STRESS AND ANXIETY CLASSIFICATION BASED ON PHYSIOLOGICAL SIGNALS USING MACHINE LEARNING

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

DOCTOR OF PHILOSOPHY

Submitted by

Shikha
(2k19/PHD/CO/502)

Under the Supervision of

Dr. Divyashikha Sethia (Supervisor)
Associate Professor
Department of Software Engineering
and
Prof. S. Indu (Co-Supervisor)
Professor
Department of Electrical and Communication Engineering

Department of Computer Science and Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India.

November, 2025

Acknowledgement

I am sincerely thankful to my supervisor, Dr. Divyashikha Sethia, Department of Software Engineering, for her invaluable guidance, constant encouragement, and insightful feedback, which helped me navigate the challenges encountered throughout the course of my research. I also extend my sincere gratitude to my co-supervisor, Prof. S. Indu, from the Department of Electronics and Communication Engineering, for her expert guidance, constructive inputs, and continuous motivation, which significantly enriched both the technical quality and overall direction of this work.

I extend my heartfelt appreciation to Prof. Prateek Sharma, Hon'ble Vice-Chancellor, Delhi Technological University, for providing a platform to pursue my doctoral research. I also wish to thank Prof. Manoj Kumar, Head, Department of Computer Science and Engineering, for ensuring the necessary facilities were available to carry out this research work. I would like to express my sincere appreciation to the Samsung Research Lab at Delhi Technological University for providing access to laboratory facilities and research devices, which played a crucial role in data collection and experimental validation throughout this research. I am also grateful to Dr. Sonia Baloni Roy, Assistant Professor at IIIT-Delhi and affiliated with the Center for Design and New Media, for her valuable guidance and support during the ethical approval process for data collection.

I am deeply indebted to my husband, Mr. Satnam Singh, whose unwavering support, patience, and belief in my abilities have been a constant source of strength throughout this journey. I am profoundly grateful to my parents, Late Mr. Khemchand and Mrs. Bimla Devi, for their unconditional love, encouragement, and blessings, which have been the foundation of all my achievements. My sincere thanks also go to my sisters, Chetna and Sakshi, and my brother-in-law, Jasbir Singh, for their continuous emotional support and motivation.

I would also like to thank my dearest friend, Dr. Durgesh Nandini, for being a pillar of strength and unwavering encouragement during both the challenges and successes of this journey.

Lastly, I extend my sincere appreciation to my lab members for their cooperation, consistent support, and valuable suggestions, which contributed meaningfully to the completion of this research.

Shikha
(2k19/PHD/CO/502)

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I, **Shikha**, hereby certify that the work which is being presented in the thesis entitled **STRESS AND ANXIETY CLASSIFICATION BASED ON PHYSIOLOGICAL SIGNALS USING MACHINE LEARNING** in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy, submitted in the Department of Computer Science and Engineering, Delhi Technological University, is an authentic record of my own work carried out under the supervision of Dr. Divyashikha Sethia, Associate Professor, Dept. of Software Engineering, DTU and Prof. S. Indu, Professor, Department of Electrical and Communication 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.

Shikha

(2k19/PhD/CO/502)

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(s) **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 **Shikha** (2k19/Phd/CO/502) has carried out her research work presented in this thesis entitled **STRESS AND ANXIETY CLASSIFICATION BASED ON PHYSIOLOGICAL SIGNALS USING MACHINE LEARNING** for the award of Doctor of Philosophy from Department of Computer Science and 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.

Supervisor
Dr. Divyashikha Sethia
Associate Professor
Department of
Software Engineering
Delhi Technological University

Co-Supervisor
Prof. S. Indu
Professor
Department of Electronics and
Communication Engineering
Delhi Technological University

Date:

Indian Examiner

ABSTRACT

Stress and anxiety significantly affect cognitive, emotional, and physiological functions. Physiological signals such as Electroencephalogram (EEG), Electrocardiogram (ECG), Electrodermal Activity (EDA), Blood Volume Pulse (BVP), and Respiration (RESP) provide objective, real-time insights compared to subjective assessments. Machine learning-based systems have recently gained attention for classifying stress and anxiety from high-dimensional physiological data, enabling real-time interventions through wearable devices. However, challenges such as redundant features, class imbalance, and high computational complexity continue to limit their scalability and practical use. To address these challenges, this thesis focuses on the following key research gaps: (1) lack of optimized channel selection in EEG-based stress classification frameworks, (2) absence of systematically collected multimodal datasets reflecting progressive stress in academic environments, (3) need for scalable and interpretable feature selection algorithms, and (4) limited exploration of reinforcement learning strategies for stress detection using physiological signals.

Firstly, this thesis presents a comprehensive literature review on automated stress and anxiety classification using physiological signals. It covers stress and anxiety theory, physiological signals, and their relationship with mental states, preprocessing methods, domain-specific feature extraction, feature selection techniques, and machine learning models.

Secondly, the thesis proposes an ensemble-based EEG stress classification framework for EEG signal-based wearable applications. It introduces the **KRAFS-ANet** framework, which stacks Bagging **K**-Nearest Neighbor, Bagging **S**upport Vector Machine, and Bagging **R**andom Forest classifiers, with an **A**rtificial Neural **Net**work (ANN) as the meta-classifier. The approach incorporates optimized channel selection

and ensemble stacking to improve accuracy while reducing computational time. The thesis validates the framework on three benchmark datasets: MAT, SAM40, and DASPS. It achieves accuracies of 98.63%, 97.25%, and 94.92%, respectively, along with consistently high F1-scores. However, this work does not address multimodal signal fusion, class imbalance, or feature interdependencies.

Thirdly, this thesis introduces the **A**cademic **S**tress **D**ataset (**ASD**), where physiological signals such as Interbeat Interval (IBI), BVP, and EDA are recorded during the Montreal Imaging Stress Task (MIST) to induce progressive mental arithmetic stress in engineering students. This work applies a hybrid feature selection to select the most informative features and further utilizes Bayesian optimization for hyperparameter tuning. The Gradient Boosting model achieves accuracy of 98.28% and 97.02% for 2-level and 3-level classification, respectively. Using only EDA and HRV features provides comparable accuracy, and SHAP-based Explainable AI (XAI) analysis further confirms them as the most informative features.

Finally, this thesis proposes and compares two efficient and effective feature selection algorithms. The first, **CorLMI-FSA**, combines **Cor**relation, **L**ogistic Regression (LR), and **M**utual **I**nformation (MI) to reduce redundancy in stress classification using EDA and HRV features from the self collected ASD dataset. It achieves highest accuracy of 96.82% for binary and 95.84% for three-level classification. The second, **ST-CIRL** (**S**MOTE**T**omek-**C**orrelated **I**nteractive **R**einforcement **L**earning), addresses class imbalance and optimizes feature selection through interactive reinforcement learning. ST-CIRL utilizes ECG, EDA, and RESP features from the Spider-phobic anxiety dataset, applies Optuna optimization, and achieves the highest accuracy of 95.35% and an F1-score of 95.49% using LightGBM, outperforming existing methods. Cross-dataset evaluation shows that CorLMI performs better in binary classification with lower runtime, while ST-CIRL achieves higher accuracy in multi-class classification.

This thesis advances intelligent, wearable stress monitoring systems and lays the foundation for real-time health assessment, stress-aware learning environments, and Human-Computer Interaction applications. It outlines future directions to enhance stress and anxiety detection by improving model generalizability, enabling real-time implementation, and integrating personalized interventions.

List of Publications

The research presented in this thesis is primarily based on the following peer-reviewed articles:

Papers Accepted/Published in International Journals

Journal 1: S. Shikha, D. Sethia, and S. Indu. Optimization of Wearable Biosensor Data for Stress Classification Using Machine Learning and Explainable AI. IEEE Access, 12:169310–169327, September 2024. **(SCIE, Impact factor: 3.4, Publisher: IEEE)**. Doi: `https://doi.org/10.1109/ACCESS.2024.3463742.` **(Published)**.

Journal 2: S. Shikha, D. Sethia, and S. Indu. KRAFS-ANet: A Novel Framework for EEG-Based Stress Classification Using Channel Selection and Optimized Ensemble Stacking. International Journal of Machine Learning and Cybernetics, 1–26, November 2024. **(SCIE, Impact factor: 3.1, Publisher: Springer)**. Doi: `https://doi.org/10.1007/s13042-024-02455-2.` **(Published)**.

Journal 3: Shikha, Divyashikha Sethia, and S. Indu. ST-CIRL: A Reinforcement Learning-Based Feature Selection Approach for Enhanced Anxiety Classification. Physiological Measurement, 13(2):025009, February 2025. **SCIE, Impact factor: 2.3, Publisher: IOP Science**. Doi: `https://doi.org/10.1088/1361-6579/adb006.` **(Published)**.

Journal 4: S. Shikha, D. Sethia, and S. Indu. A Systematic Review on Physiology-Based Anxiety Detection Using Machine Learning. Biomedical Physics & Engineering Express, May 2025. **ESCI, Impact factor: 1.3, Publisher: IOP Science**. Doi: https://doi.org/10.1088/2057-1976/add5fc. **(Published)**.

Papers Accepted/Published in International Conferences

Conference 1: Shikha, Divyashikha Sethia, and S. Indu. "CorLMI-FSA: An Efficient Feature Selection Approach for Stress Classification Using Physiological Signals." *In Proceedings of the 5th IEEE International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies (ICAECT), 2025.* Doi: https://doi.org/10.1109/ICAECT63952.2025.10958862. **(Published)**.

Conference 2: Shikha, Divyashikha Sethia, and S. Indu. "A Review on Feature Selection Techniques for Anxiety Classification from Physiological Signals." *In Proceedings of the IEEE International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT), 2025.* Doi: https://doi.org/10.1109/CE2CT64011.2025.10939412. **(Published)**.

Conference 3: Shikha, Divyashikha Sethia, and S. Indu. "Investigating the Effect of Sliding Window Length for EEG-based Anxiety Classification." *In Proceedings of the IEEE Asia Pacific Conference on Innovation in Technology (APCIT), 2024.* Doi: https://doi.org/10.1109/APCIT62007.2024.10673545. **(Published)**.

Conference 4: Shikha et al. "The Effectiveness of Advance Deep Learning Architectures for Classification of Stress using Raw EEG Data." *In Proceedings of the IEEE Asia Pacific Conference on Innovation in Technology (APCIT), 2024.* Doi: https://doi.org/10.1109/APCIT62007.2024.10673521. **(Published)**.

Conference 5: Shikha et al. "Body Sensor-Based Multimodal Nurse Stress Detection Using Machine Learning." *In Proceedings of the 16th IEEE International Conference on COMmunication Systems & NETworkS (COMSNETS), 2024.* Doi: https://doi.org/10.1109/COMSNETS59351.2024.10427006. **(Published)**.

Conference 6: Shikha Divyashikha Sethia, and S. Indu. "Ensemble Classifier for EEG-Based Stress Classification: An Empirical Study on Stacking Classifiers." *In Proceedings of the Springer International Conference on Computational Electronics for Wireless Communications (ICCWC), 2023.* Doi: https://doi.org/10.1007/978-981-97-1943-3_34. **(Published)**.

Conference 7: Shikha, Divyashikha Sethia, and S. Indu "Efficient Stress Detection System using Feature Selection and Ensemble Learning Method" *In Proceedings of the IEEE International Conference on Communication, Security and Artificial Intelligence. (ICCSAI), 2022.* **(Presented)**.

Conference 8: Shikha et al. "HRV and GSR as Viable Physiological Markers for Mental Health Recognition" *In Proceedings of the 14th IEEE International Conference on COMmunication Systems & NETworkS (COMSNETS), 2022.* Doi: 10.1109/COMSNETS53615.2022.9668439. **(Published)**.

Conference 9: Shikha, et al. "Stacked Sparse Autoencoder and Machine Learning Based Anxiety Classification Using EEG Signals" *In Proceedings of the 1st ACM International Conference on AI-ML Systems, 2021.* Doi: https://doi.org/10.1145/3486001.3486227. **(Published)**.

Contents

List of Tables

List of Figures

List of Abbreviations

ACC	Accelerometer
ADB	Adaptive Boosting
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ASD	Academic Stress Dataset
ASR	Artifact Subspace Reconstruction
BPF	Butterworth Bandpass Filters
BT	Bagged Trees
BVP	Blood Volume Pulse
CNN	Convolutional Neural Networks
CNS	Central Nervous System
CVNNI	Coefficient of Variation of NN Intervals
CVSD	Coefficient of Variation of Successive Differences
CorLMI	Correlation-Logistic Mutual Information
DL	Deep Learning
DNN	Deep Neural Networks
DT	Decision Tree
DWT	Discrete Wavelet Transform
EBM	Explainable Boosting Machine
ECG	Electrocardiogram
EDA	Electrodermal Activity
EMD	Empirical Mode Decomposition
FD	Frequency Domain
FFT	Fast Fourier Transform

FSA	Feature selection algorithms
GA	Genetic Algorithm
GBM	Gradient Boosting Machine
GRU	Gated Recurrent Unit
GSR	Galvanic Skin Response
HAM-A	Hamilton Anxiety Rating Scale
HCI	Human-Computer Interaction
HF	High Frequency
HFnu	High-Frequency Power in Normalized Units
HPF	High-Pass Filters
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Inter-Beat Interval
ICA	Independent Component Analysis
IMFs	Intrinsic Mode Functions
IRL	Interactive Reinforcement Learning
KNN	K-Nearest Neighbour
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LF	Low Frequency
LFnu	Low-Frequency Power in Normalized Units
LPF	Low-Pass Filters
LR	Logistic Regression
LSTM	Long Short-Term Memory networks
MAF	Moving Average Filters
MDP	Markov Decision Process
MI	Mutual Information
MIST	Montreal Imaging Stress Task
ML	Machine Learning
MLP	Multilayer Perceptron
NB	Naive Bayes

NCA	Neighborhood Component Analysis
NMI	Normalized Mutual Information
PBP	Probabilistic Binary Pattern
PCC	Pearson Correlation Coefficient
PI	Permutation Importance
PNS	Parasympathetic Nervous System
PPG	Photoplethysmography
PSD	Power Spectral Density
PSO	Particle Swarm Optimization
RESP	Respiration
RF	Random Forests
RFE	Recursive Feature Elimination
RL	Reinforcement Learning
RMSSD	Root Mean Square of Successive Differences
RNN	Recurrent Neural Network
RP	Relative Power
RR	Respiration Rate
RRV	Respiratory Rate Variability
SBS	Sequential Backward Selection
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SCWT	Stroop Color Word Test
SFS	Sequential Forward Selection
SHAP	Shapley Additive Explanations
SMOTE	Synthetic Minority Over-Sampling Technique
SNS	Sympathetic Nervous System
SSAE	Stacked Sparse Autoencoder
ST	Skin Temperature
ST-CIRL	SMOTETomek-Correlated Interactive Reinforcement Learning
SVM	Support Vector Machines
SampEn	Sample Entropy

TD	Time Domain
TINN	Triangular Interpolation of NN Interval Histogram
TSST	Trier Social Stress Test
VHF	Very High-Frequency
WHO	World Health Organization
WOA	Whale Optimization Algorithm
XAI	Explainable AI
XGB	Extreme Gradient Boosting
mRMR	Maximum Relevance Minimum Redundancy

Chapter 1

Introduction

In today's fast-paced and demanding world, stress has become an unavoidable part of life [217]. The pressures of balancing personal and professional responsibilities, meeting societal expectations, and managing uncertainty contribute to significant emotional and physical strain. While moderate stress can improve focus and motivation, prolonged or unmanaged stress has serious consequences for both mental and physical health, including cardiovascular diseases, metabolic disorders, weakened immune function, and anxiety disorders [217] [250]. Despite these risks, stress is often overlooked or dismissed, leading individuals to ignore early symptoms until the condition becomes severe. Anxiety, which is closely linked to stress, manifests as persistent worry, apprehension, and heightened emotional responses [31]. Unlike stress, which typically arises from external pressures, anxiety is often internally driven and can occur without an immediate trigger. Anxiety disorders, such as generalized anxiety disorder and social anxiety disorder, further exacerbate these issues by intensifying distress and impairing an individual's ability to function in daily life [226]. Both stress and anxiety activate the autonomic nervous system, leading to physiological changes that, when prolonged, contribute to serious health issues [219].

Over the years, researchers have explored various means to monitor and assess stress and anxiety through different approaches effectively:

1. **Self-Reported Psychological Measures:** Standardized questionnaires such as the Perceived Stress Scale (PSS) [125] and the State-Trait Anxiety Inventory (STAI) [236] are widely used due to their ease of administration and cost-

effectiveness. However, these assessments are less reliable because they rely on memory, personal perception, and social influences, which can make continuous stress monitoring ineffective [13].

2. **Behavioural Indicators:** Stress and anxiety can also be inferred from facial expressions [54], speech patterns [32], posture, and movement dynamics [179]. While these indicators are non-invasive and easily captured using cameras and microphones, they remain susceptible to manipulation. Individuals may consciously alter their expressions or tone of voice in social situations, reducing the classification reliability of stress and anxiety. Furthermore, environmental factors, such as lighting conditions and background noise, further affect the accuracy of physical assessments.

3. **Physiological Indicators:** A more objective and reliable approach involves analyzing physiological responses that reflect autonomic nervous system activity. These measures include Electroencephalogram (EEG) [27], Electrocardiogram (ECG) [171], Heart Rate Variability (HRV), Electrodermal Activity (EDA) [149], Heart Rate (HR) [103], Blood Volume Pulse (BVP) [69], and Respiration (RESP) [18]. Unlike self-reports and behavioral indicators, physiological indicators provide continuous, real-time monitoring, ensuring higher accuracy and reliability. One key advantage of physiological signals is that they cannot be consciously controlled or easily masked, making them a more objective measure of stress and anxiety [202]. While individuals may suppress facial expressions or adjust speech patterns, autonomic responses, such as Heart Rate Variability (HRV) and EDA, remain involuntary, reflecting their true emotional state. It makes physiological monitoring valuable for real-time stress detection and continuous assessment in diverse environments.

   Moreover, integrating multiple physiological signals through a multimodal approach further enhances stress classification by compensating for noise, artifacts, and individual variability [62]. However, despite its advantages, physiological monitoring relies on specialized equipment and complex signal processing, which pose challenges for large-scale implementation.

This thesis focuses on leveraging physiological indicators for stress and anxiety classification. It examines the effectiveness of single and multimodal physiological signals in capturing autonomic nervous system responses and improving stress classification accuracy.

1.1 Motivation and Background

1.1.1 Why is Stress and Anxiety Classification Important?

Stress and anxiety-related disorders have become a major public health concern, affecting millions of individuals worldwide. According to the World Health Organization (WHO), over 301 million individuals, including 58 million children and adolescents, suffered from anxiety disorders in 2019 [181]. The COVID-19 pandemic further exacerbated these conditions, increasing global anxiety and depression rates by 25% [264].

Beyond health implications, stress and anxiety impact multiple sectors, including education and the workplace [161]. In an academic environment, high stress levels hinder learning outcomes, cognitive function, and long-term resilience [2]. In professional environments, excessive workplace stress reduces productivity, job satisfaction, and decision-making abilities [136]. Given the widespread effects of these conditions, there is a critical need for effective and automated detection methods using physiological signals to detect stress and anxiety early, enabling timely intervention and prevention.

1.1.2 Why Are Physiological Signals a Reliable Alternative for Stress and Anxiety Classification?

Traditional assessment methods, such as self-reports and behavioral observations, have inherent limitations. They are often subjective, prone to bias, and incapable of providing real-time or continuous monitoring. These constraints reduce their effectiveness in accurately detecting stress and anxiety levels.

In contrast, physiological signals offer objective, continuous, and real-time insights

into an individual's emotional state. Signals such as HRV, EDA, and RESP reflect the activity of the autonomic nervous system, which governs involuntary physiological responses to stress. Unlike behavioral indicators, individuals cannot easily control or manipulate these responses, making them less susceptible to social desirability bias and self-report inaccuracies [81]. Due to their involuntary and consistent nature, physiological signals provide a more reliable and unbiased basis for stress and anxiety classification, supporting the development of automated and scalable mental health monitoring systems.

1.1.3 Background for Automated Stress and Anxiety recognition using physiological signals?

Automated stress and anxiety recognition systems using physiological signals aim to assess an individual's emotional state in real-time without relying on subjective self-reports. These systems analyze physiological signals that reflect underlying autonomic and neural responses to stressors [85]. Given these signals' complex and high-dimensional nature, machine learning algorithms play a crucial role in enabling automation. They facilitate extracting discriminative features, identifying patterns, and classifying stress and anxiety levels [86, 81]. By training on large physiological datasets, these models can capture subtle signal variations that may not be evident through traditional analysis methods.

An automated recognition system typically follows a structured pipeline involving signal acquisition, preprocessing, feature engineering, and classification, as shown in Figure 1.1. Wearable devices initially record raw physiological data and then clean the signals to remove artifacts and noise. Feature extraction methods convert the cleaned signals into informative representations, while feature selection techniques such as correlation and mutual information reduce redundancy and retain the most relevant inputs. Furthermore, combining multiple physiological signals in a multimodal framework enhances the reliability and accuracy of classification. This integration captures complementary stress-related patterns across different modalities, resulting in a more robust and comprehensive evaluation of the individual's stress or anxiety state [279].

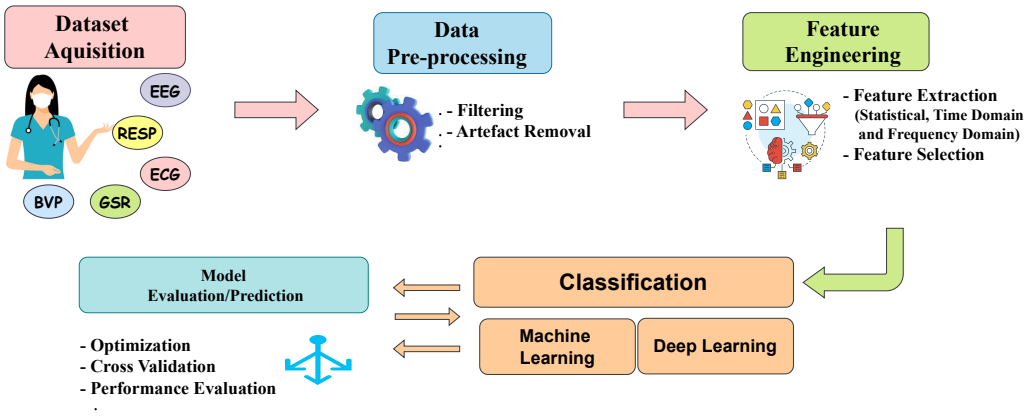


Figure 1.1: Stress and anxiety classification methodology.

Various machine learning and deep learning models, including Support Vector Machines (SVM), Random Forests (RF), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks, can be employed. Further, these models can be optimized using cross-validation and evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score [8]. These trained models enable real-time automated stress and anxiety monitoring in academic, clinical, and workplace environments.

1.1.4 What Are the Challenges in Using Physiological Signals for Stress and Anxiety Classification?

Despite numerous advantages, physiological signals pose several challenges in stress and anxiety classification.

1. **Signal Noise and Artifacts:** Movement, environmental interference, and sensor limitations introduce noise and artifacts, distorting data accuracy. While offering non-intrusive monitoring, wearable devices require advanced preprocessing techniques to mitigate these issues [200].

2. **High-Dimensionality of Physiological Signals:** The multi-domain nature of physiological signals—including time, frequency, and time-frequency domains—leads to a high-dimensional feature space [200]. This increased dimensionality introduces redundancy, reduces model interpretability, and

increases computational complexity, necessitating effective feature selection techniques.

3. **Limited Availability of Large, Well-Labeled Datasets:** The scarcity of publicly available, well-annotated physiological datasets restricts the successful application of advanced deep learning models [4]. This limitation affects model generalization and performance, making it difficult to leverage deep architectures effectively.

4. **Real-Time Processing and Computational Efficiency:** Ensuring real-time stress classification while optimizing computational efficiency for wearable and mobile applications remains a significant challenge [202].

Despite these challenges, advancements in sensor technology, machine learning models, and signal processing techniques continue to enhance the feasibility and accuracy of physiological stress classification. These developments have enabled a wide range of real-world applications, including mental health monitoring, workplace stress management, stress-aware learning environments, and human-computer interaction systems. Such applications highlight the practical significance of developing accurate, efficient, and real-time classification models. Chapter 7 provides a detailed exploration of these application areas.

1.2 Research Gaps

Several challenges remain in the automated classification of stress and anxiety using physiological signals. The key research gaps are as follows:

1. **Research Gap 1: Trade-off between accuracy and efficiency**.

   Developing accurate yet efficient classification models for wearable devices is challenging due to the high computational demands of processing physiological signals [148]. Complex models improve accuracy but are unsuitable for real-time deployment due to resource constraints, while lightweight models compromise performance [268] [148]. Existing approaches often focus on accuracy without optimizing efficiency, limiting practical applications [26] [271]. Therefore, there

is a need for computationally efficient models that maintain high classification performance while being feasible for real-time wearable implementation.

2. **Research Gap 2: Lack of feature selection techniques**

   Existing studies have largely overlooked the role of feature selection in stress and anxiety classification. Most approaches rely on high-dimensional feature sets, increasing model complexity and reducing efficiency [140] [175]. Developing advanced feature selection methods can enhance classification accuracy while minimizing computational overhead [188].

3. **Research Gap 3: Identifying relevant physiological signals and their most contributive features**

   Determining the most relevant physiological signals, such as EEG, ECG, and EDA, for stress classification remains a challenge [209][51]. Additionally, identifying the most contributive features within each signal domain is crucial for improving stress and anxiety classification accuracy and reducing model complexity. Explainable AI (XAI) techniques can be leveraged to interpret model decisions and highlight the most impactful features, which enables a more informed feature selection process.

4. **Research Gap 4: Limited application of ensemble stacking models for stress and anxiety classification**

   While traditional machine learning models have been used for stress and anxiety classification, the exploration of ensemble stacking techniques remains limited [76]. Stacking can enhance classification performance by leveraging the strengths of multiple models, yet its potential in stress and anxiety classification is underexplored [100]. Addressing this gap can improve model robustness and accuracy.

Table 1.1: Summary of open stress and anxiety datasets

Dataset	Modality	Domain	Stress Inducer
MAT [280]	Single	Stress	Mental arithmetic task
SAM40 [84]	Single	Stress	Stroop color-word test, arithmetic problem solving, mirror image identification, relaxation
CLAS [150]	Multi	Stress	Mathematical problems, Stroop test, logical problems, video and image viewing
SRAD [99]	Multi	Stress	Driving under different situations
DASPS [30]	Single	Anxiety	Face-to-face psychological stimuli
Spider-Phobic [109]	Multi	Anxiety	Spider phobic videos

5. **Research Gap 5: Lack of open datasets in an academic environment**

   Existing stress datasets, such as CLAS [150], SAM40 [84], and MAT [280], primarily focus on binary classification tasks, such as stressed vs. non-stressed, limiting the ability to study multi-level stress progression. Additionally, these datasets do not capture progressive stress or mental workload increases within real-world academic environments, particularly among engineering students. Furthermore, tasks like the Montreal Imaging Stress Task (MIST) [58] are designed to induce progressively increasing stress; however, no publicly available physiological dataset currently integrates MIST with real-world academic settings for engineering students. While EEG signals provide detailed insights into cognitive and emotional stress, their intrusive and cumbersome setup makes them impractical for continuous or daily use among students. Hence, to address these limitations, there is a growing need to focus on physiological signals collected through wearable devices, such as EDA, BVP, and Interbeat Interval (IBI), which offer non-invasive, real-time stress monitoring suitable for academic environments. Table 1.1 summarises existing datasets briefly, and section 3.2 provides detailed information.

6. **Research Gap 6: Limited exploration of deep learning algorithms for stress and anxiety classification**

   While traditional stress and anxiety classification uses machine learning

algorithms, the application of deep learning remains underexplored [48] [232]. Deep learning can automatically extract complex patterns from physiological signals, potentially improving classification accuracy. Further research is needed to harness its full potential in this domain.

7. **Research Gap 7: Lack of automated EEG channel selection**

   Most existing studies rely on manual channel selection based on predefined brain regions, which can be subjective and inefficient. There is a need for automated, data-driven channel selection to reduce dimensionality and improve classification performance [225][76][15].

1.3 Problem Statement

Effective stress and anxiety classification using physiological signals faces several challenges, including EEG channel selection, feature dimensionality, and dataset availability. This thesis aims to overcome four key challenges: 1) Existing studies often rely on manual EEG channel selection, which introduces subjectivity and limits reproducibility, particularly in real-time wearable applications. 2) The multi-domain nature of physiological signals results in high-dimensional feature spaces, increasing redundancy and reducing model interpretability. Developing efficient feature selection techniques is crucial to enhancing classification performance while maintaining computational efficiency. 3) Many classification models prioritize accuracy but neglect efficiency, making them impractical for wearable devices. Achieving a balance between performance and efficiency remains a key challenge. 4) There is a lack of publicly available datasets for classifying stress and anxiety in an academic environment.

1.4 Research Objectives and Contribution of the Thesis

The primary objective of this thesis is to design and develop robust and efficient machine learning-based systems for the classification of stress and anxiety using physiological signals.

1. **Objective 1: Literature Review -** To perform a systematic literature review

on anxiety detection based on Physiological signals using machine learning techniques.

Contribution: The thesis conducts a comprehensive review of physiological signal-based anxiety classification using machine learning. The review explores the multi-domain nature of physiological signals and the challenges posed by high-dimensional feature sets. It examines various machine and deep learning models and identifies key research gaps for classifying stress and anxiety. (Completed by **Conference 2** and **Journal 4**)

2. **Objective 2: Methodology -** To develop a better scientific methodology to reduce the complexity of the stress and anxiety detection model and perform a comparative analysis of the proposed approach with existing stress predictive models.

   Contribution: The thesis proposes a novel Ensemble stacking **KRAFS-ANet** framework, which integrates *bagging K-Nearest Neighbour (KNN), bagging Random Forest (RF), and bagging Support Vector Machine (SVM) with Artificial Neural Network (ANN) meta-classifiers*. This framework is specifically designed to enhance the performance of stress and anxiety classification using EEG signals from the MAT [280], SAM40 [84], and DASPS [30] datasets. It also performs channel selection to identify the most informative EEG channels. Furthermore, the thesis compares the proposed KRAFS-ANet framework with existing studies for channel selection and stacking approaches to demonstrate its effectiveness and advancements [223] [148] [76] [15]. (Completed by **Journal 2**)

3. **Objective 3: Efficient Algorithm -** To propose an efficient stress and anxiety detection algorithm and perform a comparative analysis of the proposed approach with existing stress predictive models.

   Contribution: The thesis proposes two efficient feature selection algorithms to enhance physiological signal-based classification. 1) The *Correlation-Logistic Mutual Information Feature Selection Algorithm* **(CorLMI-FSA)** effectively reduces redundancy and extracts the most relevant features, demonstrating its utility for stress classification. The performance is evaluated on self-

collected *Academic Stress Dataset* (**ASD**) using EDA and HRV, which are identified as the most reliable signals for stress classification in an academic environment. Further, this work compares the proposed work with existing state-of-the-art work [51][45][42]. 2) The *Correlated Interactive Reinforcement Learning Feature Selection Algorithm* **(CIRL-FSA)** is proposed to model feature selection as a dynamic decision-making process using interactive reinforcement learning. This work extends CIRL-FSA to the *SMOTETomek-Correlated Interactive Reinforcement Learning* **(STCIRL-FSA)** to address class imbalance in the Spiderphobic dataset [109]. This extension integrates the SMOTETomek resampling technique with interactive reinforcement learning-based feature selection. STCIRL-FSA employs EDA, HRV, and RESP signals to capture diverse autonomic responses and enhance classification reliability. It achieves the highest classification performance with state-of-the-art accuracy and efficiency. (Completed by **Conference 1 and Journal 3**)

4. **Objective 4: Data Collection -** To propose a new study to collect data and automated stress detection in an academic environment using wearable biosignals.

   Contribution: The thesis proposes an *Academic Stress Dataset* (**ASD**), which collects physiological signals such as EDA, IBI (Interbeat Interval), and BVP from wearable devices across different stress levels and meditation phases. A key motivation for collecting this data is the lack of datasets that specifically explore the combination of IBI-derived HRV, BVP, and EDA signals during (**MIST** [58]) within an academic setting for engineering students. This work identifies the most contributing features and illustrates the impact of meditation on stress reduction through visualizations of physiological data collected from wearable devices among college students. (completed by **Journal 1**)

1.5 Thesis Overview

The following is the structure for the rest of the thesis:

- **Chapter 2 - Technical Background for Physiological Signals-based Stress and**

Anxiety Detection: presents the background information for Stress and Anxiety, physiological signals, and their correlations. Further, it describes various data preprocessing techniques and feature selection algorithms for automated stress and anxiety classification.

- **Chapter 3 - Literature Review:** presents the literature survey of the existing state of automated stress and anxiety recognition using physiological signals. It discusses the openly available datasets and their limitations. Further, it describes preprocessing, feature extraction, and selection-related studies. It presents the contribution of different ML and DL models and discusses the challenges associated with ML and DL stress and anxiety recognition techniques. This chapter also outlines the different evaluation measures required to assess this thesis's proposed framework and algorithms.

- **Chapter 4 - KRAFS-ANet: A Scientific Methodology for EEG-Based Stress Classification:** proposes an EEG-based novel **KRAFS-ANet** framework for better scientific methodology using channel selection and optimized ensemble stacking. This chapter proposes an efficient methodology for stress and anxiety classification using EEG-based MAT [280], SAM40 [84], and DASPS [30] datasets. It strategically employs channel selection to identify the most informative channels and utilizes an ensemble stacking technique, integrating bagging classifiers with an Artificial Neural Network (ANN) meta-classifier. Additionally, it compares previous state-of-the-art work and validates the performance of the proposed framework on multiple datasets [223] [148] [76] [15].

- **Chapter 5 - Academic Stress Dataset Collection and Wearable-Based Stress Classification Optimization:** proposes a *Academic Stress Dataset* (**ASD**) collected using the Empatica E4 wearable device [137], [218] to gather physiological signals from college students during the Montreal Image Stress Task (MIST) [58]. It also investigates the effectiveness of meditation audio in reducing stress levels after academic exposure. Furthermore, the XAI [104] analysis identifies the most significant physiological features contributing to stress classification.

