Sustainable Development Goals (SDGs) goals aim to reduce the burden of mental diseases by promoting mental health and well-being through comprehensive health strategies, reducing inequalities, and fostering supportive environments. Lately, automatic social media depression detection has gained attention as a promising field of study. However, we suggest a technique for using a model that learns to answer a mental health questionnaire and utilizing it to produce population-level predictions because there is a dearth of comprehensive annotated data. The purpose of this paper is to efficiently predict mental health conditions like depression and anxiety using data from psychological instruments. In this study, we propose an improved hybrid model employing DT+kNN and RF+NN which were applied on three datasets. Further, we propose a novel prediction model, DA-TabSVM, that can efficiently predict depression and anxiety using multiclass classification. The results show that the DT+kNN hybrid model achieves a maximum accuracy of 95% for predicting depression and 97% for predicting anxiety. Similarly, the RF+NN hybrid model achieves a maximum accuracy of 98% for predicting depression and anxiety. The performance of DA-TabSVM reached up to 97% in terms of accuracy on Dataset 1, 99% on Dataset 2, and 99% on Dataset 3 for depression detection and similarly high when detecting anxiety with a range reaching up to an accuracy of 100% on Dataset 3. Unlike standard strategies that may struggle with intricate feature relationships or be less flexible when dealing with different datasets, our hybrids make use of the complementary skills of many algorithms. Our findings demonstrate the usefulness of hybrid classifiers in mental health evaluation, advancing our knowledge in this crucial field.

© 2025 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering

Keywords: Mental Health; Machine Learning; Deep Learning; Performance Analysis

1877-0509 © 2025 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering
10.1016/j.procs.2025.04.368

From Virus to Vulnerability: A Systematic Literature Review of Psychological Aftershocks Following the COVID-19 Pandemic

Priyanka Arora(✉) and Sonika Dahiya

Software Engineering, Delhi Technological University, New Delhi 110042, India
priyankaarora1019@gmail.com

Abstract. The COVID-19 pandemic has caused an unprecedented global health crisis, both physically and mentally. As the world continues to deal with its consequences, the mental impact of the pandemic is increasingly becoming a cause for concern. This paper presents a systematic literature review of peer-reviewed articles between 2020 and 2025 on the mental health effects of COVID-19, such as depression, anxiety, and sleep problems. Data were collected from prominent academic databases, adhering to PRISMA standards for study selection and synthesis. It discusses the effectiveness and limitations of widely applied mental health screening instruments. This work emphasizes the importance of interdisciplinary strategies that integrate psychological, social, and technological insights. The results indicate an urgent need for a holistic, multi-disciplinary approach to deal with post-COVID mental health issues. There is underutilization of machine learning (ML) and deep learning (DL) methods in explaining and responding to these issues. Integration of ML and DL models for more precise, scalable, and personalized mental health treatments is the need of the hour. This includes the establishment of preventive interventions, enhanced access to mental health care, and the incorporation of real-time monitoring systems. This paper urges a worldwide emphasis on mental health recovery and its crucial function in personal and societal resilience in the post-pandemic era. The crisis has brought to the forefront the necessity to improve mental health systems worldwide. This study aligns with Sustainable Development Goal 3 (Good Health and Well-being), which promotes increased access to mental health care, general well-being, and efficient management of long-term psychological impacts during and after public health emergencies.

Keywords: Post-COVID-19 Mental Health · Long COVID · Depression · Anxiety

1 Introduction

The global outbreak of COVID-19, beyond its disastrous effect on international public health and economies, has unleashed a simultaneous and quieter crisis, a pandemic of mental health disorders [1]. While the initial response has primarily been to the biological

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
A. Swaroop et al. (Eds.): ICDAM 2025, LNNS 1607, pp. 267–284, 2026.
https://doi.org/10.1007/978-3-032-03751-0_22

Copyrighted material

THE TIMES GROUP

IEEE
UP SECTION (INDIA)

BENNETT UNIVERSITY
THE TIMES GROUP

CERTIFICATE OF PRESENTATION

This is presented to

PRIYANKA ARORA, DTU

for **presenting** the paper title

Exploring Deep Learning Architectures for Mental Health Assessment: A Comparative Analysis

in **Global AI Summit 2024, International Conference on Artificial Intelligence and Emerging** organized by **School of Computer Science and Engineering and Technology, Bennett University, India** from **4th-6th Sep, 2024,** Technically Co-Sponsored by IEEE UP Section. We highly appreciate your participation in the conference.