- **Chapter 6 - Efficient Feature Selection Algorithm for Stress and Anxiety Classification:** addresses the challenges of redundant features, class imbalance, and real-world applicability in physiological stress and anxiety classification. First, it proposes the Correlation-Logistic Mutual Information Feature Selection Algorithm **(CorLMI-FSA)**, which dynamically adjusts redundancy weights through a logistic function to capture nonlinear feature dependencies. CorLMI-FSA is validated on the self-collected *Academic Stress Dataset (ASD)* using EDA and HRV features and compared against state-of-the-art methods. The chapter further introduces the Correlated Interactive Reinforcement Learning Feature Selection Algorithm **(CIRL-FSA)**, which formulates feature selection as a dynamic decision-making problem using reinforcement learning and meta-descriptive statistics. To address class imbalance in the Spiderphobic dataset [109], CIRL-FSA is extended to **STCIRL-FSA** by integrating the SMOTE-Tomek resampling technique. STCIRL-FSA enhances anxiety classification performance by selecting compact and informative feature subsets, with machine learning models optimized using Optuna. Finally, the chapter presents a cross-dataset evaluation to assess the generalizability of both proposed algorithms.

- **Chapter 7 - Conclusion, Future Work & Social Impact:** provides Conclusion, Future Work & Social Impact. This chapter summarizes the conclusions from this research work and highlights the potential future work and the social applications in this area.

Chapter 2

Technical Background for Physiological Signals-based Stress and Anxiety Detection

Stress and anxiety elicit a complex set of physiological responses primarily governed by the Autonomic Nervous System (ANS) and the Central Nervous System (CNS) [123]. These responses result in observable changes in cardiovascular activity, respiratory patterns, and electrodermal activity. The ANS helps the body stay balanced by adjusting these responses based on emotional and environmental changes. It comprises two main branches: the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS). The SNS activates the body's fight-or-flight response, increasing heart rate, respiration, and sweat gland activity to prepare for immediate action. Conversely, the PNS supports the rest-and-digest state by slowing the heart rate, reducing respiration, and promoting relaxation to restore balance. Chronic stress or prolonged anxiety can disrupt this regulatory balance, leading to autonomic dysregulation and increasing the risk of long-term health complications such as hypertension, cardiovascular disease, and anxiety disorders [167].

To objectively assess these states, researchers rely on physiological signals such as Electrodermal Activity (EDA), Heart Rate Variability (HRV), Respiration (RESP), Electrocardiogram (ECG), Blood Volume Pressure (BVP), and Electroencephalogram (EEG) [85]. These signals provide real-time insights into autonomic and neural activity,

Table 2.1: Experimental stress-inducing tasks.

Test	Study	Description
Trier Social Stress Test (TSST) [129]	[79][23][88] [176][175] [248][24]	It is also known as the public speaking task. Participants deliver a speech on a specified topic within a limited time. Afterward, they perform verbal calculations. Both tasks are completed in front of an audience to induce stress.
Stroop Color Word Test (SCWT) [213]	[29][188]	Participants view color names printed in incongruent font colors and must name the font color rather than reading the word. This task measures cognitive interference and stress response.
Montreal Imaging Stress Task (MIST) [58]	[276][265] [222][158] [265][222]	The MIST consists of three stages: rest, control, and experiment. In the rest stage, participants view a static screen. During the control stage, they solve simple arithmetic problems. The experiment stage includes time-constrained complex math problems to induce stress.

making them valuable for stress and anxiety classification. Machine Learning (ML) models can detect stress-induced physiological patterns by analyzing signal variations and support real-time mental health monitoring [81]. A critical component of this research involves the use of standardized assessment tasks that induce measurable physiological responses. These controlled experimental methods are benchmarks for evaluating stress and anxiety classification models, ensuring their reliability and validity.

This chapter presents the technical foundation for physiological signal-based stress and anxiety classification. It introduces standardized stress and anxiety inducing tasks and assessment tests used in an experimental environment. It then provides an overview of relevant physiological responses and commonly used signals. Finally, the chapter outlines essential preprocessing techniques, feature extraction methods, and feature selection algorithms that enhance signal quality and improve classification accuracy.

2.1 Stress and Anxiety Inducing Tasks

Inducing stress and anxiety in controlled settings helps researchers systematically evaluate their effects on mental and physiological health. This process enables accurate measurement of physiological signals and behavioral responses necessary for effective

Table 2.2: Experimental anxiety-inducing tasks.

Task	Study	Description
Spiderphobic Video Task	[109][80]	Participants diagnosed with spider phobia are shown video clips of spiders. The task is designed to trigger fear and anxiety responses specific to phobic stimuli.
Public Speaking	[207][218] [218][9]	Participants prepare and deliver a speech on a given topic under time constraints and in front of an audience or evaluators. This task reliably induces social anxiety and performance-related stress.
Face-to-Face Interaction Task	[31][165]	Recite a situation to participants and then instruct them to recall it.
Cycling Task	[277]	The participant rides a stationary bike in two stages. During the first stage, the participant rides at a comfortable speed. In the second stage, the participant imagines competing with someone riding at a speed above 80 km/h.
Arithmetic Task	[130][197]	Participants solve arithmetic problems, such as subtraction, multiplication, and addition, within a time limit. This task induces cognitive load and performance-related anxiety.

interventions.

1. **Stress Inducing Tasks:** Stress assessment tasks are widely used in research and clinical settings to evaluate stress levels and understand how stressors affect physiological and psychological processes. Table 2.1 shows standard experimental tasks that induce stress under controlled conditions by exposing participants to cognitive, emotional, or social challenges. Stress-inducing tasks, such as the MIST [58] and the TSST [129], are designed to provoke physiological stress responses progressively or acutely.

2. **Anxiety Inducing Tasks:** Anxiety-related tasks vary widely, as no universally standardized experimental task exists for anxiety evaluation. Studies incorporate different approaches, including cognitive interference tasks [130][197], social pressure tasks [218][9], and emotion-inducing stimuli [31][109], to elicit anxiety-related responses.

Table 2.3: Stress and anxiety assessment tests [198].

Questionnaire	Domain	Description
Perceived Stress Scale (PSS) [125]	Stress	A self-reported questionnaire that evaluates how individuals perceive stress in their daily lives [125].
State-Trait Anxiety Inventory (STAI) [236]	Anxiety	A self-report questionnaire that measures both state (temporary) and trait (long-term) anxiety to assess an individual's anxiety levels [236].
Hamilton Anxiety Rating Scale (HAM-A) [146]	Anxiety	A clinician-administered scale that evaluates the severity of anxiety symptoms based on psychological and physical indicators [146].

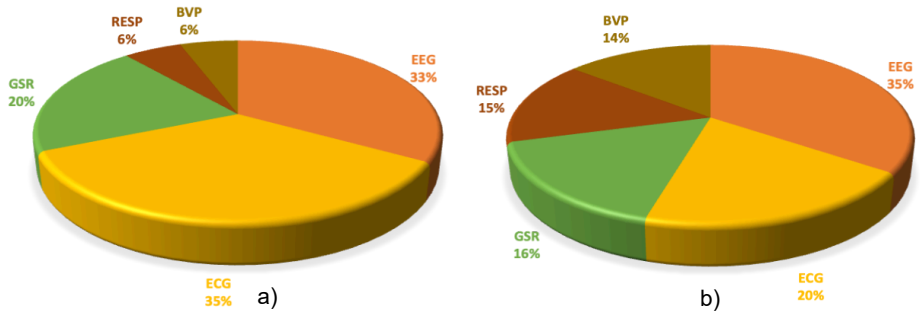


Figure 2.1: Distribution of physiological signals used in a) stress and b) anxiety classification studies.

2.2 Stress and Anxiety Assessment Tests

Self-reported assessments rely on subjective evaluations of perceived stress or anxiety levels, often through standardized questionnaires and clinician-administered scales. These assessments help quantify emotional and psychological responses, but they are influenced by individual perception biases, which can affect reliability. To mitigate these biases, researchers typically first expose participants to stress-inducing or anxiety-inducing tasks before administering questionnaires. This approach ensures that self-reports correspond to real-time physiological reactions rather than general perceptions. Additionally, it allows researchers to correlate physiological data with subjective ratings, enhancing the validity of stress and anxiety assessments. Table 2.3 summarizes commonly used stress and anxiety assessment tests.

Table 2.4: List of features extracted from different physiological signals.

Signals	Domain	Features
EEG	Time Domain (TD)	Statistical: Mean, Median, Variance, Standard Deviation, Peak-to-peak; Hjorth Features: Complexity, Mobility, and Activity; Higuchi features [31] [140]
	Frequency Domain (FD)	Power Spectral Density (PSD), Relative Power (RP), Band power, Band Ratio, [277][24]
	Entropy	Spectral Entropy, Sample Entropy, Shannon entropy [244]
ECG	Statistical	Mean_HRV, Median_HRV, Standard Deviation_HRV
	Time Domain (TD)	Mean Heart Rate, Std_HR, Max, Min HR, RR Interval, RMSSD, NN50 Count, Triangular Index, TINN, CVSD, CVNNI, SampEn
	Frequency domain (FD)	Power Spectral Density (PSD), LF, HF, LF/HF ratio, LFnu, HFnu, Total Power [170] [197] [82]
EDA	Statistical	Statistical features applied to: SCR and SCL signal [80]
	Tonic (SCL)	Slowly fluctuating baseline patterns in the EDA signal, reflecting continuous variations in skin conductance levels and representing overall arousal state and tonic sympathetic activity.
	Phasic (SCR)	Rapid and short-term fluctuations in EDA signal represent momentary responses to stimuli or events and reflect the phasic sympathetic activity
RESP	Temporal	Statistical and Respiratory Rate Variability (RRV) features [80]
BVP	Statistical Features	Mean BVP, standard deviation, maximum, minimum, power VLF BVP, power LF BVP, power HF BVP

2.3 Physiological Signals for Stress and Anxiety Recognition

Physiological signals serve as biological indicators that capture the body's responses to external and internal stimuli, offering valuable insights into the functioning of underlying physiological systems [85]. Figure 2.1 shows how physiological signals are used in stress and anxiety classification, highlighting which signals are most commonly utilized in existing research. Physiological signals exhibit complex patterns and require advanced preprocessing techniques to minimize noise and extract meaningful information [254] [133]. Effective analysis of physiological signals relies on feature extraction, which transforms raw signal data into numerical representations suitable for ML models. This process plays a critical role in stress and anxiety classification by identifying signal characteristics that contribute to accurate classification.

Table 2.5: Relationship between physiological signals, their features, and corresponding responses to anxiety and stress.

Signals	Extracted Feature	Anxiety Response		Stress Response	
EEG	PSD (α wave)	[61], [47]	$\downarrow$	[12], [38], [157], [61], [67]	$\downarrow$
	PSD (β wave)	[1], [47]	$\uparrow$	[194], [91], [98], [224], [152]	$\uparrow$
	PSD (θ wave)	[1], [194], [270], [126], [59]	$\uparrow$	[92], [159]	$\downarrow$
	PSD (δ wave)	[177], [86]	$\uparrow$	[184], [246]	$\downarrow$
	θ/β ratio	[7], [256]	$\uparrow$	[7], [53], [258]	$\uparrow$
	δ/β ratio	[8]	$\uparrow$	—	–
	β/α ratio	[194]	$\uparrow$	[220]	$\uparrow$
ECG	Heart Rate (HR)	[109], [145] $\uparrow$			
	Heart Rate Variability (HRV)	[170], [5] $\downarrow$			
	LF/HF ratio	[277] $\uparrow$			
	Low Frequency (LF)	[277] $\uparrow$			
	High Frequency (HF)	[277] $\downarrow$			
	RMSSD	[277] $\downarrow$			
EDA	Skin Conductance Response (SCR)	[228] $\uparrow$			
	Skin Conductance Level (SCL)	[228] $\uparrow$			
RESP	Respiration Rate (RR)	[85] $\uparrow$			

Note: PSD = Power Spectral Density, HR = Heart Rate, HRV = Heart Rate Variability, LF = Low Frequency, HF = High Frequency, RMSSD = Root Mean Square of Successive Differences.

This thesis extracts features in two primary domains, each capturing distinct signal properties:

1. *Time-Domain Features:* Computed directly from raw signals, these features describe amplitude variations, variability, and trends over time, making them computationally efficient and interpretable.

2. *Frequency-Domain Features:* Obtained by applying Fast Fourier Transform (FFT) to convert signals into the frequency domain, these features reveal periodicity and rhythmic activity, aiding in identifying stress-related patterns.

Table 2.4 presents the key physiological features associated with anxiety. Furthermore, Table 2.5 presents the relationship between physiological signals and their response to stress and anxiety. A detailed explanation of each signal and its influence is provided in the following subsections.

2.3.1 Electroencephalographic signals (EEG)

Electroencephalography (EEG) is a widely used neuroimaging technique that records electrical activity in the brain through electrodes placed on the scalp [168]. Due to its high temporal resolution, non-invasiveness, affordability, and lack of radiation exposure, EEG has become a practical tool for real-time stress and anxiety monitoring [252]. Researchers have increasingly employed EEG-based approaches to assess stress and anxiety, as these conditions significantly alter brain activity patterns [64, 107, 24].

As the central organ for cognitive and emotional processing, the brain consists of distinct regions that contribute to various functions [89], as illustrated in Figure 2.2. Specifically, the human brain is divided into four distinct regions [230]: The *frontal lobe*, particularly the prefrontal cortex, is responsible for higher-order cognitive tasks such as decision-making, problem-solving, attention, and emotional regulation [117]. The *temporal lobe* governs auditory processing, language comprehension, and memory formation while also aiding in the recognition and interpretation of emotions [243]. The parietal lobe integrates sensory information related to touch, spatial awareness, and proprioception, enabling coordinated movement and perception of the external

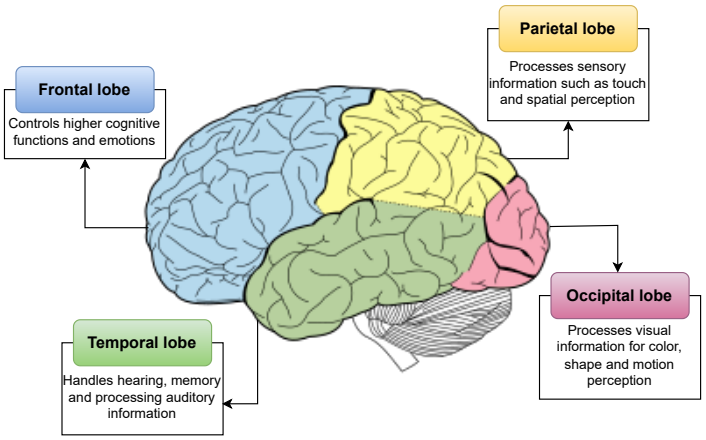


Figure 2.2: Illustrating human brain regions and key functional areas involved in cognitive and emotional processing.

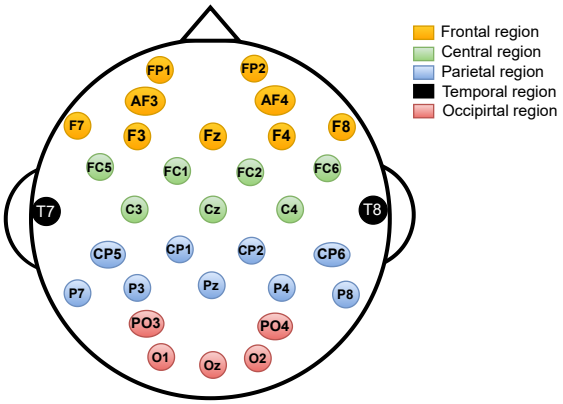


Figure 2.3: 32 channel EEG headset montage mapping over brain regions.

environment [230]. Finally, the occipital lobe primarily involves visual processing, allowing the brain to interpret shapes, colors, and motion [34].

Further, to capture the neural activity associated with these functions, EEG electrodes are placed according to the 10-20 international system [111], as illustrated in Figure 2.3. Frontal electrodes (Fp1, Fp2, F3, F4, F7, F8) record cognitive and executive functions [215], temporal electrodes (T3, T4, T5, T6) capture auditory and memory-related activity [33], and parietal and occipital electrodes (P3, P4, Pz, O1, O2) monitor sensory integration and visual processing [111]. These electrode placements comprehensively understand brain activity across different cognitive and perceptual domains, supporting various neurological and psychological studies. These studies confirm that frontal channels consistently provide discriminative features for stress-

related EEG classification. At the same time, the involvement of parietal and occipital regions often depends on the nature of the task and the features extracted.

Several studies highlight that stress predominantly affects the frontal brain region. Sharma et al. [223] reported significant activity in channels Fp1, Fp2, F3, F4, F7, F8, and Fz. Varshney et al. [247] achieved higher classification accuracy using entropy features from O2, Fz, Cz, and Pz with RNN-based models, emphasizing the importance of frontal and parietal areas. Further, Wang et al. [253] selected F8, F3, AF3, and O2 for mental arithmetic task recognition and achieved an accuracy of 97.11% using an SVM classifier. These studies confirm that frontal channels consistently provide discriminative features for stress-related EEG classification. At the same time, the involvement of parietal and occipital regions often depends on the nature of the task and the features extracted.

EEG signals are further categorized into five primary frequency bands, each associated with distinct neural functions [10] [148], as illustrated in Figure 2.4. Delta (δ) waves (0.1 - 4Hz) are the slowest and are predominantly observed during deep sleep and unconscious states [17]. Theta (θ) waves (4–8 Hz) are linked to drowsiness, light sleep, and relaxation, often appearing during meditative or introspective states [214]. Alpha (α) waves (8–12 Hz) are associated with a relaxed yet alert state, commonly present when an individual is awake but calm [36]. Beta (β) waves (12–30 Hz) reflect active cognitive processing, problem-solving, and focused attention [135]. Lastly, Gamma (γ) waves (>30 Hz) are the fastest frequency band and are associated with heightened cognitive function, sensory processing, and states of high mental engagement [102].

2.3.1.1 Significance of EEG for Stress and Anxiety

EEG is a valuable tool for understanding stress and anxiety as it provides direct insights into neural activity and cognitive processing. Unlike peripheral physiological signals, ECG, EDA, and HRV, which primarily reflect autonomic responses, EEG enables a localized assessment of brain activity changes under stress-inducing conditions [182] [37] [85]. Studies have shown that increased beta activity, altered theta-to-beta ratios, and reduced alpha power are commonly associated with heightened stress and anxiety

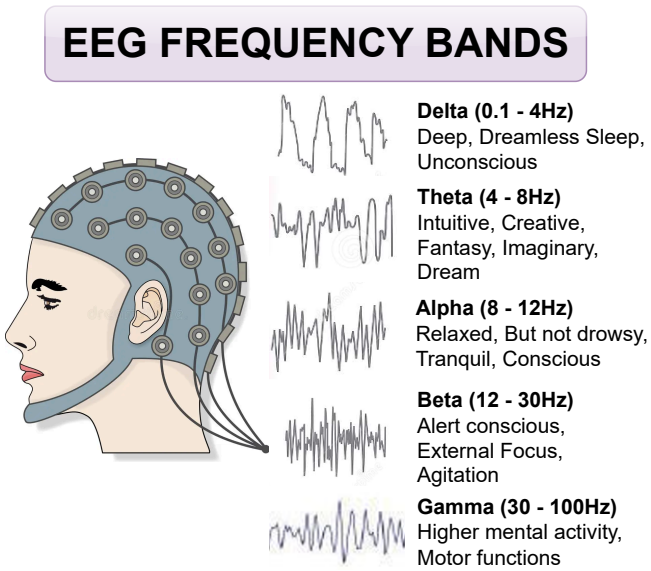


Figure 2.4: EEG frequency bands and their corresponding brain state.

states [61] [47][194] [55]. These EEG-based markers are viable for differentiating stress and improving classification accuracy [124].

2.3.1.2 Features from EEG signals

EEG feature extraction is essential for analyzing brain activity patterns associated with stress and anxiety. These features are typically derived from distinct domains, including time and frequency, each capturing unique signal features.

- **Time Domain (TD):** Time-domain features capture the EEG signal's amplitude variations and temporal properties. These features provide essential insights into signal dynamics and include the following features [86].

    1. **Statistical Features:** Statistical features represent the distribution and central tendency of the EEG signal, providing insights into its overall amplitude variations and variability.

        (a) *Mean:* Represents the average amplitude of the EEG signal over a given period.

$$Mean = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.1}$$

Equation 2.1 computes the mean where N is the total number of EEG samples, and x_i represents the amplitude at time i.

(b) *Variance:* Measures the dispersion of signal values around the mean.

$$Variance = \frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2 \tag{2.2}$$

Equation 2.2 computes variance where σ^2 represents the variance, S is the mean, N is the total number of EEG samples, and x_i denotes the amplitude at time i.

(c) *Standard Deviation:* Quantifies the degree of fluctuation in the EEG signal, indicating signal stability.

$$StandardDeviation = \sqrt{variance} \tag{2.3}$$

Equation 2.3 computes standard deviation where σ represents the standard deviation, and σ^2 is the variance.

(d) *Peak-to-Peak:* Represents the difference between the maximum and minimum values of the signal, as shown in Equation 2.4.

$$Peak - to - Peak = peak_{high} - peak_{low} \tag{2.4}$$

where P_{high} is the highest amplitude in the signal, and P_{low} is the lowest amplitude in the signal.

(e) *Skewness:* It measures the asymmetry of the data distribution, indicating negative or positive skewness. Equation 2.5 shows skewness formula:

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2)}{(\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2)^{\frac{3}{2}}} \tag{2.5}$$

where N is the total number of EEG samples, x_i is the amplitude at time i, and S is the mean value of the signal.

(f) *Kurtosis:* It measures the tailedness of the data distribution, identifying

whether the data has heavy or light tails. Equation 2.6 shows kurtosis formula:

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^4)}{(\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2)^2} - 3 \tag{2.6}$$

where N is the total number of EEG samples, x_i is the amplitude at time i, and S is the mean value of the signal.

2. **Hjorth Features:** It provides key insights into the signal by measuring activity, mobility, and complexity, making them useful for EEG analysis.

   (a) *Activity:* Activity measures the signal's variance, indicating its overall power level, which can be categorized as low or high [31] [154]. Equation 2.7 shows activity formula, where y(i) is the EEG signal:

   $$Activity = Var(y(i)) \tag{2.7}$$

   (b) *Mobility:* It is calculated as the square root of the variance of the first derivative divided by the variance of the signal, providing an estimate of its frequency content [31] [154]. Equation 2.8 shows mobility formula:

   $$\text{Mobility} = \sqrt{\frac{var(\frac{dy}{dt})}{var(y(t))}} \tag{2.8}$$

   where $y(t)$ is the EEG signal, $\frac{dy}{dt}$ is its first derivative, and var$(\cdot)$ denotes variance.

   (c) *Complexity:* It is defined as the ratio of the mobility of the signal's first derivative to the mobility of the original signal. It quantifies how the frequency content varies over time, reflecting the signal's structural intricacy [31] [154]. Equation 2.9 shows complexity formula:

   $$\text{Complexity} = \frac{\text{Mobility}(\frac{dy}{dt})}{\text{Mobility}(y(t))} \tag{2.9}$$

   where $y(t)$ is the EEG signal, and $\frac{dy}{dt}$ is its first derivative.

3. **Higuchi Features:** It quantifies the complexity of a signal by analyzing its self-similarity across different time scales. It is particularly useful for

measuring irregularities in EEG signals and detecting subtle changes in brain activity [185].

$$HFD(k) = \frac{log(N)}{log(\frac{N}{k.L(k)})'} \tag{2.10}$$

Equation 2.10 computes higuchi where $HFD(k)$ represents the Higuchi Fractal Dimension at scale k, N is the total length of the EEG signal, and $L(k)$ is the average length of the curve over scale k.

- **Frequency Domain (FD):** Fast Fourier Transform (FFT) converts the time-domain EEG signal into the frequency domain. Researchers extract frequency domain features from the power spectrum, representing how signal power is distributed across different frequency bands. The following features are commonly used for EEG analysis:

  1. **Power Spectral Density (PSD):** It quantifies the distribution of signal power across different frequencies [91] [29] [25]. Researchers commonly compute it using the Welch method to ensure a reliable estimation of power variations over time [258] [20] [15] [15].
  2. **Spectral Entropy:** Measures the randomness in the power spectrum, indicating how evenly power is distributed across frequency bands. A higher value suggests a more complex and unpredictable signal [15] [206]. Equation 2.11 represents the formula for spectral entropy.

$$SE = -\sum_{f=4}^{f=N} P\bar{S}D(F)log(P\bar{S}D(F)) \tag{2.11}$$

  where $P\bar{S}D(f)$ is the normalized PSD at frequency f.

  3. **Sample Entropy:** Estimates the complexity of a time-series signal by quantifying the likelihood of similar patterns occurring. It is useful for analyzing EEG signal irregularities [100] [223]. Equation 2.12 represents the formula for sample entropy.

$$SampEn(m,r,N) = -\ln\left(\frac{A}{B}\right) \tag{2.12}$$

where m is the embedding dimension, r is the tolerance threshold, N is the total number of data points, A is the number of matching patterns of length $m+1$, and B is the number of matching patterns of length m.

4. **Shannon Entropy:** Represents the uncertainty or disorder in a signal by evaluating the probability distribution of its values. Higher entropy indicates a more unpredictable signal [223]. Equation 2.13 represents the formula for Shannon entropy.

$$SE = -\sum_{i=1}^{N} P(x_i)\log P(x_i) \tag{2.13}$$

where N is the total number of discrete signal values, and $P(x_i)$ is the probability of occurrence of signal value x_i.

5. **Relative Power:** Computes the proportion of power in a specific frequency band relative to the total power of the signal, helping to analyze dominant brainwave activity [10] [25]. Equation 2.14 represents the formula for relative power.

$$RP = \frac{power(band)}{power(allbands)} * 100 \tag{2.14}$$

6. **Band Power:** Band Power quantifies the total signal power within a specific EEG frequency band, providing insights into cognitive and emotional states.

7. **Band Ratio:** Calculates the ratio between the bandpower of different EEG frequency bands, such as θ/β ratio, to assess cognitive and emotional states. This feature is commonly used in stress and anxiety classification [10] [148] [187].

2.3.1.3 Related work of EEG on Stress and Anxiety

EEG-based studies consistently demonstrate that stress and anxiety influence brainwave activity in distinct ways. Elsadek et al. [67] reported that alpha activity increases during

relaxed states but significantly decreases under stress, indicating its sensitivity to stress-related neural changes. Similarly, Giannakakis et al. [85] observed that stress-induced stimuli cause a substantial rise in spectral power, particularly in the frontal and central brain regions, reinforcing their role in stress processing. In contrast, Katmah et al. [124] showed that gamma oscillations exhibit inconsistent responses across stress and relaxed conditions, suggesting that gamma activity may not reliably serve as a biomarker for stress classification.

Further investigations highlight the beta frequency band's association with stress responses. Sharma et al. [224] established that beta power significantly increases during stressful situations, emphasizing its link to emotional intensity and stress perception. Attallah et al. [26] demonstrated that frontal electrodes effectively classify stress, underscoring the prefrontal cortex's role in stress and anxiety regulation. Additionally, Masood et al. [152] identified a strong correlation between fast beta activity and high-stress phases, showing that beta oscillations intensify under increased stress levels.

For anxiety-related EEG patterns, studies consistently report an increase in delta and theta activity in individuals diagnosed with anxiety disorders [177] [189]. Falconer et al. [72] further confirmed that frequent transitions between frequency bands distinguish anxiety states from normal and stress conditions. These findings collectively suggest that while beta activation and alpha suppression characterize stress responses, anxiety is marked by increased slow-wave activity and greater instability across frequency bands.

2.3.2 Electrodermal Activity (EDA)

EDA, also known as Galvanic Skin Response (GSR), is a reliable measure for predicting stress levels by analyzing the skin's electrical conductance, which reflects sympathetic nervous system activation through sweat gland activity [144] [176]. Tonic and phasic activity are the two information sources within a raw EDA signal, as shown in Figure 2.5.

1. *Tonic activity* represents the baseline level of skin conductance influenced by factors such as skin dryness, moisture, and temperature. The Skin Conductance Level (SCL), a common measure of tonic activity, indicates overall physiological arousal in the absence of specific stimuli [137] [88].

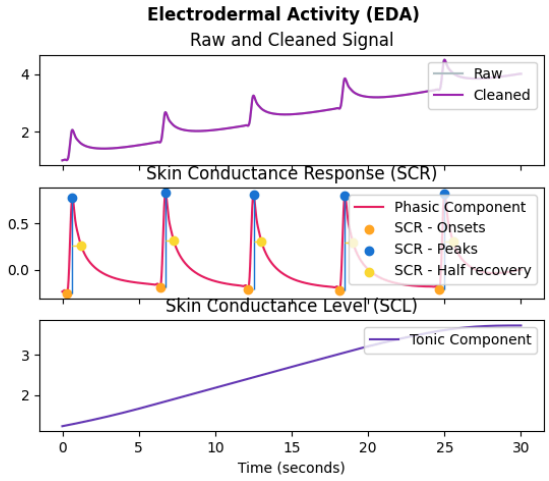


Figure 2.5: Figure 2.7 shows the decomposition of a raw EDA signal into its phasic and tonic components.

2. *Phasic activity* captures rapid changes in conductance, known as Skin Conductance Responses (SCR), which occur in response to emotionally stimulating events. These short-term peaks reflect momentary changes in arousal caused by sympathetic nervous system activation [137] [88].

2.3.2.1 Significance of EDA for Stress and Anxiety

EDA is a key physiological marker for assessing stress and anxiety, as it directly reflects sympathetic nervous system activation through sweat gland activity [137] [134]. Stress and anxiety trigger an increase in skin conductance, resulting in elevated SCL and more frequent SCR [106]. The increased EDA activity indicates physiological arousal associated with the body's fight-or-flight response [113]. Studies have shown that individuals experiencing acute stress or anxiety exhibit increased EDA [175] [160], while lower conductance levels are observed in relaxed states [218]. Due to its high sensitivity to autonomic changes, EDA is widely used in stress classification models and real-time mental health monitoring.

2.3.2.2 Features from EDA Signal

Extracting meaningful features from the EDA signal is essential for assessing autonomic responses related to stress and anxiety. These features are derived from tonic (SCL)

and phasic (SCR) components, capturing variations in skin conductance over time. The following section outlines key EDA features used in stress and anxiety classification.

- **Statistical Features:** Statistical features quantify variations in the EDA signal by analyzing its distribution, central tendency, and dispersion over time.

    1. **Mean:** Represents the average EDA value over a defined period, serving as a baseline indicator of skin conductance and overall physiological arousal [176] [137].
    2. **Variance:** Quantifies the dispersion of EDA signal values over a specified period, indicating the extent of fluctuation in skin conductance [176] [137].
    3. **Skewness:** Evaluates the asymmetry of the EDA signal distribution relative to its mean. A positive skew indicates a longer right tail, whereas a negative skew reflects a longer left tail in the data distribution [176] [137].
    4. **Kurtosis:** Assesses the sharpness of the EDA signal distribution, indicating the presence of outliers or extreme values. Higher kurtosis signifies more pronounced peaks and outliers, while lower kurtosis suggests a flatter distribution [144][248].
    5. **Standard Deviation:** Measures the extent of variation in EDA values from the mean. A higher standard deviation indicates greater fluctuations in skin conductance, reflecting increased physiological variability [144] [109].

- **Tonic Features:** Slowly fluctuating baseline patterns in the EDA signal reflect continuous variations in SCLs and represent overall arousal state and tonic sympathetic activity.

    1. **Tonic Component:** Represents the slow, baseline variations in skin conductance, typically extracted using low-pass filtering. It reflects the overall level of physiological arousal over time [155] [137].
    2. **SCL (Tonic) Slope:** Represents the rate of change in the tonic component of the EDA signal, typically determined by fitting a linear model to the baseline. It indicates the trend or directional shift in baseline skin conductance over time [218] [176].

3. **Mean SCL Value:** Represents the average level of the tonic component over a specified period, providing a baseline measure of skin conductance.
4. **Variance of SCL:** Quantifies the average squared deviation of SCL values from the mean, indicating the extent of fluctuation in the tonic component of skin conductance [218].
5. **Standard Deviation of SCL:** Calculates the square root of the variance of SCL, representing the degree of dispersion in the tonic component of skin conductance [218] [144].

- **Phasic Features:** Rapid and short-term fluctuations in the EDA signal represent momentary responses to stimuli or events and reflect the phasic sympathetic activity.

  1. **SCR (Phasic) Peaks:** Represents the amplitude and timing of response peaks in the phasic component of the EDA signal, indicating rapid changes in skin conductance due to stimuli. Higher peaks reflect stronger physiological responses to specific events [218].
  2. **Mean SCL Value:** Represents the average level of the tonic component over a defined period, providing a baseline measure of skin conductance [144].
  3. **Variance of SCL:** Quantifies the dispersion of SCL values by measuring the average squared deviation from the mean, indicating fluctuations in the tonic component of skin conductance [144].
  4. **Standard Deviation of SCL:** Represents the square root of the variance of SCL, indicating the degree of fluctuation in the tonic component of skin conductance [155].

2.3.2.3 Related work of EDA on Stress and Anxiety

Several studies have shown EDA as a reliable physiological marker for stress classification. In various studies, EDA is also referred to as GSR. Can et al. [42] analyze EDA, heart activity, and acceleration data in a study involving 21 participants and show that SCL and SCR significantly vary across stress conditions. Their findings

highlight the effectiveness of combining EDA with other physiological signals for stress classification. Kumar et al. [134] examined EDA and multiple physiological signals and confirmed that EDA-based features effectively differentiate stress, amusement, and neutral states. Similarly, Garg et al. [79] validated that statistical EDA features such as standard deviation, mean, minimum, and maximum values are useful stress indicators. The study also emphasized that EDA, combined with other physiological signals, effectively classifies stress.

Further, Albertetti et al. [14] investigated EDA responses to emotional, intellectual, physical, and pain-induced stressors in 6 participants, demonstrating that phasic and tonic EDA features accurately capture stress responses. Milstein et al. [156] compared Empatica E4's EDA and HRV measurements in 30 participants and showed that Empatica E4 reliably records mean IBI and HR, but struggles with HRV accuracy under movement conditions. Similarly, Zhu et al. [278] conducted a large-scale comparative study using multiple datasets and concluded that EDA-based stress classification achieves high classification accuracy, particularly when combined with HRV. Their findings suggest that 30-second sliding window processing enhances EDA-based stress classification.

In anxiety studies, Markiewicz et al. [149] observed that EDA increases when a person experiences anxiety, when the sympathetic nervous system activates the "fight or flight" response. This physiological reaction represents the body's adaptive mechanism for coping with potential threats. Najafpour et al. [173] and Singh et al. [234] further confirmed that elevated EDA levels correlate with emotional arousal, making it an effective indicator of anxiety intensity.

2.3.3 Heart Rate (HR) and Heart Rate Variability (HRV)

Heart Rate (HR) is defined as the number of heartbeats per minute and is a primary indicator of cardiac function and ANS activity [50]. Researchers often analyze the Inter-Beat Interval (IBI) to measure HR, which represents the time between consecutive heartbeats. On the other hand, Heart Rate Variability (HRV) also focuses on the variation in IBI over time. It reflects how the ANS dynamically adapts to internal and external stimuli, such as stress and anxiety [139][138]. While HR presents a measure

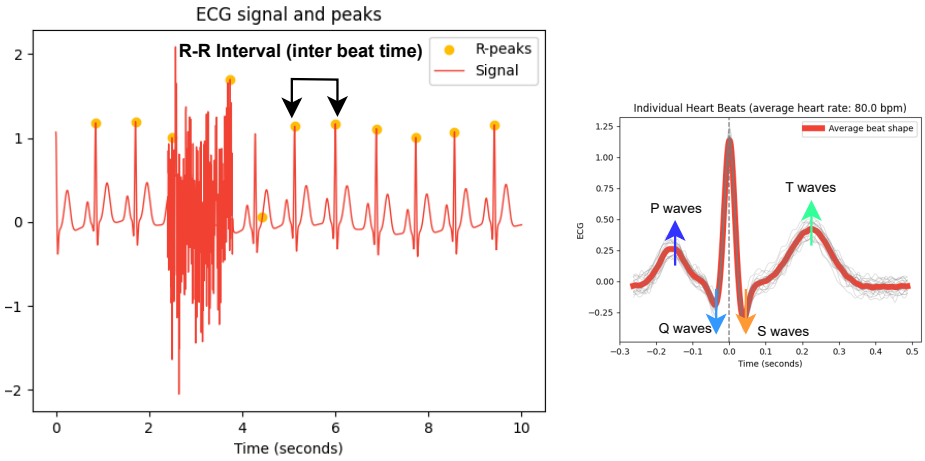


Figure 2.6: A normal Electrocardiogram (ECG) [196].

of overall cardiac activity, HRV quantifies the subtle variations in the timing between consecutive heartbeats that provide deeper insights into emotional and physiological states. HR and HRV are reliable markers for assessing mental health and stress responses [85].

Researchers commonly measure the Inter Beat Interval (IBI) using two key physiological sensing methods:

1. **Electrocardiography (ECG):** ECG devices record the electrical activity of the heart and produce a waveform comprising four primary components: the baseline, P wave, QRS complex, and T wave, as illustrated in Figure 2.6. Researchers calculate the RR interval—the time between successive R-peaks—from this waveform, which directly corresponds to the IBI [90]. ECG offers precise cardiac measurements but typically requires careful sensor placement and is more prone to motion artifacts in mobile settings.

2. Photoplethysmography (PPG): PPG sensors measure blood volume fluctuations in peripheral tissues through a non-invasive optical technique. Similar to ECG, PPG signals allow the extraction of IBI and subsequent calculation of HR and HRV. Due to its simplicity, lower cost, and integration into wearable devices, PPG supports continuous and unobtrusive monitoring of stress and anxiety, making it highly suitable for real-world applications [81][116].

2.3.3.1 Significance of HR and HRV for Stress and Anxiety

HR and HRV are important physiological indicators for detecting stress and anxiety as they reflect ANS activity [128][5][51]. During stressful or anxious states, sympathetic activation increases while parasympathetic control decreases. This leads to elevated HR, reduced HRV, and shorter IBIs [109][145]. These responses indicate heightened physiological arousal associated with the fight-or-flight mechanism.

Low HRV is a well-known marker of chronic stress and anxiety disorders and reflects poor autonomic flexibility and limited recovery capacity [128]. A decrease in IBI also highlights autonomic imbalance under such conditions [156]. Extracting and analyzing HR and HRV features improves the accuracy of stress and anxiety classification models across different physiological sensing methods. Devices such as ECG and PPG provide signals used to derive HR and HRV. ECG records cardiac activity but requires precise sensor placement and is more susceptible to environmental noise and motion artifacts, posing challenges for wearable applications. However, PPG offers a simpler setup and greater suitability for continuous monitoring [42].

2.3.3.2 Features from HR and HRV

Researchers extract HR and HRV features from IBIs obtained through ECG or PPG sensors. These features are categorized into different domains, including time, frequency, and non-linear measures, each providing valuable insights into stress and anxiety classification.

- **Time Domain:** It directly analyzes the ECG waveform, capturing variations in HR, HRV, and signal amplitudes over time. These features provide essential insights into ANS activity and cardiac responses under stress and anxiety conditions.

    1. **Mean Heart Rate:** Computes the average number of heartbeats per minute using RR intervals [57] [156].
    2. **Standard Deviation of Heart Rate (Std_HR):** Measures the variability in heart rate values over the recording period, indicating fluctuations in cardiac activity [51] [156].

3. **Maximum and Minimum Heart Rate (Max, Min HR):** Represents the highest and lowest heart rate recorded within a given period, reflecting the range of cardiac activity [144].
4. **RR Interval:** Represents the time between two successive R-wave peaks in the ECG signal, serving as a key indicator of heart rate variability (HRV) [144] [235].
5. **Root Mean Square of Successive Differences (RMSSD):** Quantifies beat-to-beat variability in heart rate by computing the square root of the mean of squared differences between successive RR intervals. It serves as a key indicator of parasympathetic nervous system activity [42] [51] [244].
6. **NN50 Count:** Represents the number of consecutive RR intervals that differ by more than 50 ms, providing an additional measure of heart rate variability (HRV) [42].
7. **Triangular Index:** Computes heart rate variability by dividing the total number of RR intervals by the height of the peak in the RR interval histogram. It provides an overall measure of HRV [42] [244].
8. **Triangular Interpolation of NN Interval Histogram (TINN):** Represents the baseline width of the RR interval histogram, obtained by interpolating the area under the histogram into a triangular shape. It serves as an indicator of overall heart rate variability (HRV).
9. **Coefficient of Variation of Successive Differences (CVSD):** Calculates the ratio of the root mean square of successive differences (RMSSD) to the mean RR interval, providing a normalized measure of heart rate variability (HRV).
10. **Coefficient of Variation of NN Intervals (CVNNI):** Computes the ratio of the standard deviation of RR intervals to the mean RR interval, expressed as a percentage, to quantify heart rate variability (HRV) [51].
11. **Sample Entropy (SampEn):** Quantifies the complexity and irregularity of the RR interval time series by calculating the negative logarithm of the conditional probability that similar patterns remain consistent in subsequent data points [51] [42].

- **Frequency-Domain Features:** Analyze power distribution across various frequency bands to provide insights into heart rate variability (HRV). These features are derived by transforming the time-domain ECG signal into the frequency domain using the FFT.

    1. **Power Spectral Density (PSD):** Represents power distribution across different frequency bands, offering insights into the energy levels within various frequency ranges of the ECG signal. It is commonly estimated using the Welch method [5] [109].
    2. **Low-Frequency Power (LF, 0.04–0.15 Hz):** Represents the combined influence of sympathetic and parasympathetic activity, providing insights into ANS modulation [51].
    3. **High-Frequency Power (HF, 0.15–0.4 Hz):** Primarily reflects parasympathetic nervous system activity and is closely associated with respiratory sinus arrhythmia, indicating vagal influence on heart rate variability (HRV) [51].
    4. **LF/HF Ratio:** Represents the balance between sympathetic and parasympathetic nervous system activity by computing the ratio of low-frequency to high-frequency power in heart rate variability (HRV) [85].
    5. **Low-Frequency Power in Normalized Units (LFnu):** Represents the power within the low-frequency range (0.04–0.15 Hz) as a percentage of total power, excluding the very high-frequency (VHF) component. It is linked to both sympathetic and parasympathetic activity, with a stronger emphasis on sympathetic modulation [86].
    6. **High-Frequency Power in Normalized Units (HFnu):** Represents the power within the high-frequency range (0.15–0.4 Hz) as a percentage of total power, excluding the very-low-frequency (VLF) component. It is primarily associated with parasympathetic nervous system activity [85].
    7. **Total Power:** Represents the overall energy in the ECG signal by summing the power across all frequency bands, providing a comprehensive measure of heart rate variability (HRV) [85].

2.3.3.3 Related work of ECG on Stress and Anxiety

ECG-derived HRV features serve as reliable indicators for stress classification, with several studies confirming their effectiveness. Castaldo et al. [46] demonstrated that ultra-short-term HRV features with a duration of 1–3 min are enough to classify mental stress accurately and identified mean NN, SDNN, mean HR, and HF power as key markers. Further, Dalmeida et al. [51] validated HR, SDNN, RMSSD, and pNN50 as stress-sensitive features and showed significant variations across different stress levels. In another study, Albertetti et al. [14] highlighted SDNN, RMSSD, HF, and LF/HF ratio as robust physiological markers across emotional, intellectual, and physical stressors. It is observed that despite the reliability of HRV-based features, recording duration, movement artifacts, and device limitations influence the accuracy of stress assessment [156] [46].

In anxiety-related research, Ihmig et al. [109] and Lueken et al. [145] observed that HRV decreases while HR increases during anxiety and mental stress. Zheng et al. [277] further show that an anxious state leads to a reduction in HF, HRV, and an increase in LF, with a concurrent rise in the LF/HF ratio and RMSSD, reinforcing their significance as key physiological markers of anxiety.

2.3.4 Blood Volume Pressure (BVP)

PPG sensors also provide Blood Volume Pressure (BVP) signals by measuring blood volume changes in peripheral tissues. Stress and anxiety directly influence the cardiovascular system, causing noticeable alterations in BVP waveforms. During stressful conditions, the sympathetic nervous system increases heart rate, blood pressure, and vascular tone, altering BVP waveforms. Figure 2.7 shows the raw and filtered BVP signals and the derived heart rate, illustrating how sympathetic activation influences the pulse waveform. Increased sympathetic activity raises pulse rates and modifies pulse waveform morphology, highlighting ANS imbalance [278] [121] [176] [207]. Additionally, stress and anxiety reduce heart rate variability (HRV), especially in individuals with chronic stress or anxiety disorders [85]. These BVP-derived variations provide crucial insights into the physiological mechanisms of stress and anxiety, emphasizing the importance of BVP analysis for stress monitoring and

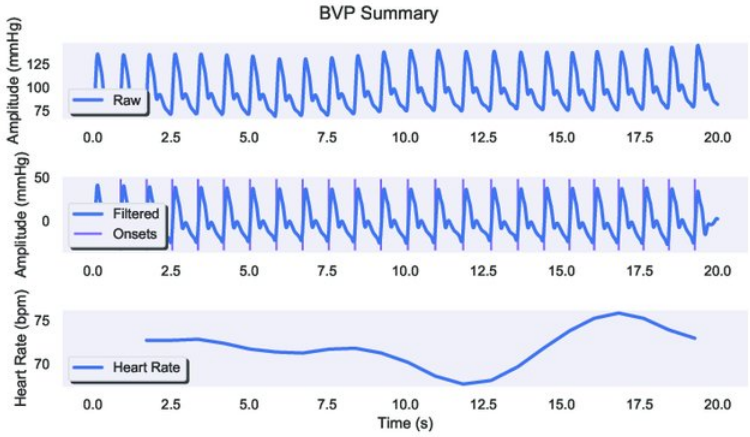


Figure 2.7: Raw Blood Volume Pulse (BVP) sensor signal, filtered signal, and its main data features (onsets and HR). [40].

assessment.

2.3.4.1 Significance of BVP for Stress and Anxiety

Blood Volume Pressure (BVP) is a key physiological marker for stress and anxiety assessment by reflecting cardiovascular autonomic regulation [85]. BVP signals capture pulse rate, vascular tone, and HRV, which are influenced by sympathetic and parasympathetic activity. Increased sympathetic activation under stress leads to higher pulse rates and altered waveform morphology, while reduced HRV is commonly observed in individuals experiencing chronic stress or anxiety [14]. Due to its non-invasive measurement and sensitivity to autonomic changes, BVP is widely used for stress classification and real-time physiological monitoring.

2.3.4.2 Features from BVP Signal

Feature extraction from BVP signals provides valuable insights into cardiovascular dynamics and ANS regulation. The following are key BVP features used in stress and anxiety classification:

1. **Mean BVP:** Represents the average BVP value over a given period, indicating baseline blood volume changes [175] [207].

2. **Standard Deviation of BVP:** Measures the variability in BVP signal amplitude, reflecting fluctuations in vascular tone [79].

3. **Maximum BVP:** Captures the highest recorded BVP value, representing peak pulse amplitude [175] [79].

4. **Minimum BVP:** Records the lowest BVP value, indicating the minimum blood volume in a pulse cycle [207].

5. **Power VLF BVP:** Quantifies the power in the very-low-frequency (VLF, <0.04 Hz) range, associated with metabolic and thermoregulatory mechanisms [175] [79] [207].

6. **Power LF BVP:** Measures power in the low-frequency (LF, 0.04–0.15 Hz) range, reflecting sympathetic and parasympathetic activity [175] [79] [207].

7. **Power HF BVP:** Represents power in the high-frequency (HF, 0.15–0.4 Hz) range, primarily linked to parasympathetic nervous system influence and respiratory activity [175] [79] [207].

2.3.4.3 Related work of BVP on Stress and Anxiety

Kumar et al. [134] analyzed BVP, ECG, EDA, and additional physiological signals from 15 subjects and observed that BVP features effectively differentiate between three stress levels, particularly when combined with HRV metrics. Similarly, Garg et al. [79] investigated multimodal stress signals, including BVP, ECG, and EDA, from 15 participants undergoing the TSST and identified that statistical BVP features such as standard deviation, mean, minimum, and maximum values significantly correlate with stress states. Their findings confirmed that statistical BVP features effectively distinguish stress from relaxation periods. Further, Nath et al. [176] explored BVP-based stress classification in 19 older adults, incorporating cortisol levels as a biomarker to enhance stress classification accuracy. The authors find that BVP amplitude and heart rate fluctuations correspond to variations in cortisol concentration during stress and recovery periods.

In anxiety studies, Rodríguez-Arce et al. [197] explored anxiety and stress recognition in academic environments using BVP, EDA, heart rate, and respiration in 21 participants. The authors conducted time-constrained arithmetic tasks as stress-inducing stimuli and found that BVP features significantly contribute to anxiety

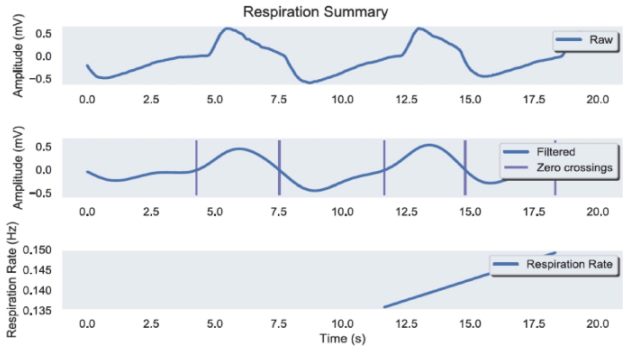


Figure 2.8: Raw and filtered RESP signal, filtered signal, and its features: zero-crossings, respiration rate [40].

classification. The results suggest that physiological signals provide a more reliable anxiety assessment than subjective questionnaires like STAI scores.

2.3.5 Respiration (RESP)

Respiration (RESP) regulates the inhalation and exhalation cycle, as illustrated in Figure 2.8. The primary respiratory metrics include breath rate and breath depth (amplitude). Respiration rate typically increases due to heightened emotional arousal during stress and anxiety, whereas relaxation leads to a slower breathing rate [87]. Additionally, stressful situations may cause temporary interruptions in breathing patterns. Additionally, Respiratory Rate Variability (RRV), which measures the variation in breathing intervals over time, is a valuable marker of ANS dynamics during stress and anxiety.

2.3.5.1 Significance of RESP for Stress and Anxiety

Respiration is a crucial physiological indicator of ANS activity, making it valuable for stress and anxiety assessment [71]. Increased emotional arousal during stressful situations typically reflects increased sympathetic activation, accelerates breathing rates, and reduces breath depth, [275]. Individuals with anxiety disorders often exhibit irregular respiratory patterns, including rapid breathing, reduced respiratory sinus arrhythmia, and intermittent breath-holding [82][152]. Conversely, slow, deep breathing is associated with relaxation and a shift toward parasympathetic dominance

[82]. Although respiration alone may not definitively indicate anxiety levels, respiratory features enhance the accuracy of stress and anxiety classification when integrated with other physiological signals. Supporting this, George et al. [82] demonstrated that combining ECG and RESP features improves the identification of individuals with anxiety disorders.

2.3.5.2 Features from RESP Signal

Respiration Rate Variability (RRV) quantifies fluctuations in respiratory cycles, offering insights into ANS regulation. Researchers derive RRV features by analyzing time-domain, frequency-domain, and non-linear characteristics, each providing distinct information about respiratory dynamics. The following sections describe the extracted RRV features for stress and anxiety classification.

- **Time-Domain Features:** It measure variations in breath-to-breath intervals.

    1. **RRV_RMSSD:** Root mean square of successive differences between consecutive breath-to-breath intervals, indicates short-term variability [109] [147].
    2. **RRV_MeanBB:** Average duration of BB intervals, representing the mean respiratory cycle length [109] [147].
    3. **RRV_SDBB:** Standard deviation of BB intervals, measuring overall variability in respiratory patterns [109] [147].
    4. **RRV_SDSD:** Standard deviation of successive differences in BB intervals, reflecting short-term fluctuations in breathing rate [109] [147].
    5. **RRV_CVBB:** Coefficient of variation of BB intervals, providing a normalized measure of respiratory cycle variability [109] [147].
    6. **RRV_CVSD:** Coefficient of variation of successive differences in BB intervals, quantifying short-term variability relative to the mean [109] [147].
    7. **RRV_MedianBB:** Median value of BB intervals, offering a robust central measure of respiration cycle length [109] [147].
    8. **RRV_MadBB:** Median absolute deviation of BB intervals, assessing dispersion while reducing sensitivity to outliers [109] [147].

9. **RRV_MCVBB:** Mean coefficient of variation of BB intervals, providing a measure of normalized variability in respiratory patterns [109] [147].

- **Frequency-Domain Features:** It measures power distribution across respiratory frequency bands.

  1. **RRV_VLF:** Power in the very low-frequency range, associated with slow respiratory oscillations and metabolic regulation [109] [147].
  2. **RRV_LF:** Power in the low-frequency range, linked to sympathetic and parasympathetic modulation of respiration [109] [147].
  3. **RRV_HF:** Power in the high-frequency range, primarily reflecting parasympathetic influence and respiratory sinus arrhythmia [109] [147].
  4. **RRV_LFHF:** Ratio of low-frequency to high-frequency power, indicating the balance between sympathetic and parasympathetic control [109] [147].
  5. **RRV_LFn:** Low-frequency power normalized to total power, providing a relative measure of sympathetic activity [109] [147].
  6. **RRV_HFn:** High-frequency power normalized to total power, representing the relative contribution of parasympathetic influence [109] [147].

- **Non-Linear Features:** It captures the complexity and irregularity of breathing patterns.

- **RRV_SD1:** Standard deviation of short-term variability in the Poincaré plot, capturing rapid fluctuations in respiration [147].

- **RRV_SD2:** Standard deviation of long-term variability in the Poincaré plot, indicating overall breathing pattern stability [147].

- **RRV_SD2SD1:** Ratio of SD2 to SD1, reflecting the balance between short-term and long-term respiratory variability [147].

- **RRV_ApEn:** Approximate entropy, measuring the complexity and predictability of respiratory patterns [147].

- **RRV_SampEn:** Sample entropy, quantifying the irregularity and randomness in respiration cycles [147].

2.3.5.3 Related work of RESP on Stress and Anxiety

Several studies highlight RESP signals as key physiological indicators of stress, with variations in respiratory rate and depth showing as strong markers for stress-induced physiological changes. Siam et al. [229] investigated stress classification in drivers using RESP and other signals such as ECG, EMG, and EDA from 10 subjects and reported that changes in respiration rate and variability significantly correlate with stress levels. Seo et al. [221] examined multimodal signals, including ECG, RESP, and facial features, in 24 participants by performing the Stroop color-word test. The study identified RESP as a critical feature in distinguishing between stress levels. Further, Huang et al. [108] investigated mental workload and respiration patterns in 18 subjects, demonstrating that respiratory rate increases significantly under high cognitive load, reflecting a stress-induced physiological response. Masood et al. [152] analyzed HRV, EEG, EDA, and RESP in 24 participants exposed to mental challenges.

In anxiety studies, George et al. [82] further demonstrated that ECG and respiration features aid in identifying individuals with anxiety disorders, utilizing the Hamilton Anxiety Scale (HAM-A) scores to classify datasets into anxiety and non-anxiety groups. Similarly, Ihmig et al. [109] showed RRV features as important to detect anxiety levels. Additionally, findings indicate that statistical features such as mean respiratory rate, standard deviation, and peak-to-peak amplitude consistently emerge as strong predictors of stress and anxiety. However, motion artifacts and inter-individual variability in breathing patterns present challenges, necessitating robust preprocessing and filtering techniques for accurate stress classification. The findings also state that while stress and anxiety increase respiratory rate, RESP information serves as a supplementary biomarker rather than a primary indicator for assessing individual stress and anxiety levels [85] [43] [114].

2.3.6 Strength and Weakness of Physiological Signals

Physiological signals are advantageous because, unlike facial emotions, they cannot be masked or conditioned by voluntary human activities. As a result, physiological signal measurements convey factual information about emotional processes rather than facial expressions. Physiological signals such as heart rate, skin temperature, and EDA are

Table 2.6: Summary of physiological signals strength and weakness.

Physiological Signals	Strength	Weakness
Electroencephalography (EEG)	Measures brain activity directly; non-invasive.	Requires multiple electrodes on the scalp, resulting in a complicated setup. Susceptible to motion artifacts, necessitating preprocessing.
Electrocardiography (ECG)	Effective for emotion recognition; provides high temporal resolution for heart rate variability analysis.	Signal quality is affected by physical movement and respiratory rate variability.
Electrodermal Activity (EDA)	Requires simple setup; allows non-contact measurement.	External temperature impacts skin conductivity.
Respiration (RESP)	Easy to install and provides insights into stress levels.	Limited range for detecting broader emotional responses.
Blood Volume Pulse (BVP)	Non-invasive; provides data on heart rate and heart rate variability.	Susceptible to noise and motion artifacts; ambient lighting conditions can impact signal quality.

cost-effective and easily captured data. Furthermore, these signals are easy to process and analyze to detect human emotions. One disadvantage of physiological signal acquisition is that data collection requires direct contact with the user's body. The user's consent and preparedness are essential to properly wear the devices or sensors. Table 2.6 shows the summarised strength and limitations of different physiological signals.

2.3.7 Wearable Sensors for Stress and Anxiety Detection

Data gathering from physiological signals using appropriate wearable sensors benefits anxiety detection assessment. Wearable technologies have the potential to provide several advantages, including improved safety, better exercise targeting, greater motivation, and better adherence. It also facilitates an improved progression of exercise programs. However, further research needs to focus on the validity and efficacy of these devices in healthcare, tracking, and therapy. Table 2.7 lists each sensor used for stress and anxiety detection with advantages and limitations.

Table 2.7: Wearable sensors for stress and anxiety detection.

Wearable sensors	Sampling Rate	Advantages	Limitations
Emotiv EPOC Headset [211], [118]	EEG - 128Hz	Non-invasive, Lightweight, provides a user-friendly framework, Cost-effective	EEG readings are easily contaminated with motion artifacts.
MindWave EEG headset [38], [172]	EEG 512Hz	Highly resistant to noise, Low cost.	Single-channel (FP1) device, Low battery life.
Zephyr BioHarness 3 [38], [77], [38]	ECG-250Hz	It can be used in ambulatory outpatient monitoring and long-range transmission.	Low battery lifetime.
CardioSport TP3 [203]	HR	Comfortable and flexible, Low cost, Records the correct RR interval.	It does not have the memory to store data.
BN-PPGED [210]	EDA & PPG-1000Hz	Lightweight and comfortable, it provides high-resolution signal waveforms.	More power-consuming.
Sleep Sense [261], [166]	BR-100Hz	Non-contact and cost-effective, with Higher accuracy.	Pairs with only a single device, No Wi-Fi connection.
CortiWatch [208]	Cortisol	A unique platform for monitoring cortisol levels utilizing human eccrine sweat and a watch-like design.	It does not adapt precisely to the skin's surface, leaving sample collection and analysis gaps.
Muse Headband [24]	EEG-256Hz	Easy to use.	More expensive.
Shimmer sensor [188] [38], [77], [238], [273]	EDA-10Hz, ECG-512Hz, EMG-1024Hz	Good for real-time applications, Low power, and Large storage.	Not waterproof.
Samsung Gear (S1, S2, S3) [42]	PPG-100Hz ST, ACC-100Hz	Easy to use, Long-lasting battery life, Waterproof.	Uncomfortable during physical activity.
LG Watch Urbane 2 [65]	HR-50Hz, ACC-50Hz	Comfortable and easy to use.	Short battery life.
Empatica E4 [137], [218], [207], [44], [119], [110]	BVP-64Hz, EDA-4Hz, ST-4Hz, ACC-32Hz	Non-intrusive, Comfortable, and Easy to wear.	Highly sensitive to motion and motion artifacts.
AutoSense [70], [127], [174]	ECG 64Hz, RESP 21.33Hz, ACC 10.67Hz	Low-power design with a lifespan of more than 10 days while continuously collecting and transferring sensory data.	Data loss/corruption; interference from physical activity.

NOTE: *EEG = Electroencephalography, ECG = Electrocardiography, HR = Heart Rate, EDA = Electrodermal Activity, BR = Breath Rate, PPG = Photoplethysmography, EDA = Galvanic Skin Response, ECG = Electrocardiography, ST = Skin Temperature, ACC = Accelerometer, BVP = Blood Volume Pressure, RESP = Respiration*

2.4 Data Preprocessing

Physiological signals are highly sensitive and can easily become contaminated with artifacts, which are unwanted components in the recorded data that do not originate from the intended source. These artifacts can arise from technical issues or the individual's behavioral and physical activities during data acquisition [83] [201] [242].

1. *Technical factors* contributing to artifacts include power line interference (50/60 Hz), malfunctioning electrodes, broken leads, or electromagnetic noise affecting the recording equipment. Poor electrode placement, excessive conductive gel, or fluctuating impedance can also introduce noise into the signals, making it challenging to extract meaningful information.

2. *Behavioral and physiological activities* of the individual can generate additional noise.

    - Eye blinks, movements, muscle activity (EMG signals), and body movements can significantly distort recordings.
    - Sweating can alter the conductivity of the electrodes, further compromising signal quality.

Artifacts must be managed effectively to ensure the reliability of data analysis. Various artifact removal techniques have been developed to address these challenges. These methods involve filtering and advanced algorithms like Independent Component Analysis (ICA).

2.4.1 Filtering

Filtering is essential for enhancing the quality of physiological signals used in stress and anxiety classification, as it can reduce noise and retain relevant frequency bands critical for effective feature extraction. Distinct stress and anxiety classification studies have employed various filtering methods:

1. **Butterworth Bandpass Filters (BPF):** BPFs are widely used for their smooth frequency response, enabling the isolation of specific frequency ranges

while minimizing distortion. Their adjustable cutoff frequencies make them particularly suitable for physiological signals; for example, a range of 4–45 Hz is commonly applied to EEG signals for stress and anxiety classification [5] [57] [227] [9].

2. **Low-Pass Filters (LPF):** LPFs remove high-frequency noise, making them useful for reducing interference in slower physiological signals like ECG or respiration [248] [109] [121] [108].

3. **High-Pass Filters (HPF):** HPFs eliminate low-frequency noise, such as baseline wander, often found in EEG and ECG signals. Frequencies below 4 Hz are often removed in EEG preprocessing and retain higher-frequency activity for stress and anxiety classification [5] [101].

4. **Moving Average Filters (MAF):** These filters smooth the signal by averaging over a sliding window, effectively reducing random noise. MAFs can be applied with window sizes tailored to the sampling rate for EEG and ECG signals, ensuring computational efficiency in real-time applications [207].

5. **Notch Filters:** These filters are designed to remove specific frequency components, such as power line interference at 50 Hz or 60 Hz. For example, a 50 Hz notch filter is often used for stress and anxiety classification in EEG and ECG signals. These filters effectively eliminate narrow-band noise without distorting the remaining signal [108] [248] [82].

2.4.2 Artifact Removal

Artifact removal is a critical preprocessing step in anxiety classification, especially when dealing with physiological signals prone to contamination by motion, environmental noise, or physiological artifacts like muscle movements or eye blinks [130]. Different techniques are employed to ensure that the signals used for feature extraction and classification are free from distortions that could affect the accuracy of ML models.

1. **Independent Component Analysis (ICA):** Independent Component Analysis (ICA) is a technique for blind source separation commonly used in preprocessing

physiological signals. It separates mixed signals into statistically independent components by applying linear decomposition [16] [29] [122]. ICA operates under specific assumptions:

- The recorded signals are mixtures of spatially stable sources, including neural, physiological, and external artifacts.
- The summation of potentials from various sources, including brain activity, eye movements, and muscle signals, occurs linearly at the sensors.
- The propagation delay from source to sensor is negligible, allowing instantaneous mixing.

2. **Empirical Mode Decomposition (EMD):** Empirical Mode Decomposition (EMD) decomposes non-linear and non-stationary signals, making it effective for processing EEG and ECG data. This adaptive technique breaks a signal into multiple Intrinsic Mode Functions (IMFs), where each IMF represents oscillatory patterns at different time scales. Unlike conventional filtering methods, EMD extracts frequency components without relying on a predefined filter bank, allowing precise separation of brainwave frequencies such as delta, theta, alpha, and beta. Some IMFs may contain artifacts, which researchers can easily discard, while artifact-free IMFs undergo further analysis. EMD extends to multiple channels through Bivariate EMD (two channels), Trivariate EMD (three channels), and Multivariate EMD (up to 32 channels), enabling enhanced signal decomposition for complex data.

2.5 Feature Extraction and Selection

Feature extraction converts raw physiological signals into meaningful numerical representations, which enables effective stress and anxiety classification. This process extracts statistical, temporal, and frequency features from EEG, ECG, EDA, and RESP signals, as discussed in Section 2.3. Further, to enhance feature extraction, signals are segmented into overlapping sliding windows, ensuring temporal continuity and preserving crucial patterns. Figure 2.9 illustrates this process, where each segment

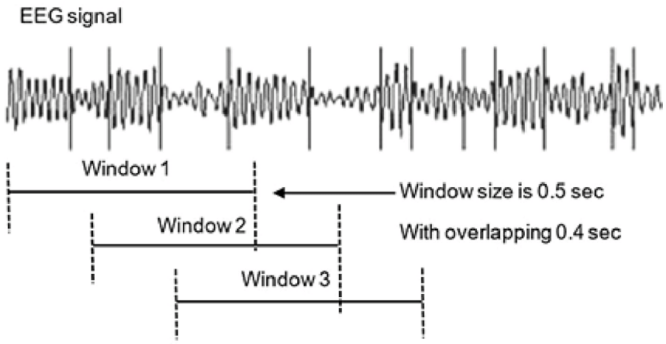


Figure 2.9: Window overlapping on EEG signal [260].

partially overlaps the previous one. This method captures short-term variations in physiological responses while maintaining sufficient data for meaningful analysis. The window size and overlap ratio are carefully selected to balance detecting rapid signal fluctuations and retaining essential temporal characteristics. Applying this approach makes feature extraction more robust, improving the model's ability to identify stress-related patterns effectively. However, the feature extraction process often results in a high-dimensional feature space, which may include irrelevant or redundant features. Retaining all features can lead to the following challenges:

- Increased computational time for analysis and model training
- The risk of overfitting, which may reduce the model's ability to generalize and lower classification accuracy.
- The issue of high-dimensional data, often called the "curse of dimensionality," can degrade model performance.

Therefore, feature selection is crucial for improving model efficiency and effectiveness by retaining only the most relevant features. Further, the feature selection techniques can broadly be categorized into Filter, Wrapper, and Embedded methods. These approaches help to identify the most important features while removing redundant or irrelevant ones.

2.5.1 Filter Method

Filter methods evaluate the relevance of features based on their statistical properties with respect to the target variable, independent of any learning algorithm. These

methods are computationally efficient and often serve as a first step in feature selection [140] [130] [176] [244].

1. *Pearson Correlation Coefficient (PCC)*: Measures the linear relationship between each feature and the target variable. Features with low correlation are removed to reduce redundancy [277][51][140].

2. *Mutual Information (MI)*: Quantifies the mutual dependence between features and the target class and captures non-linear relationships [115][197][93].

3. *Minimum Redundancy Maximum Relevance (mRMR)*: Selects features that are highly relevant to the target but minimally redundant among themselves [52][163].

4. *F-Score and p-Value Tests*: Perform feature ranking based on statistical significance through ANOVA F-test and t-test [100][223][257][175].

5. *ReliefF*: Assigns feature weights based on instance-based learning to identify relevant features [186].

2.5.2 Wrapper Method

Wrapper methods evaluate the performance of different feature subsets by training a model and choosing the subset that yields the highest accuracy. These methods are computationally intensive but often result in better performance as they are tailored to the classifier [109] [24] [165].

1. *Recursive Feature Elimination (RFE)*: Recursively removes the least important features based on the model's weight coefficients to find the optimal subset [165][248].

2. *Genetic Algorithm (GA)*: A population-based optimization technique that evolves feature subsets over generations to maximize classification performance [151].

3. *Particle Swarm Optimization (PSO)*: A swarm intelligence algorithm that searches for optimal feature subsets by simulating the movement of particles in a search space [268].

4. *Whale Optimization Algorithm (WOA)*: Inspired by the hunting behavior of humpback whales, WOA explores and exploits the search space to identify the best feature subset [223].

5. *Sequential Backward Selection (SBS)*: Starts with all features and iteratively removes the least useful features to improve performance [268][109][24][259].

2.5.3 Embedded Method

Embedded methods integrate feature selection into the model training process, leveraging regularization or internal metrics to select features during model building [207][137].

1. *Random Forest (RF)*: Computes feature importance scores based on Gini impurity reduction across decision trees during model training [207].

2. *Bagged Trees (BT)*: Aggregates feature importance across bagged decision trees, reflecting their contribution to overall prediction accuracy [96].

3. *Least Absolute Shrinkage and Selection Operator (LASSO)*: Performs embedded feature selection by penalizing and shrinking coefficients to zero during training [137][180].

2.5.4 Hybrid Method

Hybrid methods integrate multiple feature selection strategies, combining filter, wrapper, or embedded approaches to leverage the strengths of each. This combination enhances feature selection performance by mitigating the limitations inherent in individual techniques [277] [82] [248] [35] [197].

2.6 Summary

This chapter establishes the technical foundation for stress and anxiety classification using physiological signals. It introduces key preprocessing techniques, including artifact removal, filtering, and automated channel selection, essential for enhancing

signal quality. Feature extraction from time and frequency domains is discussed, highlighting its role in identifying stress- and anxiety-related biomarkers. Additionally, various FSAs are explored, emphasizing their importance in reducing redundancy and improving model performance.