Prof. (Dr.) Abhay Bansal
Dean, School of Computer Science Engineering and Technology
Dean, International Relations & Corporate Outreach
General Chair, Global AI Summit 2024
Bennett University, India

Prof. (Dr.) Ajith Abraham
Vice Chancellor
Co-Patron, Global AI Summit 2024
Bennett University, India

THE TIMES GROUP

IEEE
UP SECTION (INDIA)

BENNETT UNIVERSITY
THE TIMES GROUP

BEST PAPER AWARD

In the session: Artificial Intelligence and Machine Learning

Awarded to

Priyanka Arora, Sonika Dahiya

for the paper titled

Exploring Deep Learning Architectures for Mental Health Assessment: A Comparative Analysis

in **Global AI Summit 2024**, **International Conference on Artificial Intelligence and Emerging Technology** organized by **School of Computer Science Engineering and Technology, Bennett University, India** from **4th – 6th Sep, 2024,** Technically Co-Sponsored by IEEE UP Section. We highly appreciate your participation in the conference.

Prof. (Dr.) Abhay Bansal
Dean, School of Computer Science Engineering and Technology
Dean, International Relations & Corporate Outreach
General Chair, Global AI Summit 2024
Bennett University, India

Prof. (Dr.) Ajith Abraham
Vice Chancellor
Co-Patron, Global AI Summit 2024
Bennett University, India

SCHOOL OF COMPUTER SCIENCE

ICMLDE
INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND DATA ENGINEERING

UPES

CERTIFICATE OF APPRECIATION

This is certified that Dr./ Prof./ Mr./ Mrs./ Miss Priyanka Arora of Delhi Technological University, India has participated/presented a paper entitled Enhancing Mental Health Diagnosis with DA-TabSVM: A Multi-Class Hybrid Approach for Detecting Depression and Anxiety

in the International Conference on Machine Learning and Data Engineering (ICMLDE-2024) held from November 28th - 29th, 2024 at School of Computer Science, UPES, Dehradun, India

Prof. Vijendra Singh
Organising Chair, UPES

Prof. Vijaysekhar Chellaboina
Dean, School of Computer Science, UPES

Dr. Ram K Sharma
Vice Chancellor, UPES

In Association with: University of Dayton, Elsevier, MeitY Government of India, Digital India

ICDAM-2025

LONDON METROPOLITAN UNIVERSITY

WSG

PORTALEGRE POLYTECHNIC UNIVERSITY
POLITÉCNICO DE PORTALEGRE

1816 SGGW

Springer

INTERNATIONAL CONFERENCE ON DATA ANALYTICS & MANAGEMENT (ICDAM - 2025)

Certificate

This is to certify that **Prof. /Dr. / Mr. / Ms. Priyanka Arora** is a presenter/co-author of the paper titled **From Virus to Vulnerability: A Systematic Literature Review of Psychological Aftershocks Following The COVID-19 Pandemic** presented at the **6th International Conference on Data Analytics and Management (ICDAM-2025),** organized jointly by London Metropolitan University, London, United Kingdom in association with the WSG University, Bydgoszcz, Poland; Portalegre Polytechnic University, Portugal & SGGW Management Institute, Poland, Europe during **13th – 15th June 2025.**

Prof. (Dr.) Bal Virdee
London Metropolitan University, UK
Convener

Prof. (Dr.) Abhishek Swaroop
Bhagwan Parshuram Institute of Technology, India
Publication Chair