Chapter 3

Literature Review of Stress and Anxiety Classification

In recent years, stress and anxiety classification has gained significant attention due to its impact on mental health, cognitive performance, and overall well-being. Self-reported questionnaires provide assessments but often lack reliability because they are subjective and prone to bias. Researchers have studied various physiological signals to improve stress and anxiety classification. These include EEG, EDA, HRV, BVP, and RESP [252]. Each signal captures different responses from the autonomic and central nervous systems. These responses provide valuable insights into an individual's physiological state.

Machine Learning (ML) and Deep Learning (DL) techniques have been widely adopted to analyze these physiological responses [228]. They provide automated and objective methods to classify stress and anxiety states with improved accuracy [85]. However, several challenges, including class imbalance, high-dimensional feature spaces, and model interpretability, still limit the generalizability of classification models [80][29][148]. To address these challenges, automated stress and anxiety classification methodologies typically follow a structured process comprising data preprocessing, feature extraction, feature selection, and classification. ML and DL models are then applied to the extracted features to perform the final classification.

This chapter presents a detailed literature review on stress and anxiety classification. It explores publicly available datasets collected using physiological signals recorded

Table 3.1: Research questions.

Research Questions	Description
RQ1	What are the publicly available datasets used for stress and anxiety classification, and what are their limitations?
RQ2	What Machine Learning and Deep Learning techniques have been explored for stress and anxiety classification using physiological signals?
RQ3	What are the key challenges in developing robust Machine Learning models for stress and anxiety classification?

during stress and anxiety inducing stimuli. It further reviews the application of signal processing and feature selection techniques across stress and anxiety classification. This chapter further analyzes traditional ML and DL approaches, highlighting their significance in physiological signal-based classification. Additionally, it investigates challenges in ML techniques for stress and anxiety classification. Finally, it discusses evaluation metrics, including accuracy, F1-score, AUC_ROC, and statistical analysis, which are utilized in this research for assessing classification performance.

3.1 Research Methodology

This literature review examines ML and DL techniques for stress and anxiety classification using physiological signals. It explores existing research on distinct datasets, classification models, and the challenges associated with physiological signal-based stress and anxiety classification. The review focuses on studies that utilize EEG, HRV, EDA, BVP, and RESP signals, analyzing their role in developing automated classification frameworks.

3.1.1 Research Questions

This research explores the following research questions (RQs), outlined in Table 3.1. These questions define the scope of the survey and establish the keywords for inclusion criteria in selecting relevant studies on stress and anxiety classification. They also aid in identifying key methodologies, datasets, and challenges addressed in existing literature.

Inclusion Criteria: The literature review was conducted by searching relevant

articles across multiple academic databases, including IEEE Digital Library, Springer, Google Scholar, Science Direct, ACM Digital Library, and PubMed, as illustrated in Figure.

Keywords: It uses the preliminary search keywords formulated based on the research focus. ***Anxiety classification, Stress classification, multimodal stress classification, multimodal anxiety classification, anxiety classification using physiological signals, stress classification using physiological signals, physiological signals and stress, physiological signals, and anxiety, anxiety detection via wearable sensors, stress detection via wearable sensors, machine learning in smart healthcare, anxiety classification using machine learning and stress classification using machine learning.***

Exclusion Criteria: This research excludes non-English articles and studies focusing solely on statistically evaluating physiological signals without incorporating classification or predictive modeling.

3.2 Open Datasets Available for Stress and Anxiety

The research particularly focuses on two types of open stress and anxiety datasets, i.e., single-modal and multimodal datasets. Table 3.2 provides a detailed explanation of each dataset and enables a direct comparison between different methodologies, facilitating the benchmarking of stress and anxiety classification techniques. Out of these datasets, this thesis utilizes four datasets: MAT [280], SAM40 [84], DASPS [30] and Spiderphobic [109] as they serve as benchmark datasets for EEG-based stress and multimodal anxiety classification.

3.2.1 Limitations of Existing Datasets

This thesis evaluates several publicly available datasets for stress and anxiety classification and identifies key limitations that impact their applicability.

1. No publicly available dataset integrates multimodal physiological signals such as HRV, EDA, and BVP in an academic environment.

Table 3.2: Open datasets available for stress and anxiety

Dataset	Domain	Year	Modality	Signal	Subjects	Stress Inducer	Channel Selection	Ensemble Stacking	Hybrid FSA	Real-time Academic Environment Assessment
MAT [280]	Stress	2019	Single modal	EEG	36	Mental arithmetic task	×	×	×	Limited
SAM40 [84]	Stress	2022	Single modal	EEG	40	Stroop color-word test, solving arithmetic questions, identification of symmetric mirror images, and a state of relaxation	×	×	×	Limited
CLAS [150]	Stress	2019	Multimodal	ECG, PPG, and EDA	62	Interactive task: 24 mathematical problems, a Stroop test, and 30 logical problems; perspective tasks: video and image	–	×	×	Limited
SRAD [99]	Stress	2005	Multimodal	EMG, ECG, EDA (foot and hand), and RESP	17	Driving in different situations	–	×	×	×
DASPS [30]	Anxiety	2019	Single modal	EEG	23	Face-to-face psychological stimuli	×	×	×	×
Spider-phobic [109]	Anxiety	2020	Multimodal	ECG, EDA, RESP	57	Spider phobic videos	–	×	×	×

2. Datasets such as MAT [280], CLAS [150], SAM40 [84] do not incorporate Montreal Imaging Stress Task (MIST) [58], which induces stress in a gradually increasing manner and is ideal for evaluating the mental workload in an academic environment.

3. Most publicly available multimodal datasets classify stress into binary levels. The datasets such as MAT [109], SAM40 [84], CLAS [150], and Spiderphobic [109] classify stress into binary levels, which restricts the ability to analyze varying degrees of stress.

4. Some datasets, such as SRAD [99], AffDataset [95], and DPPS [66], have relatively small sample sizes, reducing the generalizability for large-scale applications.

5. No datasets explicitly examine the effect of interventions such as meditation and relaxation on physiological responses to stress and anxiety.

3.3 Automated Stress and Anxiety Detection Using Physiological Signals

Machine learning has emerged as a powerful technique for the automated detection of stress and anxiety using physiological signals [29][81]. Researchers leverage various physiological signals to train classification models capable of identifying stress- or

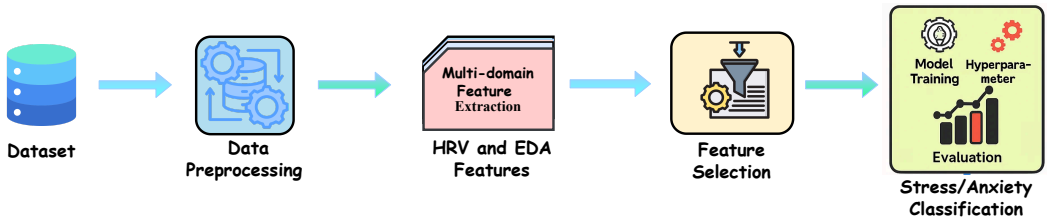


Figure 3.1: Classification Methodology for Stress and Anxiety Classification

anxiety-related patterns in distinct scenarios. However, developing such models requires a systematic methodology that ensures data quality and model reliability. This methodology typically involves different phases, including data preprocessing, feature extraction, feature selection, and model training, as shown in Figure 3.1. Each step directly influences the overall performance of the classification system.

Chapter 2 presents a detailed overview of commonly used techniques for data preprocessing and feature engineering in the context of stress and anxiety classification. The subsequent subsections describe each phase of this methodology and review related work that implement these techniques to physiological signal-based stress and anxiety classification.

3.3.1 Data Preprocessing

Data preprocessing plays a critical role in preparing physiological signals for reliable analysis. Various preprocessing techniques utilised in the studies are detailed in Section 2.3. This section focuses on reviewing the preprocessing strategies adopted in previous research.

3.3.1.1 Related work of Data Preprocessing on Stress and Anxiety

Several studies emphasize the importance of filtering techniques to eliminate noise and unwanted frequency components from raw physiological signals. Seo et al. [221] applied a bandpass filter from 1.5 to 150 Hz for ECG and a low-pass filter with a cutoff of 0.5 Hz for RESP signals for stress classification. Similarly, Nath et al. [176] used a low-pass Butterworth filter of a 10 Hz cutoff for BVP signals and a 1 Hz cutoff for EDA to reduce motion artifacts while preserving essential physiological information.

In anxiety studies, Vaz et al. [248] applied a 4th-order Butterworth bandpass filter

for ECG, EDA, and EMG signals. Similarly, Grigoras et al. [251] employed 0.5-12 Hz low-pass filters for ECG signals to maintain heart rate variability information, which is essential for detecting anxiety states. For the EDA signal, Vaz et al. [248] applied a 5 Hz low-pass Butterworth filter to focus on the tonic component of the EDA signal, effectively removing high-frequency noise and capturing slow-varying physiological responses associated with anxiety. Daneshmand et al. [52] employed an FIR pass-band filter (4-45 Hz) to preprocess EEG data.

In addition to ICA, Artifact Subspace Reconstruction (ASR) is another robust technique that removes transient noise from EEG signals. Lee et al. [137] utilized ASR to remove motion and environmental artifacts during the classification of driver anxiety, which improved the quality of EEG data and led to more reliable anxiety classification results. ASR is effective in real-time systems, where continuous signal quality control is essential for accurately detecting emotional states. Li et al. [140] effectively utilized ICA in their anxiety classification study. ICA is particularly used for EEG signals because it can efficiently separate neural activity from non-invasive noise.

For signals such as EDA, ECG, and PPG, where artifacts typically originate from motion and environmental factors, filtering techniques like bandpass filtering are often paired with artifact removal processes. Vaz et al. [248] applied a combination of Butterworth filters and Notch filters to handle artifacts in ECG and EMG signals, resulting in better signal quality for classification tasks. Puli et al. [188] demonstrated that combining multimodal Kalman filters with bandpass filtering significantly reduced motion artifacts, especially during real-time classification of anxiety in dynamic environments. This technique is particularly useful in wearable devices, where motion is inevitable and can heavily distort the signal quality.

3.3.2 Feature Extraction and Selection

After preprocessing the physiological signals, the feature extraction process captures discriminative patterns associated with stress and anxiety. Section 2.2 describes the detailed feature extraction from each physiological signal. Applying feature selection becomes essential as this process often generates a high-dimensional feature space that may contain irrelevant or redundant features. This step reduces dimensionality,

improves classification performance, and enhances model interpretability. Section 2.4 discusses various feature selection algorithms, such as filter, wrapper, embedded, and hybrid methods, in detail to refine physiological signal data and optimize the performance of stress and anxiety classification models. This section presents related work that implements these algorithms for physiological signal-based stress and anxiety classification.

3.3.2.1 Related Work of Feature Selection in Stress Classification

Studies applying filter methods have effectively ranked features based on statistical significance. Hemakom et al. [100] used ANOVA F-value selection to refine ECG and EEG features, enabling an SVM classifier to achieve 87.50% accuracy. Sharma et al. [223] applied F-score selection on EEG entropy-based features, enhancing classification performance with RF, SVM, and evolutionary algorithms. Wen et al. [257] employed P-value selection to refine EEG-based FD and Time-Frequency Domain features, increasing SVM accuracy to 99%. Jebelli et al. [115] applied a Correlation-Based Wrapper Method, identifying EEG features correlated with salivary cortisol levels, which led to 80.32% accuracy with Gaussian SVM. These studies highlight the importance of statistical ranking techniques in stress classification, ensuring that the most relevant biomarkers contribute to classification models.

Wrapper-based approaches have further refined feature selection in stress classification. Mustafa et al. [169] applied a wrapper-based selection method, reducing EEG feature dimensionality and improving classification accuracy across Naïve Bayes, SVM, and MLP classifiers (91.52%). Saputra et al. [192] used K-best selection to reduce 900 EEG features, leading to 95.36% accuracy with SVM. Xiong et al. [268] employed SBS and PSO to optimize feature selection for EEG and ECG, improving classification performance across Quadratic SVM, KNN, and Decision Tree models. These findings show that wrapper techniques efficiently select the most discriminative stress-related features, improving model robustness.

Several studies have integrated embedded feature selection within model training to improve computational efficiency. Attallah et al. [26] applied PCA-based dimensionality reduction, demonstrating that a minimal number of frontal EEG

electrodes can achieve high-accuracy stress classification (Cubic SVM: 99.91%, KNN: 99.98%). Halim et al. [93] combined PCA, MI, and Random Subset Feature Selection, achieving 97.95% accuracy with an ensembled model. Doma et al. [63] applied PCA for EEG-based stress classification, identifying frequency-domain features as crucial for stress classification. Furthermore, hybrid feature selection methods, which combine filter, wrapper, and optimization-based techniques, have further enhanced stress classification models. Hasan et al. [96] combined Boruta selection with PCA, increasing KNN classification accuracy to 73.38%.

3.3.2.2 Related Work of Feature Selection in Anxiety Classification

Similar to stress classification, researchers have explored various FSA techniques for anxiety classification. Studies using filter methods have efficiently ranked anxiety-related EEG features. Li et al. [140] extracted 90 EEG features, selecting 30 using correlation coefficients, which resulted in 62.56% accuracy with SVM. Klados et al. [130] applied CfsSubsetEval on 466 EEG features, improving Naïve Bayes classification accuracy to 93.75%. Nath et al. [176] employed P-value selection to refine 24 features from wristband sensor data, achieving 92% accuracy with Random Forest. Tripathy et al. [244] applied ANOVA to ECG signals, selecting 187 out of 200 features, leading to 79.17% accuracy with a Decision Tree. Daneshmand et al. [52] implemented mRMR for EEG-based anxiety classification, reducing 28 features to 15 and obtaining 100% accuracy with KNN. These studies demonstrate that filter-based selection methods effectively rank anxiety-related features, allowing classifiers to distinguish between anxiety levels more accurately.

Wrapper methods have also shown promise in anxiety classification. Ihmig et al. [109] applied Sequential Forward Selection (SFS) to retain six out of 25 features from ECG, RESP, and EDA, achieving 89.8% accuracy with Bagged Trees. Arsalan et al. [24] employed forward selection, selecting five out of ten EEG features, leading to 76.92% accuracy with MLP. Muhammad et al. [165] reduced 60 EEG features to 18 using RFE, improving classification accuracy to 94.9% with Random Forest. Wen et al. [259] applied backward selection, refining 11 heart rate features to four, which improved SVM classification accuracy to 90.5%.

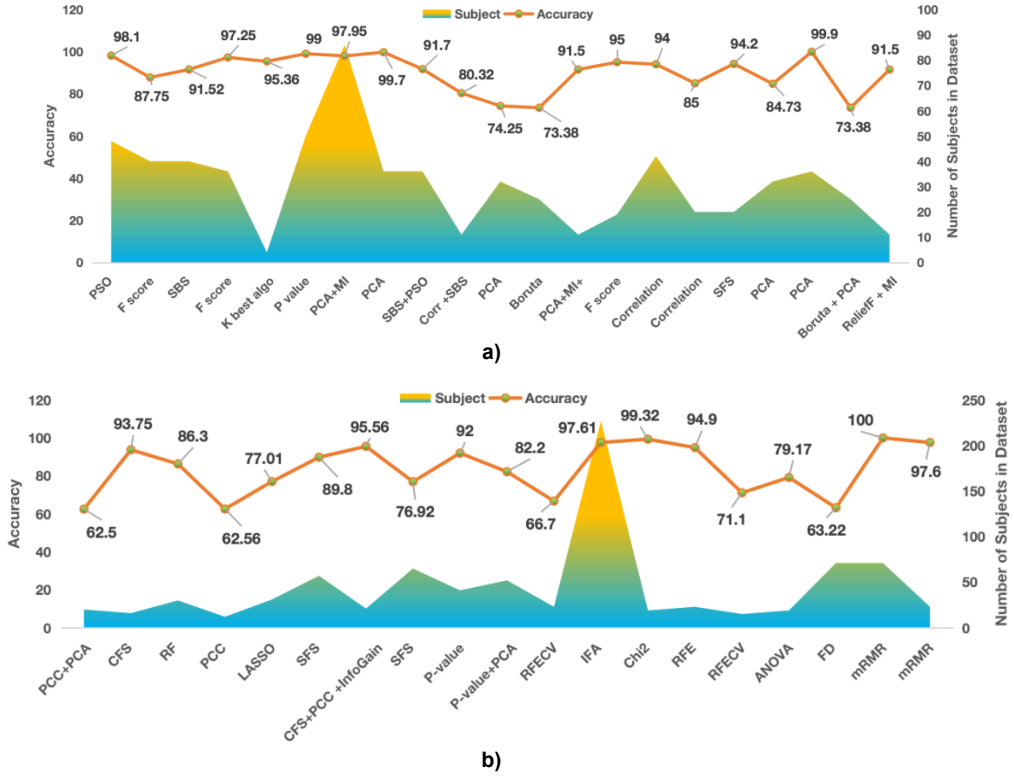


Figure 3.2: Accuracy of different feature selection algorithms for a) stress and b) anxiety classification, along with the number of subjects in each dataset.

A metaheuristic feature selection approach has also been introduced for anxiety classification. Mo et al. [162] implemented the Improved Fireworks Algorithm for multimodal feature selection (heart rate, respiration rate, behavioral, audio, text, and questionnaire data), selecting 19 features and achieving 97.61% accuracy with Adaptive Boosting (ADB).

For embedded methods, Salkevicius et al. [207] applied Random Forest-based embedded selection, reducing 33 multimodal features to ten, achieving 86.3% accuracy with SVM. Lee et al. [137] used LASSO feature selection for EEG, PPG, EDA, and pupil size data, selecting 31 out of 445 features, achieving 77.01% accuracy with Logistic Regression.

Hybrid approaches have also demonstrated effectiveness in anxiety classification. Zheng et al. [277] applied Pearson Correlation and PCA for PPG and EEG feature reduction, achieving 62.5% accuracy with KNN. George et al. combined P-value selection and PCA, reducing 23 features to 11, reaching 82.2% accuracy with SVM.

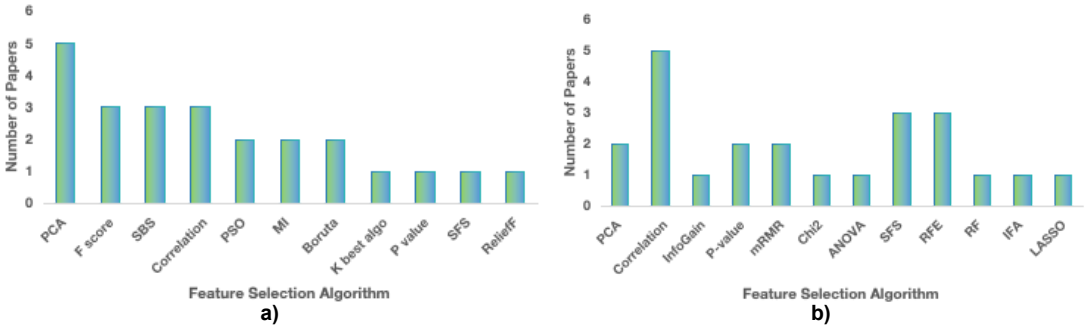


Figure 3.3: Frequency of feature selection algorithms utilized in a) stress and b) anxiety classification studies.

Baygin et al. [35] integrated Neighborhood Component Analysis (NCA) and Chi-square tests, improving KNN classification accuracy to 99.32% using ECG. Vaz et al. [248] employed unsupervised FSA and RFECV, reducing 109 ECG, EDA, and EMG features to 34, achieving 71.1% accuracy with Logistic Regression. Lastly, Rodríguez-Arce et al.[197] focused on multimodal physiological signals in an academic environment, applying CfsSubsetEval, CorrelationAttributeEval, and InfoGainAttributeEval, selecting 13 out of 21 features, which led to 95.56% accuracy with KNN.

To further illustrate the impact of feature selection algorithms on stress and anxiety classification, Figures 3.2 and 3.3 present a comparative analysis of different FSAs used in existing studies. Figure 3.2 highlights the classification accuracy achieved by various FSAs and the number of subjects in each dataset, providing insights into their effectiveness in optimizing feature subsets. Further, Figure 3.3 visualizes the frequency of feature selection techniques employed in stress and anxiety classification research, highlighting the most commonly utilized methods.

3.3.3 Machine Learning Techniques for Stress Classification

Machine Learning (ML) techniques enable the identification of temporal patterns in physiological signals to classify stress levels. The training phase ensures accurate classification and requires extensive and diverse datasets to develop robust models. This process involves mapping selected features to corresponding stress levels. High-quality training data enhances model generalization, allowing its application across

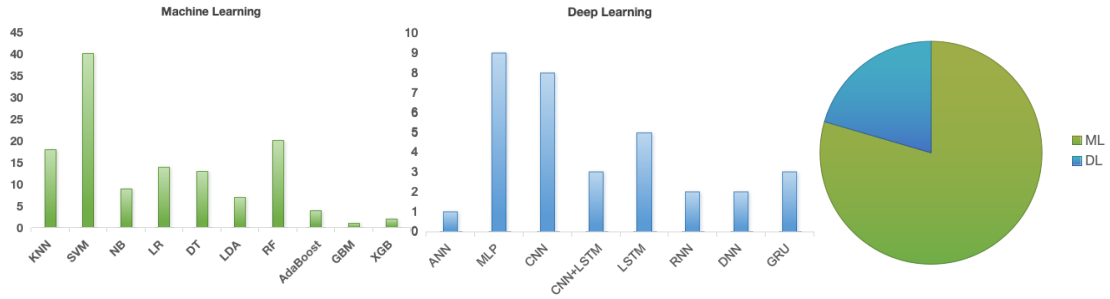


Figure 3.4: Distribution of traditional ML and DL algorithms for stress classification. (a) ML algorithm utilization. (b) DL algorithm utilization. (c) Comparison of ML vs. DL utilization.

diverse datasets and experimental conditions.

Although researchers widely use ML for stress classification, physiological signal-based approaches face several challenges, including class imbalance, feature redundancy, and model complexity, which affect classification performance. To improve classification accuracy, researchers integrate techniques such as data balancing methods, feature selection, and model optimization within supervised learning frameworks to enhance stress level differentiation. Figure 3.4 presents a comparative analysis of ML and DL algorithms utilized for stress classification. Figure 3.4(a) illustrates the Distribution of ML algorithms, highlighting the prevalence of Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN) in stress classification tasks. Figure 3.4(b) depicts the Distribution of DL models, emphasizing the widespread adoption of Convolution Neural Network (CNN), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN) models in handling physiological signals. Lastly, Figure 3.4(c) compares ML and DL utilization, indicating that ML-based approaches remain more widely employed than DL techniques in stress classification research. Observations show that SVM consistently demonstrates high accuracy among all models, making it one of the most utilized classifiers for distinguishing distinct stress levels. Table 3.3 summarises the ML studies for physiological signals-based stress classification.

Table 3.3: Overview of stress classification studies using traditional ML.

Author	Task	Preprocessing	Dataset	Signal	Subjects	Features	Algorithms	Results
Badr et al.[29]	SCWT	ICA, PSD	Private	EEG	22	FD	KNN, LDA, SVM, NB, DT	KNN: Delta 99.46%, SVM: Alpha 99.98%, SVM: Beta 99.96%, SVM: Theta 99.85%
Marthinsen et al.[151]	Stroop, Arithmetic, Mirror	–	Public	EEG	40	FD	SVM, CNN	CNN: 84.18%, SVM: 87.50%, SVM with wavelet scattering: 87.50%
Suryawanshi et al.[239]	Cognitive Task	FFT	Private	EEG	–	FD	SVM, KNN, DT, ANN	SVM: 91%
Hemakom et al.[100]	Mental arithmetic task	EEG: LPF 200Hz, HPF 1Hz; ECG: HPF 1Hz	Private	EEG + ECG	40	FD	NB, LR, ADB, KNN, RF, SVM, RBF-SVM	No stress vs Low stress: SVM 86.13%, No stress vs High stress: SVM 87.75%, No stress vs (low & high) Stress: SVM 87.50%
Mustafa et al.[169]	Questionnaire + PSS	90% overlap window	Private	EEG	40	FD	NB, SVM, LR, RF, MLP, ADB	2-class: 91.52%, 3-class: 88.47%, 4-class: 87.36%
Sharma et al.[223]	Mental arithmetic task	Notch filter: 50Hz, LPF and HPF	Public	EEG	36	Entropy-based features	RF, J48, NB, KNN, SVM	4s: 94.02%, 8s: 97.26%
Halim et al.[93]	Driving Videos	BPF, removed 0-4Hz	Private	EEG	86	TD and FD	SVM, RF, NN	RF: 87.23%, NN: 89.00%, SVM: 91.60%, Ensembled: 97.95%
Attallah et al.[26]	Mental arithmetic task	Butterworth IIR Filter, Wavelet decomposition, PCA	Public	EEG	36	TD and FD	RF, KNN, SVM, LDA	LDA: 98.6%, Linear SVM: 94.5%, Cubic SVM: 99.7%, KNN: 99%, RF: 98.91%
Ahammed et al.[3]	Mental arithmetic task	HPF: 0.5 Hz, LPF: 45 Hz, Notch: 50 Hz	Public	EEG	36	Multivariate Sample Entropy	SVM + Statistical Analysis	SVM: 90%
Xiong et al.[268]	Mental arithmetic task	HPF: 0.5 Hz, LPF: 45 Hz, ICA	Public	EEG, ECG	36	TD, and FD	SVM, KNN, DT	SVM (quadratic): 80.6%, KNN 91.7%, DT 86.1%
Garg et al.[79]	TSST	10s sliding window	Public	BVP, ECG, EDA, EMG, BT, RP, Acceleration	15	TD	RF, SVM, KNN, LDA, ADB	NB: 60%, KNN: 80%, SVM: 81%, MLP: 83%, RF: 85%, GBM: 85%
Dalmeida et al.[51]	Driving Task	Minmax, replaces missing values	Private	ECG, HRV	16	TD and FD	NB, KNN, SVM, MLP, RF, GBM	NB: 60%, KNN: 80%, SVM: 81%, MLP: 83%, RF: 85%, GBM: 85%
Deng et al.[60]	Driving Task	5-minute segments	Private	ECG, EMG, EDA	10	FD	ANN, SVM, NB, KNN	SVM: 78.46%, NB: 75.38%, KNN: 72.31%
Pankajavalli et al.[180]	Driving Task	Not Mentioned	Private	ECG, EMG, EDA, RR	–	TD, FD and Statistical	SVM, NB, DT, LR, KNN	SVM: 100%

NOTE: *LDA: Linear Discriminant Analysis, RBF: Radial Basis Function; TD: Time Domain; FD: Frequency Domain*

3.3.3.1 Related Work of Traditional ML Models on Stress Classification

Several studies demonstrate the effectiveness of traditional ML techniques for stress classification using physiological signals. Among traditional ML models, SVM is one of the most widely used classifiers, consistently achieving high accuracy across multiple studies. Badr et al. [29] extracted PSD-based Frequency Domain (FD) features from EEG signals and applied SVM to classify stress, achieving an accuracy of 99.98% in the alpha band and 99.96% in the beta band using SVM. Similarly, Marthinsen et al. [151] applied wavelet scattering-based features from EEG signals and observed that SVM achieves an accuracy of 87.50%, outperforming the CNN model with an accuracy of 84.18%. Hemakom et al. [100] combined EEG and ECG signals and employed Naïve Bayes (NB), Logistic Regression (LR), Adaptive Boosting (ADB), KNN, RF, SVM, and RBF-SVM. The authors achieved the highest accuracy of 87.75% in distinguishing high-stress levels using SVM. Additionally, Attallah et al. [26] utilized feature selection techniques along with traditional classifiers and achieved the highest accuracy of 99.7% using SVM.

Furthermore, ensemble-based methods demonstrate robust classification accuracy in stress classification. Dalmeida et al. [51] utilized ECG-derived HRV features and observed the highest accuracy of 85% using RF and Gradient Boosting Machine (GBM), respectively. Similarly, Garg et al. [79] examined multimodal signals, including BVP, ECG, EDA, EMG, BT, RP, and Acceleration, and observed the highest accuracy of 85% using RF and GBM, outperforming other classifiers such as KNN with an accuracy of 80%, SVM with an accuracy of 81% and MLP with an accuracy of 83%.

Moreover, several studies explored the integration of multimodal physiological signals to enhance classification performance. Xiong et al. [268] utilized EEG and ECG features, achieving an accuracy of 91.7%, using KNN. Additionally, Pankajavalli et al. [180] utilized SVM along with other traditional classifiers for multimodal physiological signal-based stress classification, achieving 100% accuracy. These results demonstrated that multimodal approaches, combined with traditional ML classifiers, significantly improve stress classification by capturing diverse autonomic responses. Mustafa et al. [169] employed NB, SVM, LR, RF, MLP, and ADB for EEG-based stress classification and achieved an accuracy of 91.52%, 88.47%, and 87.36% for binary, three-class, and

four-class classification, respectively.

3.3.3.2 Related Work of DL Models on Stress Classification

Deep learning (DL) approaches have been extensively applied in stress classification, utilizing various physiological signals such as EEG, ECG, EDA, PPG, and respiration. Researchers have demonstrated the effectiveness of CNN-based models in stress classification. Gil-Martin et al. [88] employed CNN on multimodal signals such as ECG, EMG, EDA, respiration, and skin temperature from the WESAD dataset [216] and achieved an accuracy of 96.62% for two-level classification and 85.03% for three-level classification. Seo et al. [221] applied Deep Neural Networks (DNN) on multimodal signals, including ECG, RESP, and facial images, and achieved an accuracy of 73.30%.

Among EEG-based studies, Bhatnagar et al. [39] utilized an EEGNet-based CNN with ReLU activation for stress classification based on academic and social stressors, achieving an accuracy of 99.45%. Malviya et al. [148] developed a CNN-BiLSTM hybrid model and employed an EEG signal from the MAT dataset [280]. The authors achieve the highest accuracy of 99.20% using CNN-BiLSTM.

Furthermore, researchers have also widely explored recurrent models such as LSTM and BiLSTM. Varshney et al. [247] utilized LSTM, BiLSTM, and Gated Recurrent Unit (GRU) models for EEG-based stress classification. The authors achieved an accuracy of 99.81%, 99.81%, and 99.43% using GRU, LSTM, and BiLSTM, respectively. Similarly, Roy et al. [199] evaluated CNN-RNN (Recurrent Neural Network), CNN-LSTM, CNN-GRU, and BiLSTM models on EEG data. The authors achieved the highest accuracy of 97.10% and 95.60% using BiLSTM and CNN-LSTM, respectively.

Further, Multimodal approaches integrating multiple physiological signals have gained good performance in recent studies. Salankar et al. [205] combined EEG and ECG signals and achieved the highest accuracy of 100% using MLP. Huang et al. [108] explored different combinations of CNN, LSTM, and ConvLSTM for multimodal stress classification, where CNN-LSTM achieved the highest accuracy of 97.8%. Nath et al. [176] employed CNN on multimodal signals from the WESAD dataset [216]. The authors achieved the highest accuracy of 87% using CNN for stress classification. Table

Table 3.4: Overview of stress classification studies using DL.

Study	Task	Preprocessing	Dataset	Signal Modality	Subjects	Channels	Feature Extraction	Algorithms	Results (% Accuracy)
Roy et al.[199]	A commercial psychological test	HPF: 1Hz	Public	EEG	48	14	FD, Time-Frequency Domain	CNN-RNN, CNN-LSTM, CNN-GRU, BiLSTM	CNN-RNN: 89.91%, CNN-LSTM: 95.60%, CNN-GRU: 95.20%, BiLSTM: 97.10%
Bhatnagar et al.[39]	Academic and social stress questionnaire	–	Private	EEG	45	23	FD	CNN (ReLU activation)	CNN: 99.45%
Hafeez et al.[91]	Mental arithmetic task	EEG Lab processing	Private	EEG	14	10	FD	LSTM, ResNet, CNN	Not reported
Salankar et al.[206]	Mental arithmetic task	EMD+VMD	Public	EEG	36	19	Statistical	SVM, MLPNN	MLPNN: 99.99%
Malviya et al.[148]	Mental arithmetic task	DWT	Public	EEG	36	19	FD	KNN, CNN, CNN-BiLSTM	KNN: 96.96%, CNN-BiLSTM: 99.20%
Al-Saggaf et al.[10]	MIST	BPF: 4–30 Hz	Private	EEG	–	128	TD and FD	CNN, DT, LR, SVM	CNN: 96.00%
Salankar et al.[205]	Mental arithmetic task	VMD	Public	EEG+ECG	36	19 EEG + 1 ECG	Statistical	SVM, MLPNN	MLP: 100% (temporal), SVM: 78% (temporal), SVM: 70.67% (frontal)
Kamińska et al.[122]	Stroop + VR relaxation	FIR, ICA, Time-Freq conversion	Private	EEG	28	32	FD	CNN, SVM, MLPNN	KNN: 91.7% (Theta), RF: 92.86% (Theta), SVM: 96.42% (Combo), MLP: 96.42% (Combo), CNN: 87.5%
Varshney et al.[247]	Mental arithmetic task	2s sliding window	Public	EEG	36	23	Entropy-based features	LSTM, BiLSTM, GRU	GRU: 99.81%, LSTM: 99.81%, BiLSTM: 99.43%
Salankar et al.[204]	Mental arithmetic task	VMD	Public	EEG+ECG	36	23	Poincare plot features	MLP, SVM	MLP: 100% (Good Performer), SVM: 78% (Bad Performer)
Arsalan et al.[23]	TSST	Onboard DRL feedback circuit	Private	EEG	28	4	FD, DASM, RASM	SVM, NB, MLP	MLP: 92.85% (2-class), 64.28% (3-class)
Kalra et al.[120]	Stroop Color-Word Test	LPF 5 Hz	Private	PPG	15	–	TD, FD	MLP, DNN, LR, SVM	DNN: 92.00%
Seo et al.[221]	Stroop Color-Word Test	10s window, BP + Notch filters	Private	TD and FD	24	–	ECG, RESP, Sequence facial features	DNN	73.30%
Gil-Martin et al.[88]	Trier Social Stress Test (TSST)	–	Public	ECG, EMG, EDA, RESP, ST	15	–	–	CNN	CNN: 96.62% (2 levels), 85.03% (3 levels)
Huang et al.[108]	Driving	EEG: Notch:50 Hz, DWT, EMD	Private	EEG, ECG, EDA, RESP	18	2	–	CNN, LSTM, ConvLSTM, XGB	CNN+LSTM: 97.80%, CNN: 96.5%, ConvLSTM: 88.0%, XGB: 74.4%
Nath et al.[176]	Trier Social Stress Test (TSST)	13 CNN filters	Public	BVP, ECG, EDA, EMG, BT, RP, Acceleration	15	–	–	CNN	CNN: 87.00%

NOTE: *TD: Time Domain; FD: Frequency Domain*

3.4 summarises the DL studies for physiological signals-based stress classification.

3.3.4 Machine Learning Techniques for Anxiety Classification

Machine Learning (ML) techniques play a significant role in anxiety classification by enabling models to differentiate between varying anxiety levels based on physiological signals. Researchers have widely adopted supervised learning approaches to train models using features extracted from EEG, ECG, EDA, and respiration signals. Traditional classifiers such as SVM, KNN, and RF remain prevalent due to their interpretability and effectiveness, as shown in Figure 3.5(a). Figure 3.5(b) presents the

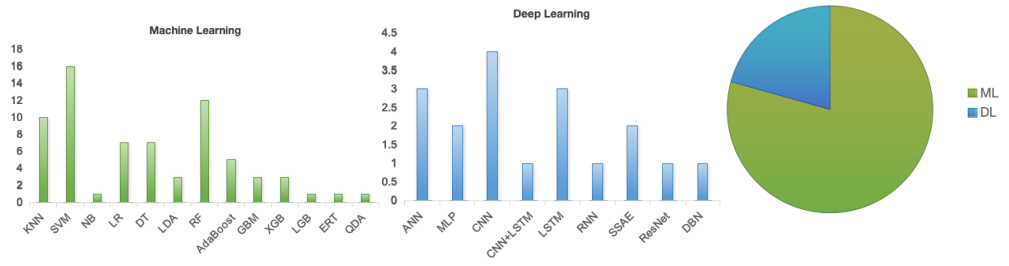


Figure 3.5: Distribution of traditional ML and DL algorithms for anxiety classification. (a) ML algorithm utilization. (b) DL algorithm utilization. (c) Comparison of ML vs. DL utilization.

DL techniques used for anxiety classification, with CNN, LSTM, and Artificial Neural Network (ANN) emerging as the most frequently utilized models. Lastly, Figure 3.5(c) compares the overall utilization of ML and DL approaches in anxiety classification, demonstrating that ML techniques remain more prevalent than DL-based methods. Additionally, recent studies have integrated feature selection techniques to enhance model performance by reducing redundancy and improving generalization.

Furthermore, multimodal approaches combining multiple physiological signals have improved classification accuracy by capturing diverse anxiety-related biomarkers. The subsequent sections explore traditional ML and DL methodologies for anxiety classification, providing insights into their effectiveness in identifying anxiety states based on physiological responses.

3.3.4.1 Related work of Traditional ML Models on Anxiety Classification

Several studies explore traditional ML techniques for anxiety classification using various physiological signals. Among the most investigated signals, EEG-based studies show that the choice of features and preprocessing techniques significantly impact classification performance. Klados et al. [130] apply NB, KNN, and SVM on EEG signals processed using ICA and bandpass filters. The results indicate that NB achieves the highest accuracy of 93.75% and outperforms KNN and SVM.

Multimodal approaches integrating multiple physiological signals further strengthen anxiety classification. Rodríguez-Arce et al. [197] demonstrated the advantages of combining EDA, HR, BR, ST, and SPO2 signals by applying KNN, SVM, Decision Tree (DT), and Logistic Regression (LR). The study achieved accuracies of 95.56%,

94.44%, 86.88%, and 87.59% using KNN, RF, SVM, and LR, respectively. Similarly, Nath et al. [175] analyzed EDA and BVP signals for 41 subjects. RF achieved the highest accuracy of 92%, significantly outperforming both SVM and LR.

ECG-derived HRV also proves highly effective in detecting anxiety-related responses. Vaz et al. [248] explored ECG, EDA, and EMG features using various classifiers and achieved the highest accuracy of 92% using ADB, 91.6% using Extreme Gradient Boosting (XGB), and 89.8% using RF. Similarly, George et al. [82] applied SVM, RF, and GBM to ECG signals, with SVM reaching an accuracy of 82.2%. Furthermore, Tripathy et al. [244] examined ECG signals and achieved an accuracy of 92.27% with XGB. Results show that tree-based ensemble methods offer superior classification performance.

Additionally, researchers have also explored HR and PPG features for anxiety classification. Puli et al. [188] investigated HR and Accelerometer (ACC) signals. The authors showed that a Kalman-filter-based approach significantly improves classification accuracy to 93%, which outperforms conventional classifiers such as KNN, SVM, RF, and ADB. Additionally, Lee et al. [137] employed a combination of EEG, PPG, and EDA signals. The authors achieved the highest accuracy of 77.01% using LR. The results show that EEG is an important physiological signal; however, integrating peripheral physiological signals further enhances classification robustness. Studies by Ihmig et al. [109] and Gazi et al. [80] utilized ECG, RESP, and EDA signals to classify anxiety responses. The results showed that tree-based classifiers effectively capture autonomic changes. Ihmig et al. [109] achieved an accuracy of 89.80% using Bagged Trees, whereas Gazi et al. [80] reported 88% accuracy using RF.

Traditional ML models consistently demonstrate strong performance in anxiety classification. SVM, KNN, and RF remain widely used due to their robustness in handling physiological data, while multimodal approaches incorporating EEG, ECG, EDA, and HRV improve classification accuracy. Table 3.5 summarises the ML studies for physiological signals-based anxiety classification.

Table 3.5: Overview of anxiety classification studies using traditional ML.

Author	Stimuli	Preprocessing	Dataset (Public/Private)	Signal	Subjects	Channels	Features	Algorithms	Accuracy
Zheng et al.[277]	Ride stationary bike on high speed during competition	PPG: 3rd order BPF, MAF	Private	EEG, PPG	20	1	TD and FD	PCA-KNN, PCA-SVM	KNN: 62.5%, SVM: 50%
Klados et al.[130]	Mental arithmetic task	LP: 0.5-45Hz, BPF: 47-53Hz, ICA	Private	EEG	16	57	–	NB, KNN, SVM	NB: 93.75%, KNN: 81.4%, SVM: 77.4%
Salkevicius et al.[207]	Public speaking	MAF	Private	BVP, EDA, ST	30	–	Statistical, Differential, Phasic, HR	SVM: RBF	SVM: 86.30%
Puli et al.[188]	SCWT	Multimodal Kalman filter, BPF:0.25-5Hz, resampled to 5Hz	Private	HR, ACC	15	–	TD	KNN, LR, SVM, DT, RF, ADB	KNN: 71%, SVM: 70%, LR: 67%, DT: 65%, RF: 71%, ADB: 71%, Proposed Kalman Filter: 93%
Sebastiao et al.[218]	Public speaking	–	Private	EDA, HR	15	–	TD, Statistical	100 Bagged Trees	95.70%
Lee et al.[137]	Anxiety inducing videos	STFT with 1-s window, 50% overlapping, PSD	Private	FD	31	31	EEG: FD, Statistical features for other signals	LR	77.01%
Rodríguez-Arce et al.[197]	Mental arithmetic task	LPF	Private	ST, EDA, HR, BR, SPO2	21	–	Statistical	KNN, SVM, DT, LR	KNN: 95.56%, SVM: 86.88%, LR: 87.59%, RF: 94.44%
Nath et al.[175]	TSST	Normalization, BVP: LPF: 1Hz, EDA: 10Hz	Private	EDA, BVP	41	–	Statistical and context-based features	RF, LR, SVM	RF: 92%, LR: 61%, SVM: 61%
George et al.[82]	Sleep study	LPF: 0.5-0.6 Hz, Notch: 50Hz, FIR	Private	ECG	52	–	Statistical and FD	SVM, RF, GBM	SVM: 82.2%, RF: 81.64%, GBM: 78%
Vaz et al.[248]	TSST	ECG: BPF, EDA: LPF: 5Hz, EMG: Notch: 50Hz	Public	ECG, EDA, EMG	15	–	TD, FD, and non-linear, phasic and tonic	LR, LDA, DT, SVM, ADB, RF, XGB	LR: 71.1%, LDA: 67.1%, DT: 75.6%, SVM: 70.2%, ADB: 92%, RF: 89.8%, XGB: 91.6%
Tripathy et al.[244]	Anxiety inducing video clips	Fourier–Bessel Domain Adaptive Wavelet Transform	Public	ECG	19	–	TD, FD	DT, Light GBM, XGB, RF, ERT	DT: 79.17%, Light GBM: 90.46%, XGB: 92.27%, RF: 88.91%, ERT: 91.92%
Daneshmand et al.[52]	Face-to-face stimulation, HAM-A	FIR: 4-45Hz	Public	EEG	23	–	Novel Chaotic map based features	KNN, DT	KNN: 100%, DT: 78%
Mo et al.[162]	Questionnaire, facial videos, audio	–	Private	PPG, behavioral, audio	227	–	FD	ADB	ADB: 97.61%
Baygin et al.[35]	Watch anxiety inducing video clips	–	Public	ECG	19	–	Probabilistic Binary Pattern	KNN+NCA, SVM+Chi2	KNN+NCA: 98.81%, SVM+Chi2: 99.94%
Wen et al.[259]	TSST	–	Private	ECG	65	–	TD and FD	SVM, NB, KNN, LDA, QDA	SVM: 90.5%, NB: 86.64%, KNN: 87.39%, LDA: 86.34%, QDA: 84.76%
Li et al.[140]	Public speech	BPF: 0.5-50Hz, ICA	Private	EEG	12	128	TD and FD	SVM	SVM: 62.56%
Jang et al.[114]	Cognitive/perceptual tasks	–	Private	ECG, EDA, RESP, Temperature	71	–	FD	LR, KNN, SVM, RF, MLP	LR: 66.30%, KNN: 60.15%, SVM: 64.66%, RF: 64.15%, MLP: 75.16%
Ihmig et al.[109]	Short spiderphobic videos	ECG: BPF: 5–12 Hz, RESP: 0.1–24 Hz, EDA: LPF: 1.5 Hz	Public	ECG, EDA, BR	57	–	FD	Bagged Trees	89.80%
Gazi et al.[80]	short spiderphobic videos	Resampled to 1Hz using FIR anti-aliasing filter, smoothed with 5-s MAF	Public	ECG, RESP, EDA	57	–	FD, Phasic and Tonic	RF	88%

NOTE: *PCA: Principle Component Analysis; TD: Time Domain; FD: Frequency Domain*

3.3.4.2 Related work of DL Models on Anxiety Classification

Deep Learning (DL) techniques have gained prominence in anxiety classification by leveraging their ability to automatically extract meaningful patterns from physiological signals. Various studies apply CNN, LSTM, and hybrid architectures to enhance classification accuracy. However, while DL models demonstrate strong performance, traditional ML models often remain competitive, particularly with multimodal physiological signals.

Several studies explore EEG-based DL approaches to classify anxiety. Mokatren et al. [164] employed CNN, SVM, and KNN on EEG recordings of Social Anxiety Disorder patients. CNN achieved the highest accuracy of 87%, outperforming SVM with an accuracy of 79% and KNN with an accuracy of 75%. In one study, Liu et al. [143] employed MSTCNN with a squeeze-and-excitation (SE) attention mechanism and CNN+LSTM. The authors achieved the highest accuracy of 99.47% using MSTCNN-SE. Similarly, Baghdadi et al. [30] extracted time and frequency domain features from EEG signals and implemented multiple classifiers, including SVM, KNN, and a Stacked Sparse Autoencoder (SSAE). The SSAE outperformed traditional ML models, achieving an accuracy of 83.35%, reinforcing the effectiveness of feature learning in enhancing classification performance. The results showed that the attention mechanism enhances feature learning by selectively focusing on critical EEG regions, leading to significant performance improvements. Similarly, Ezzi et al. [9] employed CNN and CNN+LSTM and obtained the highest accuracy of 96.43%.

Multimodal physiological signal approaches have also shown enhanced classification accuracy. Aristizabal et al. [22] integrated ECG, RESP, EDA, and cortisol features using a neural network. The authors achieved the highest accuracy of 96.05% using a Neural Network. However, in some cases, traditional ML models outperformed DL Models. Henry et al. [101] evaluated ECG and BVP signals with DL models such as ResNet and MLP but achieved a higher accuracy of 81.1% using SVM. The authors achieved an accuracy of 71.5% and 64.2% using ResNet and MLP, respectively.

DL techniques have also been applied to ECG and EDA signals to capture autonomic nervous system activity associated with anxiety. Grigoraș et al. [251] employed a 1D CNN on ECG signals and achieved the highest accuracy of 83.29%. Additionally,

Table 3.6: Overview of anxiety classification studies using DL.

Author	Stimuli	Preprocessing	Dataset (Public/Private)	Signal	Subjects	Channels	Features	Algorithms	Accuracy
Xie et al.[267]	Stroop task	BPF:0.5-100Hz, notch: 50Hz	Private	EEG	20	31	Brain network features, prefrontal lateralization	CNN, RF, LR, MLP	CNN: 67.67%, RF: 87.69%, LR: 70.76%, MLP: 56.92%
Baghdadi et al.[30]	Face-to-face Stimulation	Finite impulse response 4-45Hz, EEGLAB artifact removal	Public	EEG	23	14	TD and FD	SVM, KNN, SSAE	SVM: 77.40%, KNN: 81.40%, SSAE: 83.37%
Arsalan et al.[24]	TSST	PSD	Private	EEG	65	4	–	MLP, RF, LR	MLP: 76.92%, RF: 87.69%, LR: 70.76%
Aristizabal et al.[22]	TSST	–	Private	ECG, RESP, EDA, Cortisol	18	–	–	Neural Network	96.05%
Muhammad et al.[165]	Face-to-face stimulation	FIR: 4-45Hz, PSD	Public	EEG	23	14	FD	DT, KNN, SVM, MLP, RF	DT: 85.39%, KNN: 84.69%, SVM: 86.98%, MLP: 88.10%, RF: 94.90%
Henry et al.[101]	Anxiety inducing video clips	BVP: 0.5-5Hz, ECG: HPF	Public	ECG, BVP	45	–	TD and FD	SVM, RF, XGB, MLP, ResNet	SVM: 81.1%, RF: 75.2%, XGB: 65.3%, MLP: 64.2%, ResNet: 71.5%
Mohan et al.[163]	Anxiety inducing video clips	FOBF filter	Public	EEG	30	–	TD and FD	CNN, ANN, KNN	Deep CNN: 97.6%, ANN: 94.5%, KNN: 80.3%
Liu et al.[143]	–	BPF: 4-30Hz, ICA	Private	EEG	81	–	MSTCNN, SE attention	MSTCNN, CNN+LSTM, Squeeze-and-Excitation Networks (SE)	MSTCNN: 99.19%, CNN+LSTM: 97.25%, MSTCNN + SE Attention: 99.47%
Mokatren et al.[164]	SAD Patients data	BPF: 1-50Hz, FFT	Private	EEG	64	34	FD	CNN, SVM, KNN	CNN: 87%, SVM: 79%, KNN: 75%
Ezzi et al.[9]	Social performance task	FIR BPF: 0.4-50Hz, spatial filters	Private	EEG	89	–	FD	CNN+LSTM, CNN	CNN: 96.43%, CNN+LSTM: 96.43%
Grigoraşi et al.	Short spider phobic videos	Bandpass filter (5-12Hz)	Public	ECG	57	–	–	1D CNN	83.29%

NOTE: *TD: Time Domain; FD: Frequency Domain*

Muhammad et al. [165] analyzed EEG signals with multiple classifiers, including RF, SVM, KNN, DT, and MLP. However, RF outperforms other models by achieving the highest accuracy of 94.90%. Similarly, Arsalan et al.[24] applied MLP, RF, and LR on EEG data, where RF achieved the highest accuracy of 87.69% and outperformed MLP with an accuracy of 76.92%. Similarly, Xie et al. [267] applied CNN, RF, LR, and MLP to EEG signals. The authors observed that RF outperformed CNN by achieving the highest accuracy of 87.69% and showed that tree-based ML models provide a robust classification for EEG data.

In conclusion, DL techniques significantly enhance classification accuracy. However, traditional ML models continue to offer competitive performance, especially in multimodal settings [267]. CNNs and hybrid networks effectively learn complex spatial and temporal dependencies in EEG data [143]. On the other hand, SVM, RF, and ensemble learning models still demonstrate superior results in certain cases,

particularly for cardiovascular and electrodermal signals [101][22]. The choice between ML and DL depends on the dataset, signal type, and preprocessing approach, highlighting the ongoing need to balance computational complexity with classification performance. Table 3.6 summarises the DL studies for physiological signals-based anxiety classification.

3.3.5 Challenges in ML-Based Stress and Anxiety classification

In previous studies, ML techniques have shown advancements in stress and anxiety classification of physiological signals. However, several challenges exist in the related work.

1. **Class Imbalance in Physiological Signal based Datasets:** Several publicly available datasets such as Spiderphobic [109] and DASPS [30] suffer from imbalanced class distributions. This class imbalance skews ML models towards predicting non-stress conditions, reducing sensitivity in detecting stress and anxiety. Researchers have explored techniques such as SMOTE [54] and cost-sensitive learning to mitigate this issue, but balancing class distributions remains an ongoing challenge.

2. **Feature Selection and Dimensionality Reduction:** Physiological signals generate high-dimensional data, which makes feature selection crucial for optimizing model efficiency and interpretability. Studies applying filter-based [115] [100][223], wrapper-based [169] [268][259], and hybrid feature selection [96] techniques demonstrate improved classification performance. However, determining the optimal feature subset and removing redundant or irrelevant features without losing valuable information remains an open challenge.

3. **Preprocessing Challenges:** It is challenging to preprocess physiological signals to remove noise and artifacts, especially in dynamic environments such as driving and walking. Artifacts such as baseline drift, motion artifacts, and power line interference can contaminate the signals, necessitating advanced filtering and preprocessing techniques for accurate analysis [51].

4. **Trade-off between Accuracy and Computational Efficiency:** DL models, particularly CNNs and LSTMs, require substantial computational power, limiting their deployment in real-time stress monitoring applications. While traditional ML models such as SVM and RF offer computational efficiency, they may struggle with capturing the complex temporal dependencies inherent in physiological signals [206][164]. The trade-off between computational efficiency and model accuracy necessitates further exploration of lightweight, real-time capable ML solutions.

5. **Lack of Explainability and Model Interpretability:** ML models for stress and anxiety classification often work as black boxes, which makes it difficult to interpret how specific physiological features contribute to classification decisions. Studies have emphasized the importance of explainability methods such as SHAP and LIME for improving model transparency [231]. However, interpretability remains challenging, particularly for DL models, which rely on hierarchical feature representations.

6. **Lack of Privacy Concerns:** Traditional ML and DL approaches used in automated stress and anxiety classification systems often lack mechanisms to ensure users' sensitive physiological data privacy during model training. This limitation increases the risk of unauthorized access, enabling potential attackers to exploit biometric and physiological information.

3.4 Answers to RQs

This thesis answers the following research questions based on the literature survey on stress and anxiety classification using physiological signals:

- **RQ1:** What are the publicly available datasets used for stress and anxiety classification, and what are their limitations?

  Ans: Several publicly available datasets support stress and anxiety classification using physiological signals. MAT [280] and SAM40 [84] provide EEG-based recordings, while CLAS [150], SRAD [99], Spiderphobic [109], and

AffDataset [95] include multimodal signals such as ECG, HRV, EDA, BVP, and RESP. However, these datasets have several limitations. Most classify stress into only two categories, restricting the analysis of varying stress levels. Additionally, no publicly available dataset integrates EEG, HRV, EDA, and BVP within an academic environment, limiting research on stress classification in educational settings. Stress and anxiety classification in academic contexts requires datasets that reflect the cognitive challenges students face. Incorporating these physiological signals in such environments would improve research on stress classification and intervention strategies. Furthermore, datasets like SRAD [99], AffDataset [95], and DPPS [66] have small sample sizes, reducing their generalizability.

- **RQ2:** What Machine Learning and Deep Learning techniques have been explored for stress and anxiety classification using physiological signals?

  Ans: Researchers have explored various ML and DL techniques for stress and anxiety classification using physiological signals. Traditional ML models such as SVM [29] [151][207][251], RF [80][248][223], KNN [29][26][277][188], and DT[268][180][52][188] have been widely used due to their ability to handle structured physiological data effectively. These models rely on feature extraction techniques from EEG, ECG, EDA, and HRV signals to classify stress and anxiety levels. Ensemble learning methods, including GBM [251][79][51], ADB [188], and XGB [248], have further improved classification accuracy by combining multiple weak classifiers into a robust predictive model. In contrast, DL approaches, such as CNN [39][199][88][267][164] and LSTM [91][247][108] networks, have gained attention for their capability to learn complex temporal and spatial patterns from raw physiological signals. Hybrid architectures combining CNNs with LSTMs [108][199][148][143][9] or autoencoders [30] have shown promise in extracting high-level features while maintaining temporal dependencies. Despite their success, DL models require large datasets and extensive computational resources, whereas traditional ML techniques offer competitive performance with fewer data and simpler implementations. The choice of model depends on the dataset characteristics,

signal type, and classification complexity, highlighting the ongoing need to balance interpretability and accuracy in stress and anxiety classification systems.

- **RQ3:** What are the key challenges in developing robust Machine Learning models for stress and anxiety classification?

  Ans: Developing robust ML models for stress and anxiety classification presents multiple interconnected challenges that affect performance and generalizability. One major issue is class imbalance, as many publicly available datasets contain skewed distributions, causing models to favor non-stress conditions and reducing their sensitivity in detecting stress and anxiety [109] [30]. In addition, the high dimensionality of physiological signals demands feature selection and dimensionality reduction to retain only the most relevant information without losing critical details. However, selecting optimal features remains complex, particularly when combined with the challenge of preprocessing physiological data. Noise, motion artifacts, and baseline drift often degrade classification performance, requiring sophisticated filtering techniques for accurate analysis [51]. Furthermore, computational efficiency is a concern, especially for DL models like CNNs and LSTMs, which require substantial processing power and limit their feasibility for real-time applications [206][164]. Model interpretability further complicates development, as many ML and DL models function as black boxes, making it difficult to understand their decision-making process and limiting trust in automated systems. Additionally, privacy concerns remain unresolved, as traditional ML and DL approaches often lack secure mechanisms to protect sensitive physiological data, increasing the risk of unauthorized access and exploitation. Addressing these challenges is crucial for improving stress and anxiety classification models' reliability, efficiency, and ethical considerations.

3.5 Evaluation Measures for ML-Based Stress and Anxiety Classification

This section outlines the evaluation metrics essential for assessing the effectiveness of stress and anxiety classification models. These measures are crucial for validating the

performance of the proposed methodologies, which address both binary and multi-level classification. It is essential to ensure the classification models' reliability, generalizability, and practical applicability, particularly in real-time scenarios involving physiological signal data. The subsequent sections discuss various performance metrics utilized in this study and their significance in stress and anxiety classification.

3.5.1 Classification Performance Measures

This thesis assesses the classification performance of the proposed stress and anxiety classification models using widely recognized evaluation metrics, including the Confusion Matrix, Accuracy, F1-score, Sensitivity, and Specificity. These metrics provide a comprehensive understanding of the model's effectiveness in distinguishing between different stress and anxiety levels. The evaluation encompasses both binary classification and multi-class classification. Each metric plays a critical role in analyzing the model's capability to handle imbalanced datasets, reduce misclassification errors, and ensure overall reliability in real-world applications.

3.5.1.1 Confusion Matrix

The Confusion Matrix is a fundamental tool for evaluating the performance of classification models. It presents a structured comparison between actual and predicted classifications, allowing for a detailed analysis of model performance. Table 3.7 shows that the matrix consists of four key components: *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), and *False Negatives* (FN). These values illustrate how accurately the model classifies stress and anxiety levels, highlighting areas where it performs well and where misclassifications occur. The confusion matrix provides the following performance metrics:

1. **Accuracy:** The proportion of correctly classified instances across all predictions provides an overall measure of the model's effectiveness.

2. **F1-score:** The harmonic mean of Precision and Recall, ensuring a balanced evaluation of classification performance, is particularly useful when handling class imbalances.

Table 3.7: Confusion Matrix representation.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

3. **Sensitivity:** The percentage of actual positive cases correctly identified by the model, indicating how effectively stress and anxiety states are detected.

4. **Specificity:** The proportion of actual negative cases correctly classified, highlighting the model's ability to distinguish non-stress and non-anxiety states.

3.5.1.2 Classification Accuracy

Classification accuracy is a fundamental metric for evaluating ML models. It measures the proportion of correctly classified instances out of the total cases. Accuracy is calculated as the sum of true positives (TP) and true negatives (TN) divided by the total number of predictions, including false positives (FP) and false negatives (FN), as shown in Equation 3.1:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Accuracy provides an overall measure of a model's performance. However, it is not always reliable for stress and anxiety classification. The physiology-based datasets often suffer from class imbalance, where non-stress cases significantly outnumber stress cases. In such scenarios, a model may achieve high accuracy by predominantly predicting the majority class while failing to detect actual stress cases effectively. It makes accuracy insufficient, as a high value does not necessarily indicate good classification performance.

False negatives are particularly concerning in stress and anxiety classification since misclassifying stressed individuals as non-stressed can prevent timely intervention. Similarly, false positives may lead to unnecessary concerns and interventions. Therefore, additional metrics such as F1-score, Sensitivity, and Specificity are used alongside accuracy to address these limitations. The F1 score balances precision

and recall, which is more suitable for imbalanced datasets. While sensitivity ensures that true stress cases are detected, specificity evaluates how well the model identifies non-stress cases.

3.5.1.3 F1-score

F1-score is a crucial metric that balances precision and recall, providing a comprehensive evaluation of a classification model's performance. It is particularly useful when dealing with imbalanced datasets, as it considers both false positives and false negatives. The F1-score is the harmonic mean of precision and recall, ensuring that both metrics contribute equally to the final score. Equation 3.2 defines F1-score:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.2}$$

A high F1 score indicates that the model achieves a good balance between precision and recall, reducing both false positives and false negatives. In stress and anxiety classification, where misclassifications can lead to incorrect assessments of mental health conditions, the F1-score provides a more reliable measure of performance than accuracy alone. A low F1 score suggests that the model struggles either with identifying positive instances or with minimizing false classifications, highlighting the need for further optimization.

3.5.1.4 Sensitivity

Sensitivity, also known as recall or the true positive rate. It measures the model's ability to identify actual positive instances correctly. It represents the proportion of correctly predicted positive cases out of all actual positive cases. A high sensitivity value indicates that the model effectively detects positive instances, while a low sensitivity value suggests that many positive cases are being misclassified as negative. Equation 3.3 defines the formula for sensitivity.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.3}$$

Sensitivity plays a crucial role in correctly identifying individuals experiencing

high stress or anxiety, especially in applications like medical diagnosis and stress or anxiety classification, where missing positive cases can have severe consequences. A model with low sensitivity may fail to detect such cases, leading to underdiagnosis or inadequate interventions. However, high sensitivity alone does not guarantee overall performance, as it must be balanced with specificity to avoid excessive false positives. Therefore, optimizing sensitivity is essential for improving the reliability of stress and anxiety classification systems while minimizing misclassifications.

3.5.1.5 Specificity

Specificity measures a model's ability to correctly identify negative cases by distinguishing between actual negatives and false positives. It is particularly important in stress and anxiety classification to avoid falsely diagnosing individuals who are not experiencing these conditions. A high specificity indicates that the model effectively excludes non-stressed or non-anxious individuals, reducing unnecessary interventions. However, the model must balance specificity with sensitivity to classify both stressed and non-stressed individuals accurately. Equation 3.4 defines formula for specificity:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.4}$$

Maintaining a balance between sensitivity and specificity is essential for developing an effective stress and anxiety classification system that minimizes false positives and false negatives.

3.5.1.6 Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

The Receiver Operating Characteristic (ROC) curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) across different classification thresholds. It visually represents the trade-off between sensitivity and specificity for a classification model. The Area Under the ROC Curve (AUC-ROC) quantifies the overall ability of the model to discriminate between classes. An AUC value of 1 indicates perfect classification, while an AUC of 0.5 suggests performance equivalent to random guessing. Higher AUC values reflect better model performance across varying threshold settings, making it a robust metric for evaluating imbalanced datasets

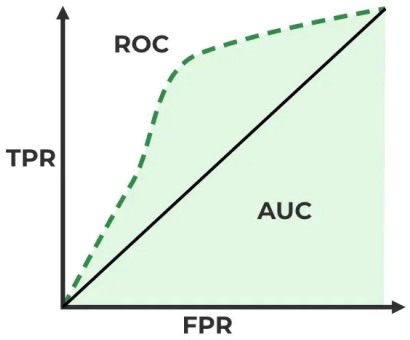


Figure 3.6: Illustration of ROC curve and the Area Under the Curve. The AUC quantifies the model's performance across all threshold levels [28].

and classification reliability. Figure 3.6 illustrates the ROC curve and the corresponding AUC. The shaded region under the ROC curve represents the area that quantifies the classifier's ability to separate the positive and negative classes.

3.5.2 Statistical Analysis

This thesis utilizes statistical analysis to assess the significance of selected features and validate model performance. Statistical analysis plays a crucial role in ML. It ensures that the extracted features meaningfully contribute to classification rather than arise from random variations in the data. Applying statistical methods allows researchers to determine whether specific physiological signals effectively distinguish between different stress and anxiety levels. This process is particularly important in feature selection, as it removes irrelevant or redundant features, enhancing model efficiency and interpretability.

This work employs Analysis of Variance (ANOVA) as a statistical technique to evaluate feature significance in stress and anxiety classification. ANOVA quantifies the variance between groups to identify distinguishing features and determine whether physiological signal differences across stress and anxiety levels are statistically significant. It produces an F-value, which measures the ratio of variance between and within groups, providing insight into the relevance of each feature. The corresponding p-value further indicates the likelihood that observed differences occur by chance, with a lower p-value (typically $x \leq y$ 0.05) confirming that a feature holds meaningful discriminative power.

3.5.3 ANALYSIS OF VARIANCE (ANOVA)

Statistical analysis plays a crucial role in validating feature selection and assessing the significance of extracted features in classification tasks. The Analysis of Variance (ANOVA) can be utilized to determine whether the means of different groups vary significantly. ANOVA helps identify features that contribute meaningfully to classification rather than being influenced by random variations by analyzing variance within and between groups. This method assumes that observations are independent, follow a normal distribution, and have equal variance across groups. ANOVA computes an F-statistic to evaluate the significance of differences among feature groups. The F-statistic is calculated using the following Equation 3.5:

$$F = \frac{MS_B}{MS_W} \tag{3.5}$$

where MS_B is the Mean Square Between groups and MS_W is the Mean Square Within groups. The p-value, derived from ANOVA, quantifies the probability that the observed differences occur due to chance. A low p-value (typically p $\leq$ 0.05 0.05) indicates that a feature significantly contributes to classification, ensuring that the model captures meaningful distinctions rather than noise. Conversely, a high p-value suggests that the feature lacks statistical relevance and may not improve classification performance. Since the F-value and p-value are inversely related, a high F-value typically corresponds to a low p-value, reinforcing the statistical importance of a feature.

3.6 Summary

This chapter reviews ML-based stress and anxiety classification, which covers publicly available datasets, classification techniques, challenges, evaluation metrics, and statistical validation methods. It explores physiological signal-based datasets for model development. It further provides a comprehensive review of related studies and presents the application of signal processing and feature selection techniques across stress and anxiety classification. It highlights the role of traditional ML models, such as SVM, RF, and KNN, along with DL approaches, like CNNs and LSTMs, for stress and anxiety

classification. While DL techniques effectively capture complex patterns in EEG signals, traditional ML models often outperform them with multimodal physiological signals. The discussion identifies distinct major challenges in stress and anxiety classification. The chapter explains accuracy, F1-score, sensitivity, and specificity to emphasize the importance of handling imbalanced data and assessing classification performance. Finally, it highlights statistical analysis using ANOVA, ensuring that selected features hold significant discriminatory power for classification.

This chapter is based on the following work:

- **J4: Shikha,** Divyashikha Sethia, and S. Indu. "A Systematic Review on Physiology-based Anxiety Detection using Machine Learning" **Biomedical physics & engineering express** (2025): 1-36. **ESCI, Impact factor: 1.3, Publisher: IOP Science**. Doi: `https://doi.org/10.1088/2057-1976/add5fc`. **(Published)**.

- **Conference 2: Shikha,** Divyashikha Sethia, and S. Indu. "A Review on Feature Selection Techniques for Anxiety Classification from Physiological Signals." *In Proceedings of the IEEE International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT), 2025.* Doi: `https://doi.org/10.1109/CE2CT64011.2025.10939412`. **(Published)**.

Chapter 4

KRAFS-ANet: A Scientific Methodology for EEG-Based Stress Classification

This chapter presents a lightweight and automated framework, **KRAFS-ANet**, for EEG-based stress classification. The framework integrates automated EEG channel selection with an ensemble stacking model to enhance classification efficiency and performance. The channel selection process identifies the most relevant EEG channels while reducing computational complexity. Additionally, the framework conducts comprehensive experiments and employs an ensemble stacking approach that combines bagging-based models as base classifiers, with an Artificial Neural Network (ANN) serving as the meta-classifier. The use of ensemble stacking also enables the exploration of prediction probability-based features to improve deep learning performance further. The chapter further demonstrates that the robustness and generalizability of KRAFS-ANet are validated across three benchmark datasets: MAT [280], SAM40 [84], and DASPS [30].

4.1 Motivation

EEG signals offer a non-invasive and objective approach for mental stress classification. However, raw EEG recordings typically include multiple channels, many of which

contribute minimal information, resulting in increased computational overhead and reduced efficiency in real-time or wearable systems. While several studies rely on manual channel selection based on neuroscience expertise, these approaches lack scalability and generalizability across datasets [223][76][91].

Furthermore, Deep learning models such as CNNs and LSTMs have demonstrated high accuracy in EEG-based stress classification. However, their significant computational demands and lack of interpretability make them less practical for real-world applications, especially in resource-constrained environments [206][247][148]. These limitations highlight the need for lightweight and interpretable frameworks that can efficiently process EEG data without compromising accuracy.

Ensemble stacking presents a promising solution that combines multiple classifiers to enhance performance while maintaining low complexity. Yet, the integration of automated, data-driven channel selection with ensemble stacking remains underexplored in EEG-based stress detection. This chapter addresses these challenges by proposing a robust and efficient KRAFS-ANet framework, which leverages Normalized Mutual Information (NMI) and Recursive Feature Elimination (RFE) for channel selection and integrates an ensemble stacking model to improve classification accuracy and system efficiency.

Contributions:

The key contributions of this chapter are as follows:

1. **Exploration of ML, DL, and Ensemble Techniques:** Conduct comprehensive experiments on the MAT dataset [280] to compare the performance of machine learning, deep learning, bagging, and stacking models for EEG-based stress classification.

2. **Development of the KRAFS-ANet Framework:** Propose a novel ensemble stacking framework that combines NMI and RFE for automated channel selection and integrates bagging KNN, bagging RF, and bagging SVM with an ANN as the meta-classifier.

3. **Validation on Multiple Benchmark Datasets:** Evaluate the proposed framework on the SAM40 [84] and DASPS [30] datasets to demonstrate its robustness and generalizability for stress and anxiety classification.

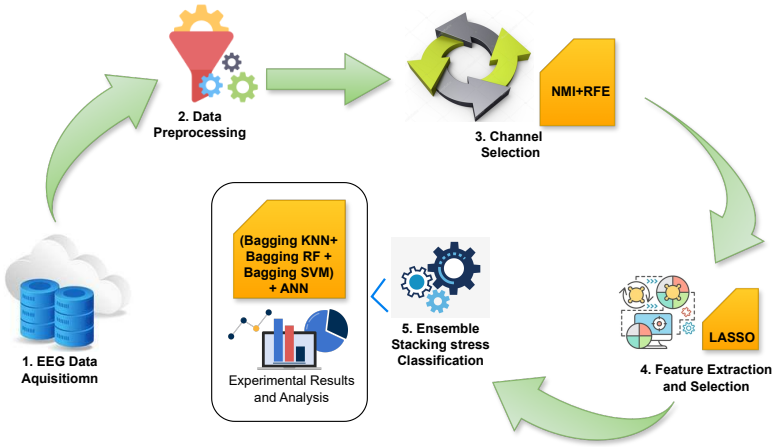


Figure 4.1: KRAFS-ANet framework: EEG-based stress classification proposed methodology with channel selection and ensemble stacking.

4.2 Experimental Methodology

This section explores the MAT dataset [280] to develop an efficient methodology that integrates a channel selection algorithm for stress classification. Further, it conducts extensive experiments with ML, DL, bagging, and various ensemble stacking techniques, as shown in Figure 4.1. It validates the effectiveness of the proposed methodology using the EEG-based SAM40[84] and DASPS[30] datasets.

4.2.1 Dataset Description

The KRAFS-ANet framework employs the MAT dataset [280] as the primary dataset and validates its performance on the secondary SAM40 [84] and DASPS [30] datasets, as summarized in Table 4.1

1. **MAT Dataset [280]:** This dataset contains EEG recordings from 36 students performing mental arithmetic tasks. Participants engage in serial subtraction of two numbers while using 23 channels of EEG signal. Each recording consists of artifact-free EEG segments, including a resting state with closed eyes and EEG data during mental tasks. It utilizes data sampled at 500Hz from 20 electrodes arranged in a 10–20 pattern across the scalp.

2. **SAM40 Dataset [84]:** This dataset includes EEG recordings from 40

Table 4.1: Summary of EEG datasets utilized for KRAFS-Anet framework.

Description/ Dataset	**MAT [280]**	**SAM40 [84]**	**DASPS [30]**
Domain	Stress	Stress	Anxiety
Participants	36	40	23
Experimental Task and relax phase	Mental arithmetic task	SCWT, mental arithmetic, mirror image and relax phase	Recitation (15 sec) and self-recall (15 sec) for six anxiety-inducing scenarios
EEG Device	Neurocom EEG System	Emotiv Epoc Flex Gel Kit	Emotiv EPOC
EEG Electrodes	19	32	14
Sampling Rate	500 Hz	128 Hz	128 Hz

participants, consisting of 14 females and 26 males, with an average age of 21.5 years. EEG signals were collected while participants engaged in three cognitive tasks: the Stroop color-word test (SCWT), solving arithmetic problems, and identifying symmetric mirror images, along with a relaxation state. The experiment aims to assess short-term stress responses triggered by these tasks. Each task runs for 25 seconds and repeats across three trials. EEG data is recorded using a 32-channel Emotiv Epoc Flex gel kit and segmented into non-overlapping 25-second epochs based on the performed tasks.

3. **DASPS Dataset [30]:** This dataset contains raw EEG recordings in .edf format collected from 23 participants. The authors capture EEG signals using a 14-channel Emotiv EPOC wireless EEG headset with a sampling rate of 128 Hz. Additionally, the dataset provides preprocessed data in .mat format. During the recording sessions, participants experienced six anxiety-inducing scenarios while keeping their eyes closed and minimizing movement. Each scenario includes two phases: a 15-second description by a psychotherapist, followed by 15 seconds in which participants recall the scenario. After completing the sessions, participants evaluated their emotional states using the SAM scale.

4.2.2 Data Preprocessing

The EEG signal preprocessing involves multiple steps to enhance signal quality and ensure reliable classification. This work applies ICA using the FastICA technique to remove physiological artifacts and separate independent signal sources. Further, it utilizes the *MNE 4* [1] Python package to process the EEG signals by applying a 5th-order Butterworth bandpass filter from the 4 to 44 Hz range. Additionally, this work applies SMOTE to the SAM40 dataset [84] to handle the class imbalance issue caused by unequal numbers of stress and relaxation segments.

4.2.3 Channel Selection

In EEG-based stress classification, each channel may contribute several features, which will lead to a substantial increase in the feature vector space. Previous research has utilized feature selection techniques to optimize the number of features without eliminating channels [26][268][197][24]. Although feature selection applied across multiple channels can achieve high accuracy in a laboratory environment. However, its practical efficiency tends to decrease in home-based or everyday applications due to longer setup times and increased user discomfort. As a result, identifying the most relevant channels for source localization becomes important for developing a more practical and user-friendly stress assessment system.

In many studies, channel selection approaches rely heavily on neuroscience expertise [223][91][137], correlating specific brain regions with tasks related to mental stress. However, researchers have applied various EEG channel selection techniques in distinct domains like emotion classification, using methods such as ReliefF [274], Maximum Relevance Minimum Redundancy (mRMR) [269], and NMI [255]. Nevertheless, limited research exists on channel selection for stress classification.

The current study utilizes an automated channel selection approach to optimize the performance of EEG-based stress classification. The proposed framework employs NMI to identify complex, non-linear relationships between EEG channels and stress labels, which provides a comprehensive measure of shared information [255]. Additionally, it implements RFE to iteratively refine the channel selection process,

[1] `https://mne.tools/stable/generated/mne.preprocessing.ICA.html`

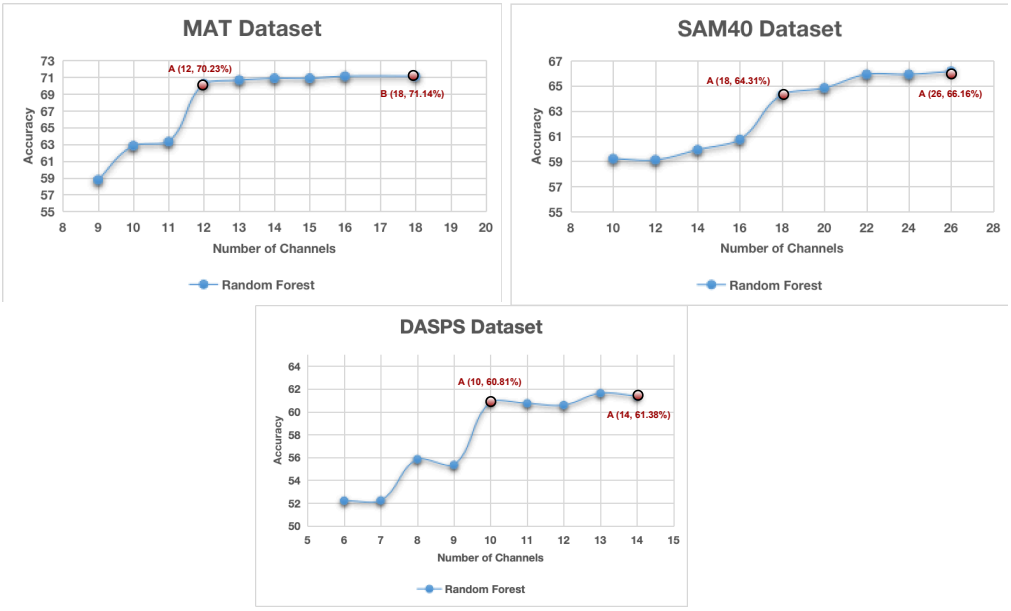


Figure 4.2: KRAFS-Anet Framework: Number of channels and accuracy rate using NMI in RF-based stress classification.

reduce redundancy, and retain the most informative channels.

4.2.3.1 NMI-RFE Channel Selection

This work utilizes preprocessed EEG signals and computes NMI for each channel corresponding to stress labels during the channel selection process. Channels with higher NMI values contain more relevant information for stress classification. The method ranks all channels in descending order based on their NMI values and selects the top channels with the highest mutual information as the most informative for classification.

Furthermore, it evaluates the average classification accuracy using the RF algorithm, which is selected based on its superior performance during preliminary experiments without channel selection. As shown in Table 4.4, RF achieves the highest accuracy compared to other classifiers. Figure 4.2 presents the variation in average classification accuracy for the MAT [280], SAM40 [84], and DASPS [31] datasets as the number of selected channels changes. The results suggest that selecting 12, 18, and 10 channels yields optimal performance for the MAT [280], SAM40 [84], and DASPS [30] datasets, respectively. It further applies the RFE algorithm across all three datasets to enhance

Table 4.2: KRAFS-Anet Framework: Summary of extracted EEG features on selected channels.

Domain	Feature Name	No of Features
Time Domain	Hjorth Parameters (Activity, Mobility, Complexity)	3
	Higuchi Fractal Dimension	1
	Peak to Peak Amplitude	1
	Mean	1
	Variance	1
	Standard Deviation	1
	Skewness	1
	Kurtosis	1
Frequency Domain	Relative Power of Theta (4–8 Hz), Alpha (8–12 Hz), Beta (12–35 Hz)	3
	Band Power Ratio of Theta/Alpha, Alpha/Beta	2
	Spectral Entropy	1

the selection process. This refinement results in a final selection of 9 channels for the MAT [280], 12 for the SAM40 [84], and 6 for the DASPS [30].

4.2.4 Feature Engineering

The KRAFS-Anet framework utilizes features from multiple domains to enhance classification accuracy. Table 4.2 presents a comprehensive summary of the extracted features. Section 2.2.1 in Chapter 2 explains the detailed description of EEG features. Further, MinMax normalization is applied to all extracted features to ensure consistent scaling.

For feature selection, the study employs the Least Absolute Shrinkage and Selection Operator (LASSO). This regression-based technique applies L1 regularization to eliminate less relevant features by driving their coefficients to zero. It enhances model efficiency by selecting the most informative features while reducing redundancy.

4.2.5 Bagging

Bagging, or Bootstrap Aggregating, is an ensemble learning technique commonly used in supervised machine learning for classification and regression tasks. Breiman et al. [262] introduced this method in 1996 to improve accuracy and reduce overfitting.

Table 4.3: KRAFS-Anet Framework: Selected hyperparameters of machine learning and deep learning models.

Model	Selected Hyperparameters
RF	criterion: 'gini', max_depth: '20', max_features: 'sqrt', min_samples_leaf: '1', min_samples_split: '2', n_estimators: '100'
DT	criterion: 'gini', max_depth: '8', min_samples_leaf: '1', min_samples_split: '2'
KNN	metric: 'minkowski', n_neighbors: '5', p: '2', weights: 'uniform'
SVM	C: '1.0', degree: '3', gamma: 'scale,' kernel: 'rbf'
Adaboost	learning_rate: '1.0', n_estimators: '150'
ANN	activation: 'tanh', learning_rate: '0.001', hidden_layer_sizes: '(100, 50)', optimizer: 'adam'
CNN	Filters: '32'; Activation Function:'Relu' and 'tanh'; pooling layer:'MaxPooling'; Dropout rate: '0.5'; Dense Layer Size: '32'

The approach generates multiple subsets of training data through bootstrapping, a process that randomly selects samples with replacement. These subsets are then trained in parallel using numerous weak classifiers, which may be of the same or different types. The final prediction is obtained by aggregating the outputs of all classifiers. In regression tasks, the model averages the predictions through a process known as soft voting. In contrast, in classification tasks, it selects the most frequently predicted class through hard voting [233][240].

This study utilizes hard voting, as represented in Equation 4.1.

$$\hat{G}_{bag}(x) = argmax_k \hat{f}_{bag}(x) \tag{4.1}$$

$\hat{G}_{bag}(x)$ denotes the predicted class label for input x, k represents each class from the set of possible classes, and $\hat{f}_{bag}(x)$ is the aggregated (bagged) estimate function for class k.[97].

4.2.6 Ensemble Stacking

Stacking is an ensemble learning approach that utilizes a meta-model to improve predictive performance. This technique integrates predictions from multiple base learners and uses a separate classifier to make the final prediction. Wolpert et al. [263] introduced stacking in 1992 to minimize bias and variance and enhance model

accuracy. The framework consists of two layers, where the first layer includes multiple base learning algorithms, and the second layer features a meta-learner that combines their outputs. It divides the dataset into training and validation sets using k-fold cross-validation, where k-1 folds train the base models, and the remaining fold serves for validation. Equation 4.2 describes how the base learners generate probability distributions for class predictions,

$$P^M(x) = (P^M(c_1 \mid x), P^M(c_2 \mid x), ..., P^M(c_m \mid x)) \tag{4.2}$$

where c denotes the class value, and $P^M(c_i \mid x)$ represents the probability of x belonging to c_i [193].

4.2.7 Hyperparameter Optimization

The hyperparameter tuning process utilizes GridSearchCV [2] to determine the optimal hyperparameters that maximize accuracy. This method systematically explores a predefined grid of hyperparameter values and evaluates each combination using stratified 10-fold cross-validation. The combination that achieves the highest performance is selected and finalized to enhance model efficiency. Table 4.3 presents the chosen hyperparameters for each algorithm.

4.3 Experimental Framework

This section illustrates the comprehensive experiment for the KRAFS-ANet framework, as shown in Figure 4.3. The experimental process follows these phases:

1. The KRAFS-Anet framework utilizes data from the publicly available MAT [280], SAM40 [84], and DASPS [31] datasets.

2. The raw EEG signals undergo preprocessing to remove noise and artifacts.

3. Further, the framework applies NMI+RFE for channel selection on the preprocessed signals. This approach selects 9 out of 20 channels from the

[2] `https://rb.gy/kvlswi`

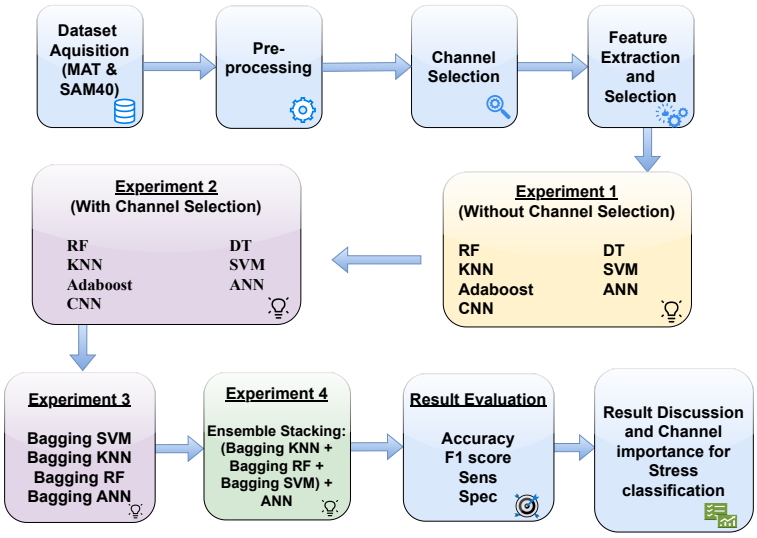


Figure 4.3: Comprehensive experiment for KRAFS-ANet Framework.

MAT dataset [280], 12 out of 32 from SAM40 [84], and 6 out of 14 from DASPS [30].

4. Feature selection takes place on both the full set of channels and the selected channels. The study employs LASSO to reduce dimensionality and enhance classification performance. It also applies MinMax normalization using the Sklearn library to ensure uniform feature scaling.

   (a) ***Experiment 1*** applies machine learning and deep learning models to all extracted and selected features. Machine learning models include SVM, KNN, RF, DT, and Adaboost, while deep learning models consist of CNN and ANN. The study optimizes models using GridSearchCV with stratified 10-fold cross-validation.

   (b) ***Experiment 2*** evaluates machine learning and deep learning models on the selected features from the selected channels. The study compares the performance of models with and without channel selection and observes better results with chosen channels. Table 4.5 shows that among machine learning models, RF, SVM, and KNN achieve the highest accuracy, while ANN performs best in deep learning. The study continues experimentation with RF, SVM, and KNN in machine learning and ANN in deep learning.

It selects ANN for its lightweight nature, universal approximation ability, and capability to maintain high accuracy even with smaller datasets. Its lower computational complexity makes it suitable for resource-constrained devices.

(c) ***Experiment 3*** introduces bagging ensemble models using the best-performing classifiers from machine learning and deep learning on selected channels. Each classifier undergoes hyperparameter tuning through GridSearchCV, leading to improved accuracy.

(d) ***Experiment 4*** implements a novel stacking ensemble model by combining bagging KNN, bagging RF, and bagging SVM. The study employs ANN as the meta-classifier and optimizes it using GridSearchCV. All experiments maintain a constant random state of 42. The final model achieves peak accuracy values of 98.63% on the MAT [280], 97.25% on the SAM40 [84], and 94.92% on the DASPS [30] datasets.

4.4 Experimental Results

This section discusses the experimental results obtained in the development of the KRAFS-ANet framework for EEG-based stress classification.

4.4.1 Performance Measure

This work evaluates the KRAFS-ANet framework using performance metrics such as confusion matrix, Accuracy (Acc), F1-score (F1), Sensitivity (Sens), and Specificity (Spec), as detailed in section 3.7. The model undergoes 10-fold cross-validation, where the dataset is divided into ten equal parts, treating each part as a test set in rotation. This approach ensures model reliability and stability. The confusion matrix further analyzes classification performance by comparing predicted and actual labels.

4.4.2 Experiment 1: Before Channel Selection

Table 4.4 presents the classification performance of machine learning and deep learning models on the MAT [280], SAM40 [84], and DASPS [31] datasets before applying

Table 4.4: KRAFS-Anet Experiment 1: Classification performance before channel selection.

	MAT [280]				SAM40 [84]				DASPS [31]			
Model	**Acc. (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**
RF	**86.81**	**85.58**	**87.69**	**83.56**	**84.44**	**84.78**	**85.65**	**83.26**	**83.17**	**82.84**	**82.42**	**84.90**
DT	74.1	73.46	73.5	75.36	68.72	69.01	69.12	68.3	67.64	66.83	66.34	68.02
KNN	**77.23**	**76.83**	**76.82**	**79.61**	**82.9**	**83.86**	**84.54**	**82.17**	**75.84**	**76.15**	**76.93**	**75.38**
SVM	**79.75**	**80.62**	**81.62**	**84.58**	**76.5**	**77.08**	**75.61**	**77.47**	**84.98**	**84.99**	**83.74**	**85.5**
Adaboost	74.03	76.48	79.45	71.59	66.25	65.49	66.49	66.02	69.01	70.93	70.34	69.73
ANN	**87.23**	**87.75**	**89.83**	**84.53**	**85.58**	**85.71**	**87.81**	**83.41**	**83.32**	**83.83**	**83.17**	**84.23**
CNN	78.05	79.98	80.02	78.03	71.5	71.12	69.59	73.54	70.37	70.09	69.4	72.62

channel selection. For MAT [280], RF, KNN, and SVM achieve the highest accuracies among machine learning models, with 86.81%, 77.23%, and 79.75%, respectively. Adaboost and DT achieve lower accuracies of 74.03% and 74.1%, respectively. Among deep learning models, ANN performs the best with an accuracy of 87.23%, while CNN achieves an accuracy of 78.05%.

SAM40 [83] shows a similar trend, where RF, KNN, and SVM outperform other machine learning models with accuracies of 84.44%, 82.9%, and 76.5%, respectively. ANN achieves the highest accuracy among deep learning models with 85.58%. Adaboost and DT achieve an accuracy of 66.25% and 68.72%, respectively. CNN provides moderate results with an accuracy of 71.5% and an F1-score of 71.12%. Figure 4.4 presents the 10-fold cross-validation results for the three datasets.

For DASPS [30], SVM and RF achieve the highest accuracies of 84.98% and 83.17%, respectively, while ANN performs similarly with 83.32%. KNN records a moderate accuracy of 75.84%, while Adaboost and DT achieve 69.01% and 67.64%, respectively. CNN continues to show lower performance, obtaining an accuracy of 70.37%, which is consistent with its results on the other datasets.

4.4.3 Experiment 2: After Channel Selection

Experiment 2 applies the NMI-RFE algorithm to the preprocessed datasets to identify the most relevant EEG channels. It results in the selection of 9 channels for the MAT [280], 12 channels for the SAM40 [84], and 6 channels for the DASPS [31]. Following channel selection, the thesis performs feature extraction and applies the LASSO feature selection algorithm to retain the most informative features from the selected channels.

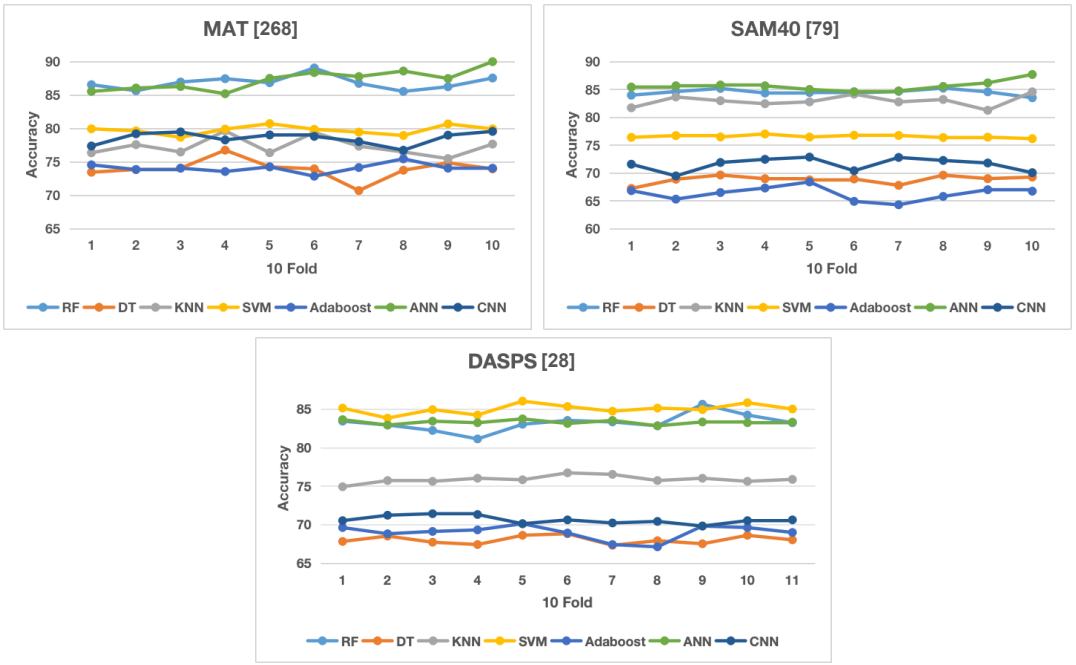


Figure 4.4: KRAFS-Anet Experiment 1: 10-fold cross-validation on classification algorithms before channel selection.

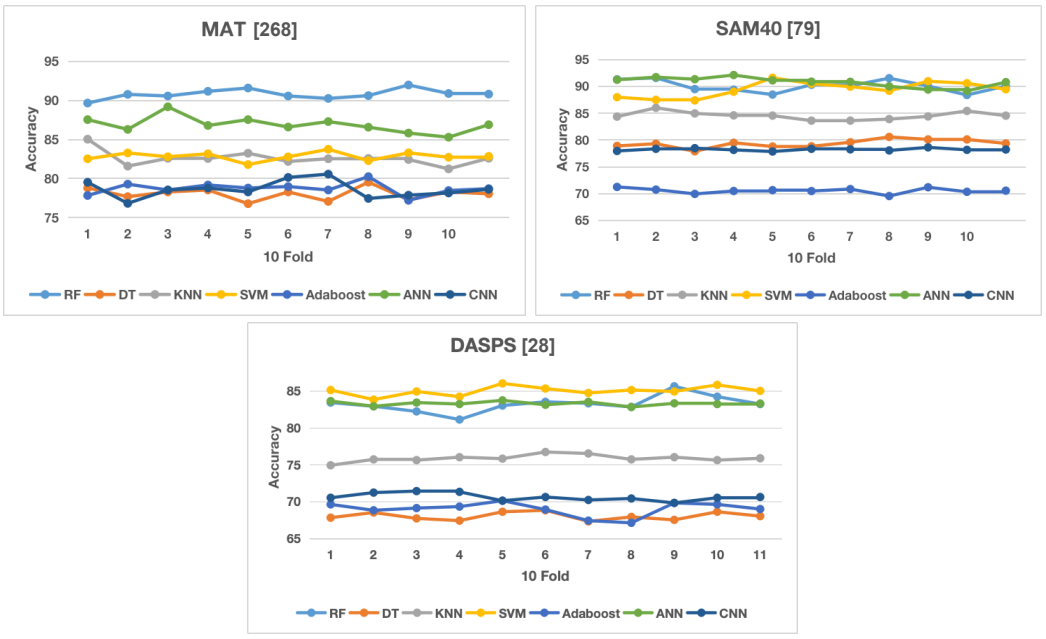


Figure 4.5: KRAFS-Anet Experiment 2: 10-fold cross-validation on classification algorithms after channel selection.

Table 4.5: KRAFS-Anet Experiment 2: Classification performance after channel selection.

	MAT [280] (9 channels)				SAM40 [84] (12 channels)				DASPS [31] (6 channels)			
Model	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**
RF	**90.76**	**90.82**	**89.32**	**91.77**	**91**	**90.85**	**92.25**	**89.81**	**87.26**	**87.25**	**86.35**	**88.17**
DT	78.02	78.04	77.43	76.36	79.3	79.56	80.12	77	74.34	74.30	73.14	73.24
KNN	**82.55**	**82.83**	**82.96**	**80.44**	**84.48**	**85.69**	**88.9**	**80.57**	**81.29**	**81.18**	**82.79**	**81.80**
SVM	**82.8**	**82.78**	**81.83**	**81.9**	**89.4**	**89.4**	**90.34**	**97.82**	**85.67**	**85.53**	**83.65**	**85.68**
Adaboost	78.67	77.06	79.1	77.22	70.5	71.32	74.72	66.83	71.27	72.54	71.30	71.84
ANN	**86.85**	**86.92**	**87.92**	**85.77**	**90.72**	**90.62**	**91.76**	**89.72**	**85.37**	**85.36**	**84.38**	**85.87**
CNN	78.58	78.82	78.23	79.86	78.17	78.43	78.69	77.66	74.67	74.61	74.06	75.19

Table 4.5 presents the classification performance of machine learning and deep learning models after channel selection, demonstrating a significant improvement in accuracy. On MAT [280], RF achieves the highest accuracy and F1-score, reaching 90.76% and 90.82%, respectively. ANN follows with an accuracy of 86.85% and an F1-score of 86.92%.

For SAM40 [84] and DASPS [30], channel selection enhances the performance of most models. Random Forest achieves the highest accuracy in both datasets, increasing from 84.44% to 91.00% in SAM40 [84] and from 83.27% to 87.26% in DASPS [30]. KNN and SVM also show substantial improvements, while ANN remains the best-performing deep learning model, achieving 90.72% in SAM40 [84] and 85.37% in DASPS [30]. Although DT and Adaboost demonstrate minor improvements after channel selection, they continue to show the lowest accuracy compared to other models. Figure 4.5 illustrates the 10-fold cross-validation results after channel selection.

Across all datasets, ANN consistently outperforms CNN, as both models operate on pre-extracted features instead of raw EEG signals. CNN performs best when applied to raw data or images, where convolutional layers can learn hierarchical spatial patterns [29]. The reliance on pre-extracted features limits CNN's ability to leverage spatial feature extraction, making ANN a more suitable choice for processing these features. As a result, ANN achieves higher accuracy than CNN in stress classification.

4.4.4 Experiment 3: Bagging Model

Recent advancements in ensemble learning have demonstrated its effectiveness in improving the performance of traditional classification algorithms. Tables 4.4 and

Table 4.6: KRAFS-Anet Experiment 3: Classification performance of the bagging models on the MAT [280] and SAM40 [84] datasets.

	MAT [280]				SAM40 [84]				DASPS [31]			
Model	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**	**Acc (%)**	**F1 (%)**	**Sens (%)**	**Spec (%)**
Bagging RF	92.25	92.32	92.54	94.32	91.63	91.48	92.6	90.72	90.45	90.77	89.53	91.45
Bagging KNN	84.62	84.41	86.34	83.3	87.61	87.91	89.36	85.82	85.77	86.17	84.50	87.18
Bagging SVM	90.18	90.44	92.31	91.54	92.85	92.8	91.61	94.1	86.79	86.43	84.49	89.07
Bagging ANN	87.23	86.75	89.83	84.53	91.36	91.2	89.35	93.37	88.82	89.44	91.02	87.44

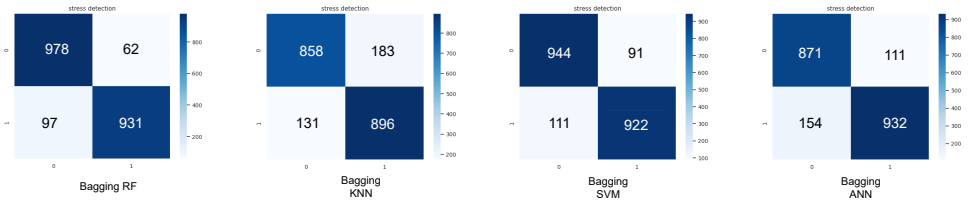


Figure 4.6: KRAFS-Anet Experiment 3: Confusion matrices of bagging models for MAT [280].

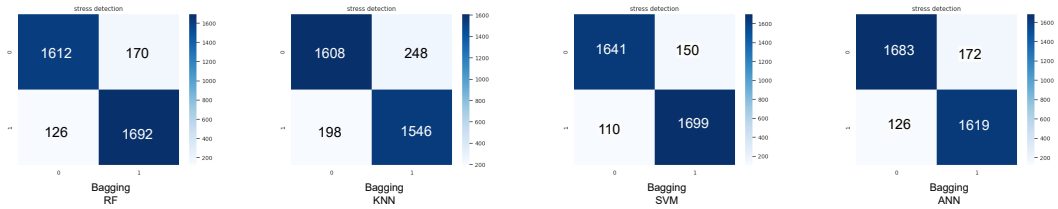


Figure 4.7: KRAFS-Anet Experiment 3: Confusion matrices of bagging model for SAM40 [84].

4.5 indicate that RF, SVM, KNN, and ANN consistently achieve the best results for EEG-based stress classification across all datasets. Based on these findings, Experiment 3 builds bagging models for these high-performing algorithms to enhance classification accuracy and model robustness.

Table 4.6 presents the classification performance of the bagging models on MAT [280], SAM40 [84], and DASPS [31] datasets. On MAT [280], the bagging RF model achieves the highest accuracy of 92.25%, followed by bagging SVM with 90.18% and bagging ANN with 87.23%. SAM40 [84] shows a similar trend, where bagging SVM outperforms other models with an accuracy of 92.85%. Bagging RF and bagging ANN also show competitive performance, achieving 91.63% and 91.36%, respectively.

For DASPS [30], bagging RF achieves the highest accuracy of 90.45%, followed by bagging ANN with 88.82%. Bagging SVM and bagging KNN also perform well,

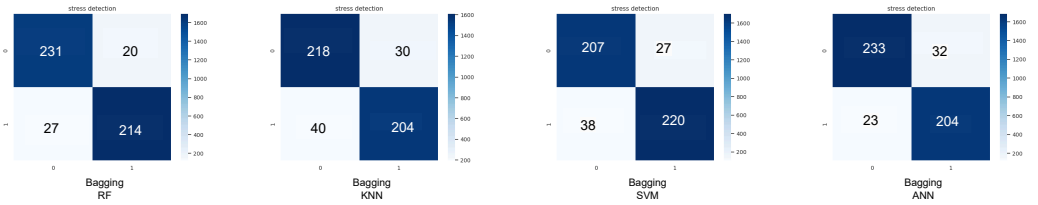


Figure 4.8: KRAFS-Anet Experiment 3: Confusion matrices of bagging models for DASPS [31].

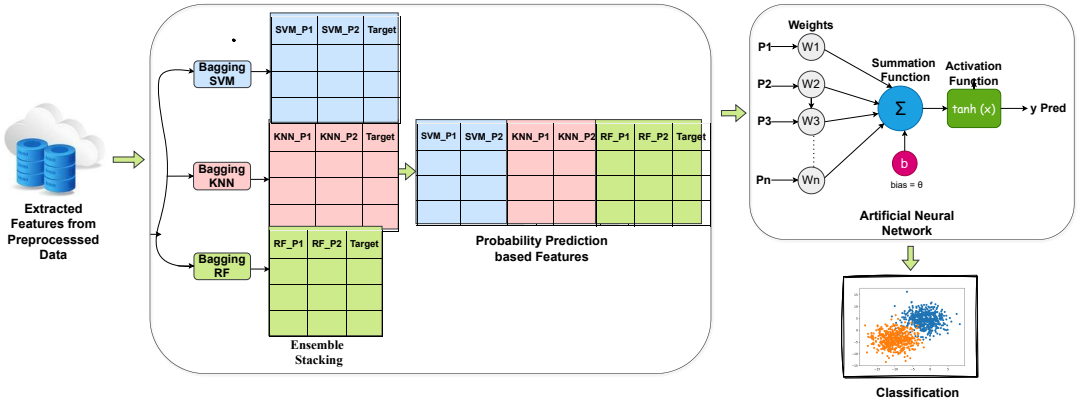


Figure 4.9: KRAFS-ANet framework: Ensemble stacking of bagging models integrated with an ANN meta-classifier

attaining accuracy of 86.79% and 85.77%, respectively. The Figures 4.6, 4.7, and 4.8 illustrate the confusion matrices. It provides insights into the stability and reliability of the bagging models across all datasets.

4.4.5 Experiment 4: Proposed KRAFS-ANet Framework

Stacking differs from bagging by integrating multiple heterogeneous algorithms into a single model using a meta-classifier, combining their predictive strengths to improve classification performance. Experiment 4 builds the KRAFS-ANet framework by merging the four most effective models from previous experiments, each optimized with its best hyperparameters.

The stacking model begins with parallel stacking, where bagging KNN, bagging RF, and bagging SVM generates prediction probability-based features (PBF). These features are then concatenated and passed into an ANN deep learning algorithm, which serves as the meta-classifier due to its superior performance over a bagging ANN. GridSearchCV optimizes the meta-classifier to achieve the highest possible accuracy. Figure 4.9

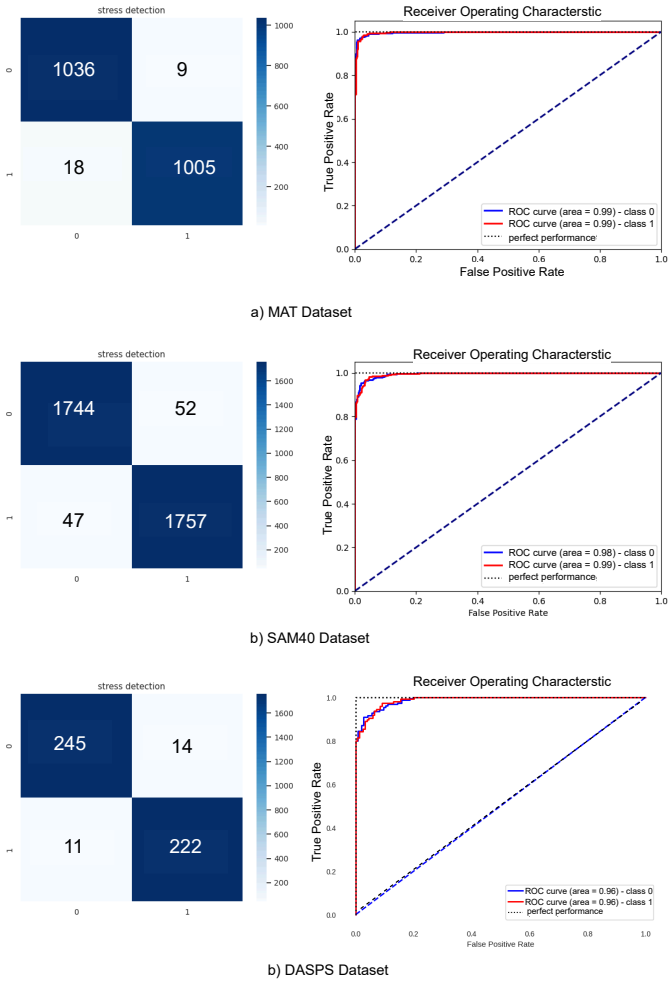


Figure 4.10: KRAFS-Anet Experiment 4: Confusion matrices and ROC curve for stress classification.

illustrates the proposed ensemble stacking model structure for stress classification.

Table 4.7 presents the classification performance of KRAFS-ANet on MAT [280] and its validation on the SAM40 [84] and DASPS [31] datasets. On MAT [280], KRAFS-ANet achieves an accuracy of 98.63%, an F1-score of 98.82%, a sensitivity of 98.23%, and a specificity of 99.11%, demonstrating its robustness in stress classification. The thesis further validates the model on SAM40 [84] and DASPS [30], achieving accuracies of 97.25% and 94.92%, respectively. The corresponding F1-scores are 97.24% for SAM40 and 95.15% for DASPS, confirming the model's ability to generalize effectively across different datasets. Figure 4.10 shows the analysis by illustrating the confusion matrix and ROC curve, providing insights into the model's

Table 4.7: KRAFS-ANet Experiment 4: Proposed framework classification performance on MAT [280], SAM40 [84], and DASPS [31] datasets.

Proposed Framework	Acc (%)	F1 (%)	Sens (%)	Spec (%)
MAT [280] (9 Channels)				
KRAFS-ANet	98.63	98.82	98.23	99.11
SAM40 [84] (12 Channels)				
KRAFS-ANet	97.25	97.24	97.39	97.12
DASPS [30] (6 Channels)				
KRAFS-ANet	94.92	95.15	95.70	94.07

classification performance and ability to distinguish stress levels accurately.

4.5 Discussion

Recent studies have explored machine learning techniques for stress classification using physiological data. However, applying these models in real-world scenarios remains challenging, highlighting the need for lightweight and adaptable approaches. This work presents an optimized ensemble stacking-based KRAFS-ANet framework for efficient stress classification. The framework integrates machine learning, bagging, and ensemble stacking techniques with a systematic channel selection approach. The study conducts experiments on the EEG-based MAT [280] and further validates the framework using the SAM40 [84] and DASPS [30] datasets.

Table 4.8 presents a comparative analysis of the selected classifiers across different evaluation metrics for all four experiments, demonstrating the contribution of each base model. The results from Experiment 2 show significant improvements over Experiment 1, primarily due to the integration of channel selection using NMI+RFE. The combination of the NMI+RFE algorithm offers two key advantages. First, it retains the most relevant channels while eliminating redundant ones, optimizing the model's ability to identify distinct stress-related patterns in EEG signals. Second, it enhances the signal-to-noise ratio, leading to a more precise and effective stress classification model.

This work identifies the most informative EEG channels by considering the non-stationary nature of EEG data and individual variations in brain activation patterns in response to stress stimuli. MAT [280] involves arithmetic tasks, where the selected

Table 4.8: KRAFS-Anet Framework: Stress classification performance of the final models in each experiment on MAT [280], SAM40 [84], and DASPS [31] datasets for KRAFS-Anet.

Experiment	Classifier	MAT [280]			SAM40 [84]			DASPS [30]		
		Train Acc (%)	Test Acc (%)	F1 (%)	Train Acc (%)	Test Acc (%)	F1 (%)	Train Acc (%)	Test Acc (%)	F1 (%)
Experiment 1	RF	91.44	86.81	85.58	88.95	84.44	84.78	85.15	83.17	82.84
	KNN	83.02	77.23	76.83	87.22	82.90	83.86	78.99	75.84	76.15
	SVM	84.80	79.75	80.62	81.90	76.50	77.08	88.51	84.98	84.99
	ANN	91.17	87.23	87.75	90.24	85.58	85.71	87.46	83.32	83.83
Experiment 2	RF	94.84	90.76	90.82	94.82	91.00	90.85	91.71	87.26	87.25
	KNN	86.47	82.55	82.83	89.54	84.48	85.69	84.20	81.29	81.18
	SVM	85.93	82.80	82.78	93.57	89.40	89.40	87.34	85.67	85.53
	ANN	89.53	86.85	86.92	95.15	90.72	90.62	90.17	85.37	85.36
Experiment 3	Bagging RF	95.58	92.25	92.32	95.88	91.63	91.48	92.15	90.45	90.77
	Bagging KNN	87.44	84.62	84.41	92.51	87.61	87.91	88.86	85.77	86.17
	Bagging SVM	92.12	90.18	90.44	96.10	92.85	92.80	90.37	86.79	86.43
	Bagging ANN	89.52	87.23	86.75	94.39	91.36	91.20	91.52	88.82	89.44
Experiment 4	**Proposed KRAFS-ANet**	**99.17**	**98.63**	**98.82**	**99.75**	**97.25**	**97.24**	**97.57**	**94.92**	**95.15**

channels include Fp1, F7, F8, Fp2, O2, P3, T3, Fz, and T6. For SAM40 [84], which incorporates arithmetic tasks, Stroop tests, and mirror image tests, the identified channels are F2, FT8, FT10, F8, PO10, CP5, P8, C3, FCz, and CP1. Similarly, for DASPS [30], where participants recite and recall anxiety-inducing scenarios, the final selected channels are AF4, F7, T7, T8, FC6, and O2, representing the most informative regions for anxiety classification.

Figure 4.11 depicts the predominant activation of the brain's right hemisphere during stress and anxiety, particularly in the frontal, temporal, parietal, and occipital lobes. In Figure 4.11 (a), the activation is evident in regions such as the frontal (Fp1, F7, F8, Fp2, Fz), temporal (T3, T6), parietal (P3), and occipital (O2) areas for MAT [280]. Previous studies have also reported right hemisphere dominance during stress, especially in tasks such as MAT [280], showing activation patterns in the frontal, temporal, parietal, and occipital regions, which align with the findings of this study [223][247][75].

Figure 4.11 (b) illustrates stress-related activation in the SAM40 [84], primarily observed in the frontal and right hemispheres. The activated regions include the frontal (F2, F8), frontotemporal (FT10, FT8), parietal-occipital (PO10), and central-parietal (CP5, P8, CP1) areas. SAM40 [84] consists of distinct tasks, indicating the involvement of both the frontal and central lobes, particularly in the right hemisphere [56][253].

Fig. 4.11 c) illustrates similar findings for DASPS [31] for anxiety classification. The selected channels also exhibit significant activation in the right hemisphere,

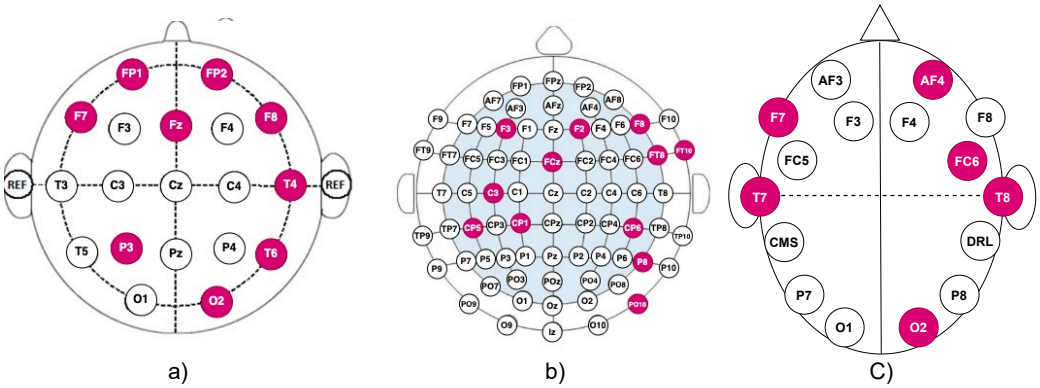


Figure 4.11: Selected EEG channel location after NMI+RFE based channel selection a) 9 channels selected for MAT [280], b) 12 channels selected for SAM40 [84], c) 6 channels selected for DASPS [31] datasets.

particularly in regions such as the frontal (AF4, F7), temporal (T7, T8), and occipital (O2) lobes. These findings align with the stress-related activation patterns observed in the other datasets, confirming the involvement of both frontal and temporal regions during stress. The selection of channels like FC6, T7, and O2 further supports the notion that stress responses vary across different tasks but consistently engage specific brain regions across datasets. Thus, these findings reveal that recognizing mental stress and anxiety for different tasks involves distinct combinations of EEG channels.

The results from Experiment 3 demonstrate that the bagging ensemble significantly improves accuracy across all stress and anxiety classification datasets. For MAT [280], bagging RF achieves notable performance with a moderate increase in accuracy. Bagging KNN also enhances classification results, although its improvement is less significant. Bagging SVM shows substantial gains across all performance metrics, likely due to the ensemble's ability to refine decision boundaries. In contrast, bagging ANN exhibits only a slight accuracy increase of 0.38%, indicating stability but limited performance gains compared to other models. SAM40 [84] also benefits from slight improvements across the machine and deep learning models, with each bagged classifier contributing to enhanced classification performance.

DASPS [30] further highlights the effectiveness of the bagging ensemble approach, with bagging RF achieving an accuracy improvement of approximately 3.5%. Bagging KNN, SVM, and ANN also show performance increases of 4.5%, 1.3%, and 3.45%, respectively, compared to their standalone counterparts. These findings demonstrate

how bagging leverages the complementary strengths of individual classifiers to enhance overall classification accuracy. The ability of RF to handle complex non-linear relationships and the localized classification insights of KNN contribute significantly to the improved performance of the ensemble model.

In Experiment 4, the KRAFS-ANet framework employs a stacking approach by integrating bagging RF, bagging SVM, and bagging KNN, with bagging ANN as the meta-classifier. This approach aggregates the strengths of individual bagged classifiers, further enhancing classification performance. The results demonstrate superior accuracy across all datasets. MAT [280] achieves the highest accuracy of 98.63% with an F1-score of 98.82%. Similarly, SAM40 [84] and DASPS [30] datasets obtain peak accuracies of 97.25% and 94.92%, with F1-scores of 97.24% and 95.15%, respectively. These findings validate the effectiveness of ensemble stacking and emphasize the importance of data-driven channel selection in improving EEG-based stress classification across different datasets.

4.5.1 KRAFS-ANet Framework: Lightweight and Efficient

The KRAFS-ANet framework incorporates several key strategies to ensure a lightweight and efficient design:

1. **Channel Selection:** The framework applies a systematic channel selection approach to identify the most informative EEG channels. This approach reduces the input data size while preserving high classification accuracy. By optimizing the number of channels, the model lowers computational complexity and minimizes memory usage.

2. **Feature Selection:** The framework employs LASSO feature selection to extract the most relevant features, reducing data dimensionality. This approach decreases computational overhead while enhancing model interpretability and efficiency.

3. **Bagging:** The framework integrates bagging with parallelization, allowing multiple base classifiers to train simultaneously. It reduces training time while maintaining computational efficiency, ensuring the ensemble model remains lightweight despite its complexity.

Table 4.9: KRAFS-ANet framework performance comparison with the base ensemble stacking model.

Model	Dataset	NMI+RFE Channel Selection	No of Channels	Total Features	Selected Features using LASSO	Acc (%)	F1 (%)	Execution Time (sec)	Data Usage (MB)
Ensemble stacking:	MAT [280]	×	20	320	158	92.48	93.32	2394.34	1266.53
Bagging(KNN+RF+SVM)	SAM40 [84]	×	32	512	160	90.59	92.01	3129.94	1881.43
+ANN	DASPS [31]	×	14	224	72	91.24	90.51	190.44	396.66
Proposed KRAFS-ANet	**MAT [280]**	✓	**9**	**144**	**16**	**98.63**	**98.82**	**878.67**	**836.31**
Framework	**SAM40 [84]**	✓	**12**	**192**	**37**	**97.25**	**97.24**	**1370.32**	**1130.15**
	DASPS [31]	✓	**6**	**96**	**17**	**94.92**	**95.15**	**103.28**	**236.07**

Time and Memory Efficiency: The KRAFS-ANet framework is evaluated based on execution time and peak memory usage to assess its lightweight design. These metrics demonstrate the model's efficiency, particularly in resource-constrained environments.

1. **Execution Time:** The total duration required for the model to complete its entire process, measured in seconds.

2. **Peak Memory Usage:** The highest amount of RAM consumed during model execution, measured in megabytes.

Table 4.9 compares the KRAFS-ANet framework, which integrates channel selection and feature selection, with a base ensemble stacking model that does not include channel selection. This comparison evaluates the impact of channel selection on accuracy, F1-score, data usage, and execution time across the MAT [280], SAM40 [84], and DASPS [30] datasets.

For MAT [280], accuracy improves by 6.2%, while execution time decreases by 63.3% and memory usage reduces by 34%. SAM40 [84] achieves a 7.3% accuracy increase, with execution time decreasing by 56.2% and memory usage dropping by 40%. Similarly, in DASPS [30], accuracy improves by 3.68% with NMI+RFE, while execution time decreases by 46.7% and memory usage reduces by 40.4%. These enhancements reduce computational overhead while enhancing accuracy and F1-score across all datasets.

4.5.2 Comparison with Existing Studies

Table 4.10 presents a comparative analysis of the proposed approach against existing studies focusing on channel selection and stacking. In a previous study on MAT [280],

Table 4.10: KRAFS-Anet framework performance comparison with existing studies.

Author	Domain	Dataset	Channel Selection Approach	No of Channels	Feature Domain	FSA	Algorithm	Acc (%)
Sharma et al. [223]	Stress	MAT [280]	Frontal and anterior frontal lobe channels	7	Time-frequency domain	Fisher Score	WOA + SVM	WOA+SVM: 97.25
Malviya et al. [148]	Stress	MAT [280]	—	19	Frequency domain	CNN	CNN+BiLSTM	CNN+BiLSTM: 99.20
Fatimah et al. [75]	Stress	MAT [280]	Analysis on different channels performance	1	Frequency domain	Kruskal-Wallis Test	LR, SVM, KNN	SVM: 98.6, LR: 95.8, KNN: 93.1
Mathur et al. [153]	Stress	SAM40 [84]	—	17	Frequency domain	Kruskal-Wallis Test	DT, KNN, SVM	DT: 87.30, KNN: 92.38, SVM: 93.38
Sim et al. [232]	Stress	SAM40 [84]	—	32	Time-frequency domain (DWT)	PCA, PCA+ICA, LDA	Rusboost, Adaboost	Rustboost: 79.08, Adaboost: 72.42
Hemakom et al. [100]	Stress	Kaggle [100]	—	8	Frequency domain	ANOVA F-value	NB+LR+AB+KNN +RF+SVM+RBF-SVM	SVM(Male):71.57, SVM(Female):62.60, SVM(Mixed): 64.08
Aldayel et al. [15]	Anxiety	DASPS [31]	Manually selected based on affected brain regions	10	Time, Frequency, Time-Frequency domain	—	KNN, Adaboost, GB, RF, LDA, SVM	RF: 87.5
Xie et al. [266]	Emotion	DEAP [132]	—	32	Entropy	Linear Discriminant Analysis	(LGB+XGB+RF)+XGB	Valence -79.06, Arousal -77.19
Proposed Work	**Stress and Anxiety**	**MAT [280], SAM40 [84], DASPS [31]**	**NMI+RFE**	**MAT: 9, SAM40: 12, DASPS: 6**	**Time and Frequency domain**	**Lasso**	**(Bagging RF + Bagging KNN + Bagging SVM) + ANN**	**MAT: 98.63, SAM40: 97.25, DASPS: 94.92**

Sharma et al. [223] selected seven channels from the frontal and anterior frontal lobes and applied the Fisher score feature selection algorithm. Using SVM optimized with the Whale Optimization Algorithm (WOA), the study achieved 97.25% accuracy for an 8-second window. These findings emphasize the frontal region's role in stress classification due to its involvement in emotional and cognitive processes. However, restricting channel selection to the frontal lobe may omit crucial information from other brain regions that contribute to stress responses.

Fatimah et al. [75] [76] conducted two studies on the MAT [280], evaluating single-channel performance and achieving a 98.6% accuracy using LR and QDA. Their findings identified C3 as the most informative channel for stress classification. While their approach demonstrates that high accuracy can be obtained with a simple model, relying on a single channel may overlook the complex and distributed brain activity associated with stress responses.

Malviya et al. [148] applied Discrete Wavelet Transform (DWT) to preprocess 19-channel EEG signals before feeding them into a stacked CNN + BiLSTM model. The CNN extracted relevant features, which the BiLSTM model further processes. Their approach achieved 99.20% accuracy, demonstrating the effectiveness of deep learning with multi-channel data. However, the high computational demands and resource requirements limit its practical application, particularly in wearable technologies that

require efficient real-time processing and user comfort. Mathur et al. [153] and Sim et al. [232] analyzed the SAM40 [84], achieving maximum accuracies of 79.08% using Rustboost and 71.57% with a stacking algorithm, respectively. In previous research, Aldayel et al. [15] analyzed the DASPS [30] and used a manual channel selection approach, selecting 10 channels based on affected brain regions and achieving 87.5% accuracy with Random Forest (RF). However, this method introduces subjectivity and may reduce the model's adaptability to diverse datasets and individual differences.

To the best of our knowledge, no prior research has investigated channel selection combined with ensemble stacking for EEG-based stress classification. This study compares the proposed approach with state-of-the-art models used in emotion classification to evaluate its effectiveness, as presented in Table 4.10. Previous studies by Xie et al. [266] and Aquino et al. [21] applied stacking techniques on the DEAP [132] dataset, utilizing 32 EEG channels. Xie et al. [266] developed an ensemble stacking model incorporating LGB, XGB, and RF for EEG-based emotion classification. Their approach achieved 79.06% accuracy for valence and 77.19% for arousal, employing Linear Discriminant Analysis (LDA) for feature selection. Similarly, Aquino et al. [21] proposed an EEG-based stacking model combining CNN, SVM, and DT, attaining an accuracy of 94.6% using all 32 channels.

In contrast, the KRAFS-ANet framework applies NMI + RFE to reduce the number of channels from 20 to 9 in the MAT [280] while incorporating an effective feature selection technique. This work further validates the framework on SAM40 [84], where 12 channels are selected, achieving an accuracy of 97.25% for stress classification. Although this approach may result in a slight accuracy reduction compared to Malviya et al. [148], it significantly enhances computational efficiency and real-world adaptability. Integrating optimized channel and feature selection with an ensemble stacking approach ensures that KRAFS-ANet surpasses existing stress classification methods in both performance and efficiency.

4.5.3 Limitations

The KRAFS-ANet framework enhances stress classification using EEG signals; however, it has certain limitations:

1. **Signal-Modality:** The framework relies solely on EEG signals, which require a complex setup and may not be practical for daily use. Multimodal wearable systems combining multiple physiological signals can offer a more practical and comprehensive approach to real-world stress assessment.

2. **Feature Interdependency:** While LASSO effectively reduces dimensionality, it may not fully capture interdependencies and interactions between features, which can affect classification performance.

4.6 Summary

The KRAFS-ANet framework incorporates an ensemble stacking model that effectively classifies EEG-based stress levels while maintaining computational efficiency through automated, data-driven channel selection. It enhances prior work by integrating KNN, RF, SVM, and ANN as meta-classifiers, optimizing performance through NMI+RFE-based channel selection and LASSO feature selection. The framework systematically selects the most informative EEG channels, reducing data dimensionality while preserving classification accuracy.

This chapter performs four comprehensive experiments for the KRAFS-ANet framework and validates the proposed methodology on the MAT [280], SAM40 [84], and DASPS [30] datasets. Experiments 1 and 2 assess model performance with and without channel selection, confirming its impact on classification accuracy. Experiments 3 and 4 explore ensemble techniques, demonstrating that ensemble stacking significantly improves classification performance. The KRAFS-ANet model achieves 98.63% accuracy on MAT [280], 97.25% on SAM40 [84], and 94.92% on DASPS [30], outperforming existing stress classification methods. This chapter concludes that effective channel selection and ensemble stacking improve EEG-based stress classification, making the framework scalable, computationally efficient, and well-suited for real-world applications.

This chapter is based on the following work:

- **J2:** **Shikha**, Divyashikha Sethia, and S. Indu. "KRAFS-ANet: A Novel Framework for EEG-based Stress Classification using Channel Selection and Optimized Ensemble Stacking" **International Journal of Machine Learning and Cybernetics** (2024): 1-22. **(SCIE, Impact factor:3.1, Publisher: Springer)**. Doi: `https://doi.org/10.1007/s13042-024-02455-2.` **(Published)**.

Chapter 5

Academic Stress Dataset Collection and Wearable-Based Stress Classification Optimization

This chapter presents the details of data collection for the Academic Stress Dataset (ASD) during the Montreal Imaging Stress Task (MIST) [58] in an academic environment for engineering students. It explores the feasibility of multimodal physiological signals for automated stress classification. This chapter focuses on collecting progressive stress data from engineering students using MIST [58] and also examines the effectiveness of meditation in alleviating stress. A hybrid feature selection approach, combining Genetic Algorithm and Mutual Information (GA+MI), is employed to identify the most significant features. Further, Bayesian optimization fine-tunes machine learning classifiers for automated stress level classification.

5.1 Motivation

Students in academic environments often experience significant stress due to examinations, deadlines, and demanding academic responsibilities. Prolonged exposure to academic stress adversely affects mental and physical health, contributing to anxiety, sleep disturbances, and reduced cognitive performance. Although numerous studies have used physiological signals for stress classification, existing datasets rarely

capture stress as a gradual and progressive condition specific to academic contexts.

To induce stress under controlled conditions, researchers have adopted various experimental tasks, including public speaking [79][152], questionnaires [24], programming contests [42], and games [65]. Among these, MIST [58] is a widely validated protocol that induces stress in a gradually increasing manner through time-constrained arithmetic challenges, making it particularly well-suited for evaluating cognitive load and mental workload in academic environments. Previous studies [276][11] have employed EEG with MIST for stress classification. While EEG offers valuable insights into brain activity, its susceptibility to noise and the discomfort of electrode placement limit its practical applicability. In contrast, wearable devices such as smartwatches offer non-invasive and continuous monitoring of physiological signals, including Electrodermal Activity (EDA), Blood Volume Pulse (BVP), and Inter-Beat Interval (IBI).

However, no existing study has combined the MIST protocol with multimodal physiological signals such as ECG, EDA, and BVP for stress classification [158][266][222][94][41][100]. Furthermore, most publicly available datasets support only binary classification, limiting the ability to capture stress severity and progression. Datasets such as MAT [280], CLAS [150], and SAM40 [84] do not incorporate the MIST task, despite its ability to model stress gradually and realistically. Additionally, despite the distinct academic pressures experienced by engineering students, no dataset has been specifically developed to investigate academic stress using physiological signals in this population.

Contributions:

The key contributions of this chapter are as follows:

1. ***Collecting Multimodal Physiological Academic Stress Dataset (ASD):*** ASD collects data by recording EDA, IBI, and BVP from the Empactica E4 smartwatch. The ASD dataset is collected from students under varying stress levels induced by the MIST, ensuring a progressive representation of stress in an academic environment. Additionally, ASD incorporates meditation phases, showing an analysis of how relaxation techniques impact physiological stress responses.

2. ***Automated Stress classification through Machine Learning:*** A hybrid feature selection approach combining Genetic Algorithm (GA) and Mutual Information (MI) is implemented to extract the most relevant physiological features. Machine learning models are then optimized using Bayesian optimization to enhance the accuracy and efficiency of automated stress classification.

3. ***Explainable AI for Feature Importance Analysis:*** Integrate SHAP (Shapley Additive Explanations) Explainable AI (XAI) to determine the most influential physiological features contributing to stress classification. It enhances model transparency and helps understand the physiological markers most indicative of stress.

4. ***Visualization of Meditation's Impact on Stress Reduction:*** Analyze the effect of meditation on stress alleviation using dynamic visualizations of physiological data. This analysis provides insights into how meditation influences physiological responses, contributing to stress management strategies for students.

5.2 Dataset Collection Procedure

The data collection procedure follows strict ethical guidelines to prioritize participant well-being. The ethics committee approves these guidelines. Before the experiment begins, researchers inform all participants about the stress-inducing nature of the study and obtain informed consent. Each participant completes the Perceived Stress Scale (PSS) [125] questionnaire to evaluate their self-reported stress levels. Participants also retain the right to withdraw from the study at any stage without penalty.

1. Participants: This study collects physiological data from 36 engineering students aged 17 to 25 at Delhi Technological University, India. Among them, 32 are male, and four are female. However, due to missing IBI data caused by movement artifacts, the final dataset includes 30 participants. Before data collection begins, participants are provided with a detailed explanation of the study protocol and its objectives.

Trained personnel closely monitor participants throughout the experiment to ensure their safety and well-being. A debriefing session follows the experiment, allowing participants to express concerns or discomfort experienced during the procedure. The

experimental protocol consists of five distinct phases: Neutral Phase, Rest Phase, Controlled Phase, Experimental Phase, and Recovery Phase.

1. **Neutral Phase:** This phase is a training stage where participants become familiar with MIST. This phase introduces the test interface, control mechanisms, and answer format. The test consists of single-digit answers ranging from 0 to 9. Participants respond using a rotary dial displayed on the screen, which they can control by clicking the left and right mouse buttons.

2. **Rest Phase:** In this phase, participants sit in a comfortable and relaxed position while minimizing movement. A blank screen remains displayed throughout this phase. The primary objective is to establish each participant's baseline physiological state before proceeding with the next stages of the experiment.

3. **Controlled Phase:** In this phase, participants solve arithmetic questions of distinct difficulty levels without any time constraints. They are encouraged to answer as many questions as possible correctly. This phase does not provide feedback. This stage is labeled as "mild stress" in the study.

4. **Experimental Phase:** This phase introduces adaptive arithmetic questions designed to induce increased stress levels through time constraints and social comparison. Each question runs on a timer, and the screen displays the participant's current performance score alongside an artificially inflated average score representing all participants. This artificially set score remains unattainable, fostering a sense of competition and increased pressure. Additionally, the task dynamically adjusts difficulty based on performance. Incorrect or delayed responses lead to an increased time limit for the next question, whereas three consecutive correct answers shorten the time limit for the following question. This phase effectively induces higher stress and is labeled "high stress".

5. **Recovery Phase:** The final phase evaluates physiological changes as participants transition back to a relaxed state. During this phase, participants engage in a stress alleviation task by listening to guided meditation audio while sitting comfortably with closed eyes. This phase is classified as "low stress" in the study.

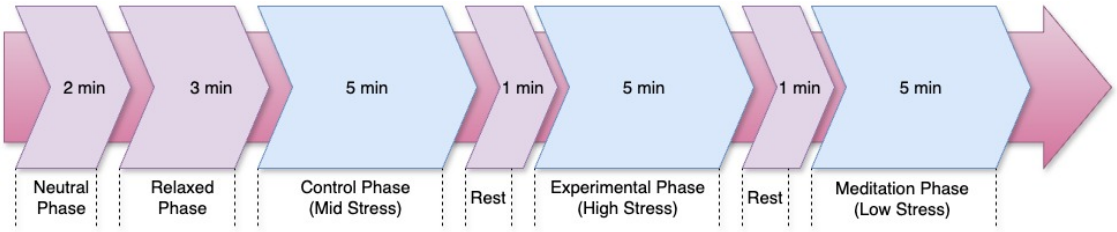


Figure 5.1: ASD acquisition procedure.

The study begins with a 2-minute training phase followed by a 3-minute neutral phase, allowing participants to familiarize themselves with the task. Next, participants complete a 5-minute control and a 5-minute experimental phase, where stress levels progressively increase. The study concludes with a 5-minute recovery phase, during which participants engage in stress alleviation activities. 1-minute rest phases are placed between the control, experimental, and rest phases to ensure proper segmentation, as shown in Figure 5.1.

2. Physiological Signal Monitoring with Empatica E4: The ASD utilizes the Empatica E4 wristwatch to monitor physiological signals throughout the experiment continuously. The device integrates four primary sensors, including a temperature sensor, an accelerometer, EDA sensors, and PPG sensors. The PPG sensor records both IBI and BVP signals. Data is transmitted via Bluetooth to a connected smartphone and monitored in real-time using the E4 Real-time app. After collection, the system uploads the data to Empatica's cloud platform, *E4 Connect*, for further processing. It generates .csv files containing Skin Temperature (ST), BVP, IBI, EDA, HR, and acceleration. The sampling rate is 4 Hz for ST and EDA and 64 Hz for IBI, HR, and BVP.

5.3 Experimental Methodology for Automated Stress Classification

This section describes the methodology illustrated in Figure 5.2. The methodology first performs data preprocessing to improve signal quality and remove potential noise. It then performs feature extraction and selection to identify the most relevant physiological features. Finally, different classifiers process the features and use Bayesian optimization to enhance predictive performance.

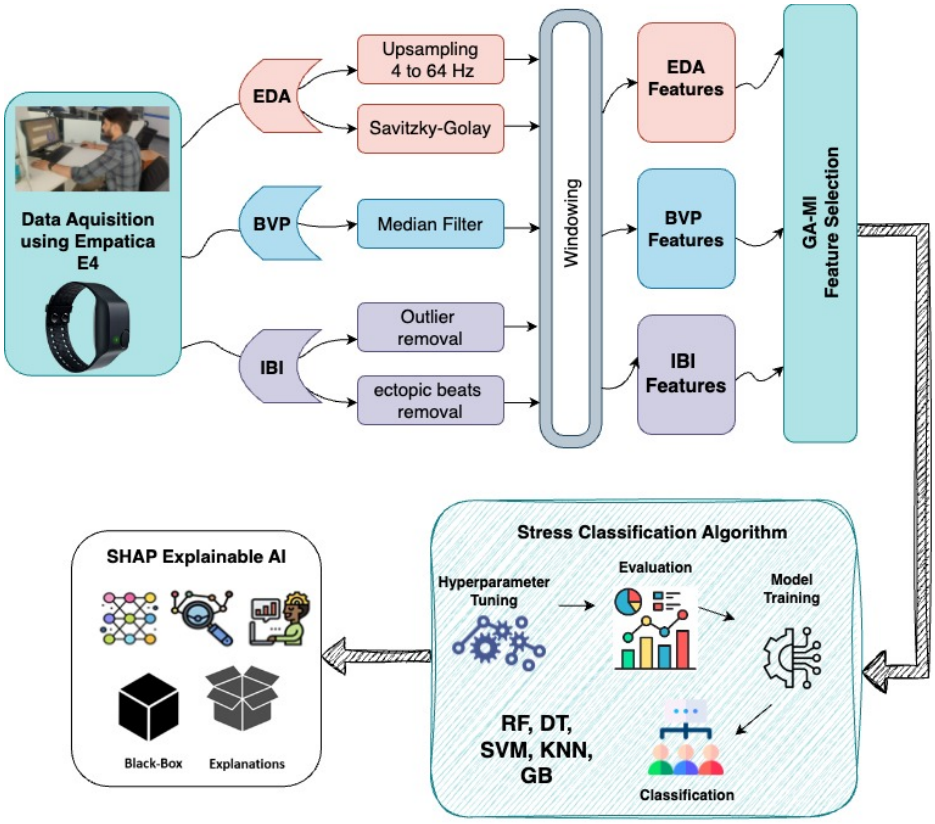


Figure 5.2: Academic Stress Dataset (ASD): Proposed methodology for stress classification in an academic environment.

5.3.1 Data Preprocessing

This work aims to analyze physiological signals recorded by the Empatica E4 and assess the accuracy of the proposed machine learning model. Although data collection occurs in a controlled laboratory environment, signal quality may still be affected by factors such as sensor detachment, outliers, and missing values. Therefore, signal preprocessing is necessary to enhance the signal-to-noise ratio and preserve meaningful information.

Labeling: Before preprocessing, the experiment segments the data based on consecutive sessions. Three-level stress classification labels the control phase as "Mid Stress," the experimental phase as "High Stress," and the recovery phase as "Low Stress."

Signal Windowing: Signal windowing helps extract detailed temporal patterns from continuous physiological data. This experiment applies a ten-second rolling window with a 50% overlap to the wristband data, ensuring maximum information retention

from the recorded signals.

1. Interbeat Interval (IBI): The IBI.csv file contains time intervals between successive heartbeats derived from the PPG sensor. The study preprocesses and analyzes the IBI data using the *hrvanalysis* [1] module in Python, which provides a comprehensive suite of HRV analysis tools. It facilitates preprocessing, R-peak detection, and feature extraction.

Data Cleaning: The IBI data consists of missing values and outliers. This experiment applies interpolation techniques to handle outliers and ensure data continuity. Wearable devices often introduce gaps in the IBI series due to discarded ectopic beats or undetected heartbeats, which can impact HRV analysis accuracy. It applies the *"malik"* method for correcting ectopic beats and improving data reliability to address this issue.

2. Electrodermal Activity (EDA): Among the five recorded features, BVP and IBI have the highest sampling frequency at 64 Hz. Therefore, the study upsamples the EDA signal from 4 Hz to 64 Hz, enabling uniform signal processing to match this frequency. To remove potential artifacts, a Savitzky-Golay filter [41] is applied with a 40-point window and a sigma of 400 ms, smoothing the signal and improving its quality. After preprocessing, only artifact-free segments proceed to the feature extraction stage.

2. Blood Volume Pressure (BVP): The Empatica E4 extracts the BVP signal from the PPG sensor. The experiment applies an HPF to remove low-frequency noise. The filter sets the cut-off frequency between 0.05 and 0.5 Hz to retain only the relevant signal components. Further, it employs a median filter to refine the signal. This filtering technique scans the signal point by point, replacing each value with the median of its neighboring entries, effectively reducing noise while preserving signal integrity.

5.3.2 Feature Extraction

The Empatica E4 [68] device provides IBI data used to extract HRV features. HRV features are computed across time, frequency, and non-linear domains using the *hrvanalysis* Python module. Similarly, Blood Volume Pulse (BVP) data undergoes statistical feature extraction using the *NumPy* [2] module. For EDA processing, the study

[1] `https://pypi.org/project/hrv-analysis/`
[2] `https://pypi.org/project/numpy/`

Table 5.1: Summary of multi-domain features extracted from EDA, IBI, and BVP signals for ASD.

Feature	Domain	Features	Feature ID
EDA	Statistical	var_gsr, mean_gsr, kurtosis_gsr, skew_gsr, std_gsr	1-5
EDA	Tonic	SCL_slope (tonic), var_scl, mean_scl, kurtosis_scl, skew_scl, slope_scl, std_scl	6-12
EDA	Phasic	scr_peaks (phasic), var_scr, mean_scr, kurtosis_scr, skew_scr, max_scr, std_scr	13-19
HRV	Time Domain	median_nni, mean_nni, pnni_50, nni_50, nni_20, pnni_20, hrv_sdnn, hrv_rmssd, hrv_sdsd, range_nni, cvnni, cvsd, max_hr, mean_hr, std_hr, min_hr, tinn, triangular_index, cvi, sd1, csi, sd2, sampen, ratio_sd1_sd2, HR_max-HR_min	20-44
HRV	Frequency Domain	total_power, lf, vlf, hf, lf_hf_ratio, hfnu, lfnu	45-51
BVP	Statistical	Mean BVP, standard deviation, maximum, minimum, power VLF BVP, power LF BVP, power HF BVP	52-58

employs the *PyEDA* [3] package in Python to extract phasic and tonic features from the preprocessed signal. It extracts 58 features from HRV, EDA, and BVP, categorizing them according to their respective domains, as summarized in Table 5.1. Chapter 2 Sections 2.2.2, 2.2.3, and 2.2.5 discuss HRV, EDA, and BVP features in detail.

5.3.3 Feature Selection

5.3.3.1 Genetic Algorithm with Mutual Information (GA-MI)

This work enhances the GA for feature selection by incorporating MI to improve stress classification. Table 5.2 presents the parameters for implementing the GA Feature Selection Approach (FSA). Additionally, Figure 5.3 illustrates the core process of the GA combined with MI for feature selection in a physiology-based stress dataset. The following section details the step-by-step procedure of this method:

1. *Initial Population Generation:* The process begins by defining a population size p, which represents the number of chromosomes in the initial population. Each chromosome c_i is a binary string of length n, where n denotes the total number of features in the physiology-based stress dataset. Each bit c_{ij} within the chromosome corresponds to a specific feature f_j. To introduce diversity, each bit c_{ij} is randomly assigned a value of *0* or *1*, ensuring that the initial population comprises a broad range of potential feature subsets.

[3] `https://pyeda.readthedocs.io/en/latest/`

Table 5.2: Parameters for implementing Genetic Algorithm.

Parameter	Description	Value
Initial Population	Number of chromosomes in the initial population	100
Evaluation Criteria	Metric used to assess chromosome fitness	Mutual Information (MI)
Reproduction Operators	Genetic operators for generating new offspring	Single-point crossover, mutation rate: 0.05
Crossover Rate	Probability of crossover occurring between chromosomes	0.8
Selection Criteria	Method for selecting chromosomes based on fitness	Fitness-proportional selection with elitism
Termination Criteria	Condition for stopping the GA iterations	100 generations or convergence

2. *Compute Mutual Information:* MI quantifies the statistical dependence between two random variables. In this step, MI is used to evaluate the relevance of a given feature set, represented by a chromosome c_i, to the target class y. A higher MI value signifies a stronger relationship between the selected features and the target class.

For a chromosome c_i, let X_{c_i} denote the subset of features selected by c_i, meaning features where the corresponding bits in c_i are set to 1. Equation 5.1 computes MI $MI(c_i, y)$ as follows:

$$MI(c_i, y) = \sum_{x \varepsilon c_i} \sum_{y \varepsilon Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{5.1}$$

where $p(x,y)$ represents the joint probability distribution of feature values $x \in X_{c_i}$ and class labels y, while $p(x)$ and $p(y)$ represent their respective marginal probability distributions. Chromosomes with the highest MI values proceed to the next phase of the algorithm.

3. *Fitness-Proportional Selection:* In this step, the selection is based on fitness proportionality, where the probability $P(c_i)$ of selecting a chromosome c_i depends on its MI value. Chromosomes with higher MI values are more likely to be chosen for reproduction, ensuring that the algorithm prioritizes features with stronger relevance to the target class. Equation 5.2 computes selection probability as follows:

$$P(c_i) = \frac{MI(c_i, y)}{\sum_{j=1}^{p} MI(c_j, y)} \tag{5.2}$$

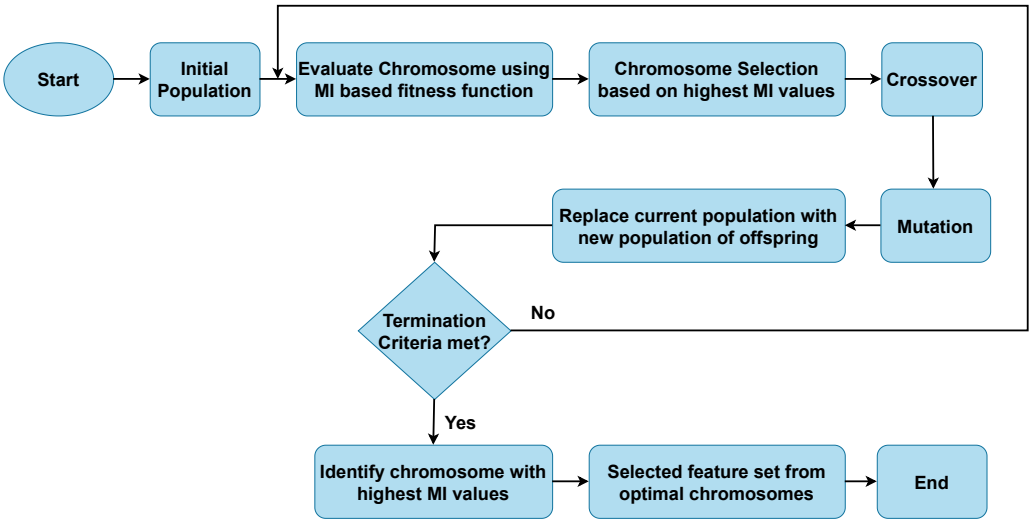


Figure 5.3: GA-MI feature selection algorithm.

where $MI(c_i, y)$ denotes the mutual information of chromosome c_i, and the denominator represents the total mutual information across all chromosomes in the population.

4. *Crossover and Mutation:* The crossover operator simulates genetic recombination by merging genetic material from two selected parent chromosomes to produce new offspring chromosomes. This process facilitates the discovery of better feature combinations by exchanging information between parents. The mutation operator randomly modifies individual bits in the offspring chromosomes. These alterations can introduce new features or eliminate irrelevant ones, enhancing diversity within the population. The mutation prevents the algorithm from becoming trapped in local optima and promotes exploration of a broader solution space.

5. *Iteration and Termination:* The algorithm repeatedly performs Steps 2 to 4, refining the population across multiple generations. The process continues until it meets the termination criterion, defined by the maximum number of generations G_{max}. The value of G_{max} is typically determined through experimental evaluation, considering the problem's complexity and the required convergence level.

6. *Optimal Solution:* The algorithm identifies the optimal feature subset by selecting the chromosome c_{best} with the highest MI value $MI(c_{best}, y)$ in the final generation. This chromosome contains the most relevant features for the target class and aims to enhance classification performance.

5.3.4 Classification Algorithm

Following feature selection, the study splits the dataset into 80% for training and 20% for testing. The study utilizes five classification algorithms to evaluate model performance: Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), XGBoost (XGB), and Gradient Boosting (GB).

5.3.4.1 Bayesian Optimization

The performance of classification algorithms depends significantly on their hyperparameters, which influence model accuracy and generalization. Hyperparameter optimization is essential for fine-tuning these parameters to achieve optimal performance. Traditionally, parameter tuning has relied on trial-and-error experimentation and subjective judgment, making it time-consuming and computationally demanding [131]. However, grid search is a widely used method for hyperparameter tuning. However, it becomes inefficient for models with a large number of hyperparameters due to its high computational cost [105] [19].

Bayesian optimization offers a more systematic and efficient approach to hyperparameter tuning [178]. Studies have shown that Bayesian optimization outperforms traditional methods in global optimization problems, particularly in hyperparameter selection [105][74][245]. Given its advantages, this work employs Bayesian optimization to fine-tune the hyperparameters of classification models for stress classification. This method enhances the search process for optimal hyperparameters, leading to improved model performance and efficiency.

5.4 Experimentation and Results

5.4.1 Preliminary Analysis and Optimization

1. Statistical Feature Analysis: This work employs ANOVA to assess the distinctiveness of the optimal feature subset obtained through GA+MI feature selection. Tables 5.3 present the statistical results for binary and three-level stress classifications.

The analysis in Tables 5.3 identifies multiple features with statistically significant effects (p-value < 0.05) in distinguishing between different stress levels. In Table 5.3a,

Table 5.3: ANOVA test for EDA+IBI+BVP feature subset using GA+MI feature selection algorithm.

(a) 2-Level Stress Classification

Features	F-Value	P-Value
Min HR	2.77E+03	0.00E+00
mean_gsr	1.53E+03	0.00E+00
Mean HR	2.77E+03	0.00E+00
LF/HF	7.45E+00	1.00E-14
RMSSD RR	4.46E+01	3.05E-03
mean_scl	2.77E+03	4.25E-04
slope_scl	6.04E+01	6.36E-03
std_scr	2.04E+01	6.51E-01
Mean BVP	6.79E+00	9.69E-01

(b) 3-Level Stress Classification

Features	F-Value	P-Value
Mean HR	6.73E+04	0.00E+00
Min HR	5.74E+03	0.00E+00
mean_gsr	5.79E+03	0.00E+00
mean_scl	3.59E+03	3.10E-23
Power HF RR	3.62E+03	4.42E-13
LF/HF	3.59E+03	2.10E-04
RMSSD RR	2.11E+03	1.25E-03
Mean BVP	1.25E+03	3.90E-11
max_scr	9.43E+02	3.08E-03
Slope scl	1.25E+03	1.20E-05
STD RR/SDNN	5.23E+01	2.87E-02

features such as Min HR, mean_gsr, and Mean HR consistently demonstrate significance across both binary and three-level classifications. Their high F-values and low p-values further indicate their strong contribution to the model's predictive performance. Additionally, Table 5.3b highlights other significant features, including LF/HF, Power HF RR, and STD RR/SDNN, which also play a crucial role in differentiating stress levels.

2. Performance Measures: The following measures evaluate the performance of machine learning classifiers in stress classification: Accuracy (Acc), F1-score (F1), and Area Under the Curve (AUC). These metrics, derived from the confusion matrix, comprehensively assess the model's effectiveness. A detailed description of these performance measures is available in Chapter 3, Section 3.7. Further, the model undergoes 10-fold cross-validation, where the dataset is divided into ten equal parts, treating each part as a test set in rotation. This approach ensures model reliability and stability.

3. Hyperparameter Tuning: This work employs Bayesian hyperparameter optimization to fine-tune classifier parameters and enhance model performance. Table 5.4 presents the optimal hyperparameters selected for the experiment. A comprehensive experimental process is conducted to ensure the reliability and consistency of

Table 5.4: List of hyperparameters used for ASD stress classification models.

Classifier	Hyperparameter	Value Range	Selected Value
RF	n_estimators	[10, 40, 60, 80, 100, 150, 200]	200
	max_depth	[2, 4, 8, 10, 12, 14]	4
	min_samples_split	[3, 5, 10, 15]	3
	random_state	-	0
DT	max_depth	[2, 4, 8, 10, 12, 14]	8
	min_samples_split	[2, 5, 10, 15, 20, 25]	10
KNN	n_neighbors	[3, 4, 5, 6, 7, 9, 12, 14, 16]	6
	metric	["euclidean", "manhattan", "minkowski"]	"minkowski"
	weights	["uniform", "distance"]	"distance"
SVM	C	[0.1, 1, 10]	1
	kernel	["linear", "rbf"]	"rbf"
	gamma	[0.01, 0.001, 0.0001, scale, auto]	0.1
XGB	n_estimators	[50, 100, 200, 500, 700]	500
	learning_rate	[0.01, 0.001, 0.1, 0.2]	0.1
	max_depth	[2, 4, 8, 10, 15]	8
GB	n_estimators	[100, 200, 500, 700, 900, 1000]	200
	learning_rate	[0.01, 0.001, 0.1, 0.2]	0.1
	max_depth	[2, 4, 8, 10, 12, 14]	4

computational results. This process identifies which hyperparameters should retain their default values and which require optimization through the Bayesian approach.

5.4.2 Classification Results

Previous studies have not explored the integration of ECG and EDA for stress classification in an academic environment during MIST. This work investigates the potential of wearable sensors, specifically IBI, BVP, and EDA biomarkers, to comprehensively assess stress responses.

The proposed stress classification methodology extracts distinctive features from a high-dimensional dataset using the GA+MI FSA. Following feature selection, machine learning classifiers, including RF, DT, SVM, KNN, XGBoost, and GB, are applied to classify stress levels. The experiment employs 10-fold cross-validation and evaluates average performance metrics to ensure reliability and stability.

Table 5.5: ASD: Performance comparison of distinct physiological signals for 2-level stress classification with 58 original feature sets.

Signals	RF			DT			SVM			KNN			XGB			GB		
	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC
EDA	76.4	77.69	84.4	76.32	75.61	78.3	55.7	55.69	60.2	60.15	59.57	69.67	78.93	77.91	87.3	**85.12**	85.64	94.63
IBI	83.84	83.62	89.42	80.37	79.11	83.65	78.72	78.88	86.52	80.5	80.33	83.98	82.92	81.4	86.83	**86.54**	86	93.02
BVP	83.91	83.92	88.54	78.61	78.75	78.37	54.7	55.03	64.29	78.83	78.73	86.66	82.88	83.97	88.29	**85.53**	85.91	91.4
EDA+IBI	93.35	92.54	97.03	90.22	90.47	95.03	83.96	83.5	87.1	83.5	81.45	87.4	92.94	92.98	96.7	**94.97**	95.35	97.86
IBI+BVP	89.76	89.67	95.02	86.22	87.14	92.64	79.71	80.78	84.98	83.88	83.75	89.75	90.00	90.1	95.69	**92.02**	92.56	96.94
EDA+BVP	88.3	88.13	93.76	87.37	87.35	91.54	62.15	62.03	67.7	80.91	82.55	87.32	88.21	88.16	94.72	**91.93**	91.91	97.38
EDA+IBI+BVP	93.37	92.43	97.83	90.54	90.87	94.15	84.54	83.21	89.26	84.59	85.63	90.41	92.23	91.67	97.36	**95.67**	95.73	98.59

Table 5.6: ASD: Performance comparison of distinct physiological signals for 2-level stress classification with selected 9 features using GA-MI FSA.

Signals	RF			DT			SVM			KNN			XGB			GB		
	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC
EDA	81.89	80.6	88.86	78.3	76.98	81.92	55.93	54.07	60.58	69.85	66.8	72.81	80.27	80.69	87.71	**88.61**	88.5	94.02
IBI	85.1	85.32	92.76	78.8	78.2	80.55	80.7	80.73	89.05	82.8	82.79	90.93	84.41	84.67	90.11	**90.86**	90.71	97.07
BVP	85.2	85.18	92.77	79.00	79.48	82.24	55.7	55.68	64.96	82.9	82.91	89.48	85.12	85.3	92.78	**87.3**	88.9	93.94
EDA+IBI	94.66	94.76	98.82	92.98	92.94	97.32	84.04	84.14	89.83	90.21	91.24	96.34	95.3	95.2	98.27	**97.07**	97.05	99.03
IBI+BVP	91.87	91.67	96.53	88.07	88.34	94.84	82.63	82.68	88.03	85.15	86.94	90.04	91.86	91.87	97.18	**94.83**	95.26	98.21
EDA+BVP	89.21	89.8	94.44	87.37	88.66	93.23	64.13	64.25	70.2	82.91	83.28	88.74	90.15	90.15	96.03	**93.42**	93.98	97.72
EDA+IBI+BVP	95.5	95.08	99.83	93.84	93.71	97.53	84.54	84	90.44	91.15	91.89	95.53	95.66	95.65	98.64	**98.28**	97.44	99.78

Table 5.7: ASD: Performance comparison of distinct physiological signals for 3-level stress classification with 58 original feature sets.

Signals	RF			DT			SVM			KNN			XGB			GB		
	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC
EDA	68.39	67.87	73.8	64.69	65.34	72.56	52.57	52.89	59.65	61.85	63.73	68.23	73.17	73.12	79.11	**75.01**	75.04	80.23
IBI	76.52	77.55	84.53	73.48	74.10	80.79	72.36	72.93	77.78	74.56	74.40	80.24	77.08	77.23	84.17	**79.74**	80.12	87.31
BVP	74.42	74.53	81.4	69.61	69.73	75.68	59.37	60.85	68.74	70.38	71.37	78.74	76.02	76.47	82.01	**79.70**	79.20	86.09
EDA+IBI	91.91	92.90	97.09	89.49	89.41	95.00	75.04	75.78	82.32	70.28	70.90	78.78	91.49	91.47	96.43	**93.87**	93.91	97.74
IBI+BVP	89.51	89.90	95.93	83.52	83.89	89.31	73.66	73.44	79.09	71.82	71.90	78.23	89.54	89.61	94.57	**91.92**	91.93	97.01
EDA+BVP	85.42	85.38	92.61	78.29	78.3	84.27	59.51	59.67	67.78	64.97	64.43	72.34	86.81	86.83	92.17	**89.43**	89.66	95.28
EDA+IBI+BVP	93.26	94.76	97.15	91.54	92.33	97.7	79.17	79.8	85.45	80.53	81.92	87.78	93.29	93.2	97.64	**95.16**	96.2	98.8

Table 5.8: ASD: Performance comparison of distinct physiological signals for 3-level stress classification with selected 11 features using GA-MI FSA.

Signals	RF			DT			SVM			KNN			XGB			GB		
	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC
EDA	71.6	70.32	77.98	67.07	68.34	73.3	52.52	52.78	59.78	67.13	67.42	72.26	75.17	75.27	81.56	**79.23**	79.98	87.76
IBI	79.93	79.27	89.48	75.88	75.56	84.84	75.26	75.08	83.1	77.3	77.63	86.04	80.04	80.41	87.72	**82.93**	82.91	89.48
BVP	77.15	77.13	83.17	70.83	70.85	77.66	64.31	64.18	72.67	72.23	73.01	80.34	78.63	78.85	85.3	**82.45**	83.46	88.11
EDA+IBI	93.12	94.23	98.21	91.76	91.98	97.78	76.68	76.23	83.32	73.07	73.87	79.77	93.16	93.28	97.16	**95.23**	96.05	98.47
IBI+BVP	90.43	90.35	96.95	84.98	84.65	92.02	75.11	75.74	81.54	73.65	73.25	79.67	90.06	90.4	95.82	**92.78**	92.24	97.63
EDA+BVP	88.48	88.35	93.7	80.62	80.57	86.45	62.98	62.62	68.96	77.77	77.92	84.21	89.03	89.02	95.6	**91.45**	91.42	96.42
EDA+IBI+BVP	95.65	96.56	99.52	93.95	94.83	98.4	80.61	80.67	87.67	89.16	90.57	95.42	95.66	95.75	98.13	**97.02**	96.53	99.13

Tables 5.5, 5.6, 5.7, and 5.8 present the experimental results for 2-level and 3-level stress classification, both with and without feature selection. These results offer critical insights into the effectiveness of the models in accurately distinguishing stress levels.

For 2-level stress classification, analyzing individual physiological signals reveals significant findings. Table 5.5 shows that EDA and BVP features perform similarly without feature selection. Notably, IBI-derived HRV features achieve the highest classification accuracy of 86.54%, highlighting their strong predictive capability.

Table 5.6 presents the classification results after applying the GA+MI FSA. The findings indicate an accuracy improvement of 3.49% for EDA and 1.77% for BVP. Additionally, IBI-derived HRV features show substantial performance gains, with accuracy increasing by 4.32% and the F1-score improving by 4.71% following feature selection. These results highlight the effectiveness of the GA+MI FSA in enhancing stress classification accuracy.

Further, the feature combination of EDA and HRV achieves an accuracy of 94.97% without feature selection, which increases to 97.07% after applying the GA+MI FSA. However, incorporating all three features, including EDA, HRV, and BVP, results in a modest accuracy improvement of 2.68%. Additionally, the study observes that RF, DT, and KNN exhibit similar performance, while XGB and GB consistently outperform other classifiers, both with and without GA+MI feature selection.

Tables 5.7 and 5.8 present the classification results before and after applying feature selection, demonstrating improved performance in three-level stress classification. This classification task adds complexity by distinguishing between low, medium, and high stress levels. While individual EDA and IBI signals yield moderate performance, their combination outperforms feature sets that include BVP. However, when using the GB classifier, incorporating EDA, IBI, and BVP achieves the highest accuracy of 97.02%, F1-score of 96.53%, and AUC score of 97.47%, indicating its effectiveness in multi-level stress classification.

The results demonstrate that applying feature selection, specifically the GA+MI FSA, significantly improves classification accuracy. This improvement underscores the importance of removing redundant features, which can negatively affect model performance. Figure 5.4 compares the test accuracy between the original feature set (F_o) and the selected feature set (F_s) for both 2-level and 3-level stress classification, highlighting the effectiveness of feature selection in enhancing model efficiency.

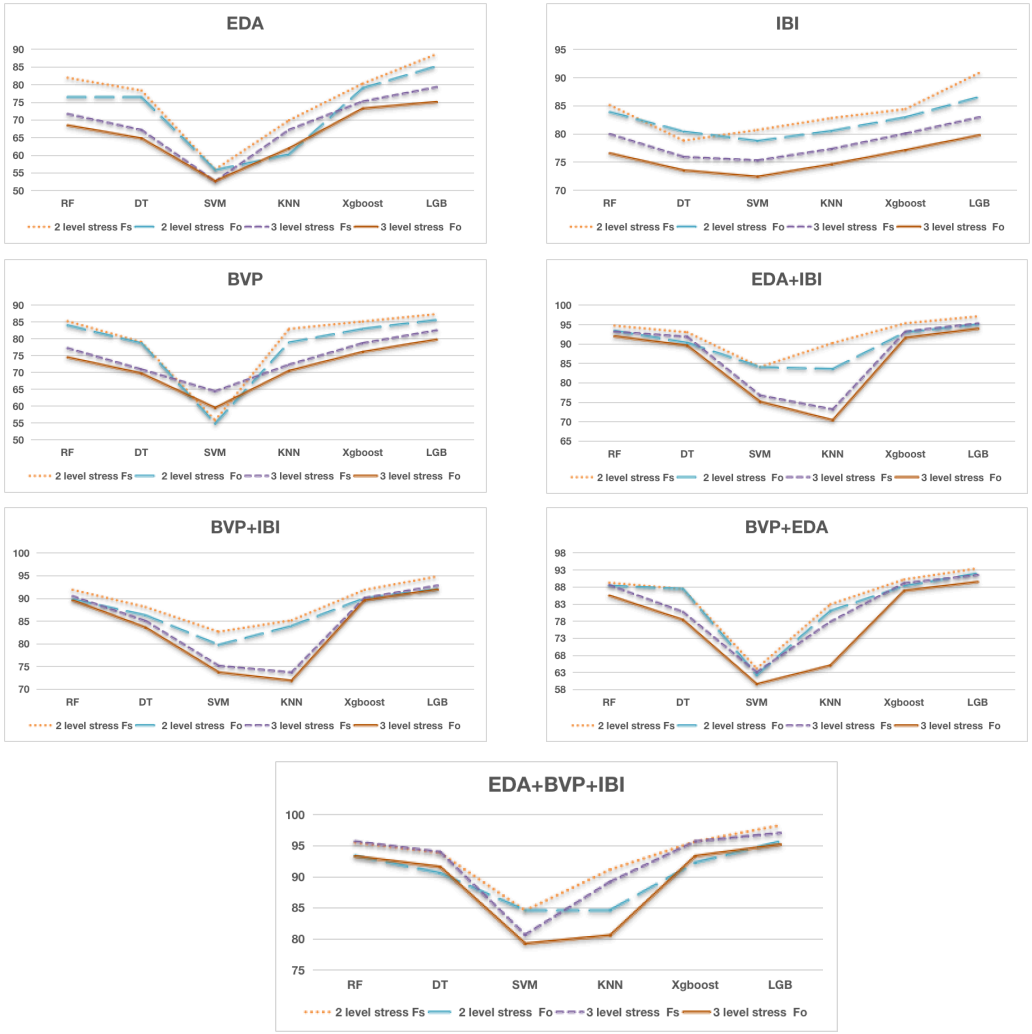


Figure 5.4: Comparison of test accuracy of different classification models for 2-level and 3-level stress classification for the original feature set (F_o) and the selected feature set (F_s).

5.4.3 Explainable AI

Figure 5.5a and 5.5c display bar plots that visualize SHAP (Shapley Additive Explanations) values, highlighting the feature importance for 2-level and 3-level stress classification, respectively. Each feature's mean absolute SHAP value across all samples determines its importance. Additionally, Figure 5.5b and 5.5d present SHAP summary plots, showcasing the top 9 features for 2-level stress classification and the top 11 features for 3-level stress classification based on the entire stress dataset. These visualizations provide insights into how each physiological feature contributes to the classification of stress levels.

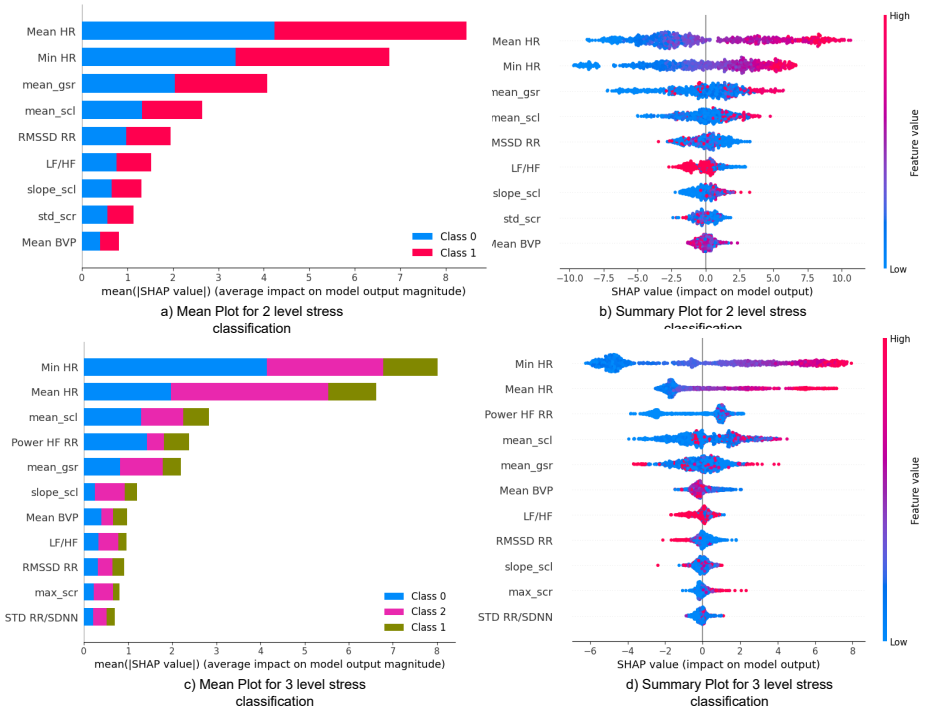


Figure 5.5: Explanation of SHAP values for ASD features.

The summary plots in Figure 5.5b and Figure 5.5d illustrate the impact and contribution of individual features in the stress classification model. The horizontal axis represents the SHAP values, indicating their influence on high or low-stress predictions. The vertical axis above zero suggests minimal impact on the classification outcome. A SHAP value of zero signifies no substantial effect on predictions, while values closer to zero indicate features with lower influence. Conversely, higher SHAP values, whether positive or negative, highlight features with strong correlations to the predicted stress levels. Each row in the plots corresponds to a specific feature, ranked based on their mean absolute SHAP values. Additionally, A stacked dot plot visualizes the distribution of SHAP values, with colors representing feature intensity. Red denotes high feature values, while blue indicates lower feature values. These visualizations provide insights into the relative importance of physiological features in distinguishing stress levels.

Furthermore, Figure 5.5 highlights mean HR, Min HR, and mean_gsr as the most influential features in stress classification. This finding aligns with the ANOVA statistical analysis results presented in Table 5.3a and Table 5.3b. The low p-values and high F-scores in the ANOVA results further confirm the significance of these features in predicting stress levels.

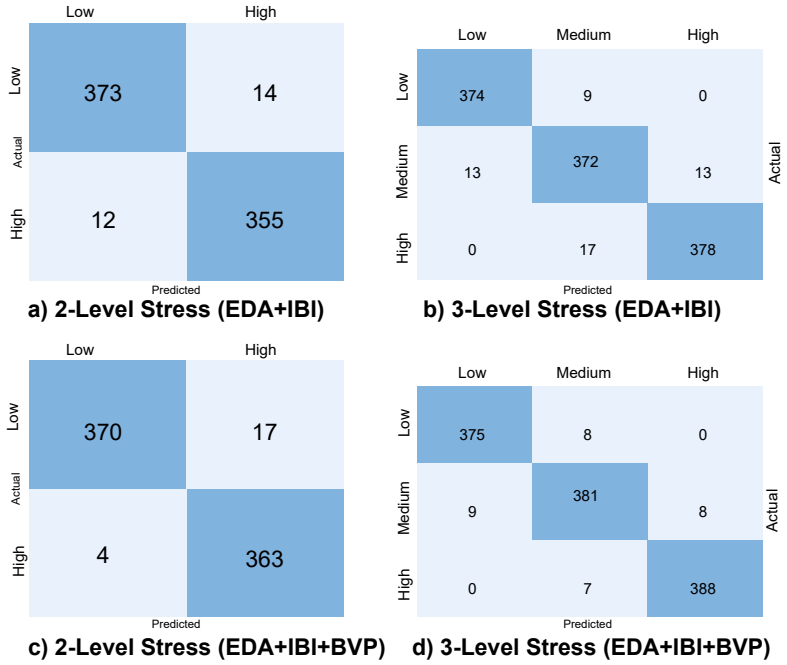


Figure 5.6: Confusion Matrix of the best-evaluated results for integrated signals using Gradient Boosting.

5.5 Discussion

The experimental findings highlight the effectiveness of EDA and IBI as key physiological indicators for stress classification, particularly when utilizing the GB classifier, and combining EDA and IBI results in an impressive 97.07%

Extending stress classification to three levels with BVP features further enhances the model's performance, achieving an accuracy of 97.02%

The findings emphasize the effectiveness of EDA and IBI in differentiating stress from non-stress states. The minimal accuracy improvement after incorporating BVP features suggests that IBI, derived from BVP, may introduce redundant information, contributing only a slight enhancement. This outcome reinforces the idea that combining EDA and IBI provides a comprehensive representation of physiological responses to stress, ensuring strong and reliable performance in stress classification.

The study also reveals that GB outperforms XGBoost in stress classification. Although XGBoost is highly effective for high-dimensional datasets with numerous features, its performance may be less optimized when applied to datasets with a limited

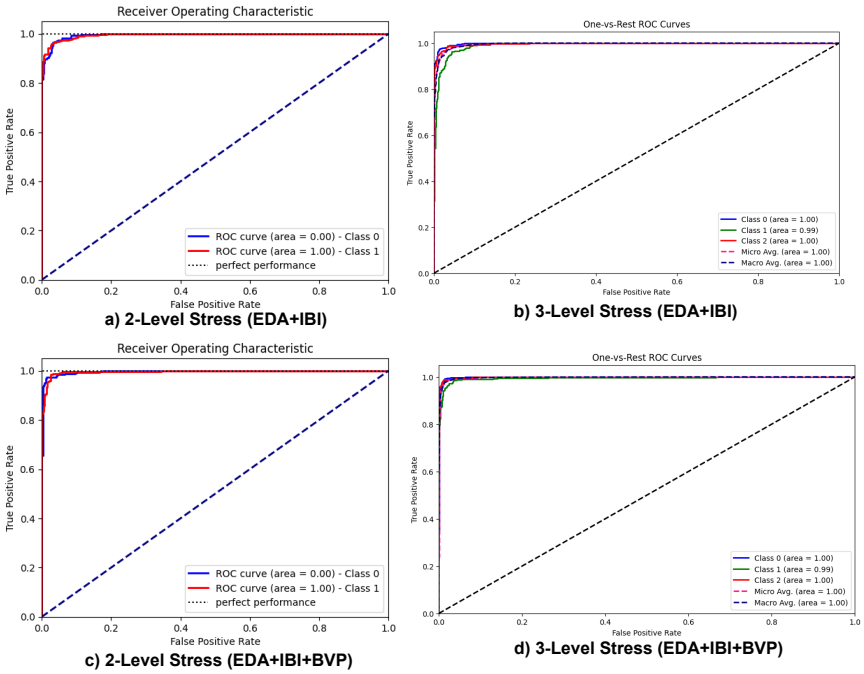


Figure 5.7: ROC of the best-evaluated results for integrated signals using Gradient Boosting.

feature set, as observed in this study. Since XGBoost relies on a broader range of features for optimal performance, its effectiveness may decrease when feature availability is restricted. In contrast, GB, less dependent on feature quantity, demonstrates superior classification performance, making it a more suitable approach for this experiment.

The ASD protocol comprises two key stages: stress classification and stress alleviation. During the stress recovery phase, participants listened to meditation audio, and the findings suggest that this intervention effectively contributes to stress reduction. The experimental results highlight IBI and EDA as key contributors to stress classification. Figure 5.8 presents the FFT spectrum, generated using Kubios analysis, to depict different stress levels. The FFT analysis examines the HRV signal derived from IBI. HRV measures the time variation between consecutive heartbeats, reflecting the activity of the ANS. The ANS regulates essential physiological functions such as heart rate, blood pressure, and respiration, influencing stress responses.

Figure 5.8 a) and 5.8 b) depict a distinct peak in the low-frequency (LF) range, reflecting Sympathetic Nervous System (SNS) activity during moderate and high-stress conditions. Additionally, a smaller peak in the high-frequency (HF) range represents Parasympathetic Nervous System (PNS) activity, indicating that despite stress, some

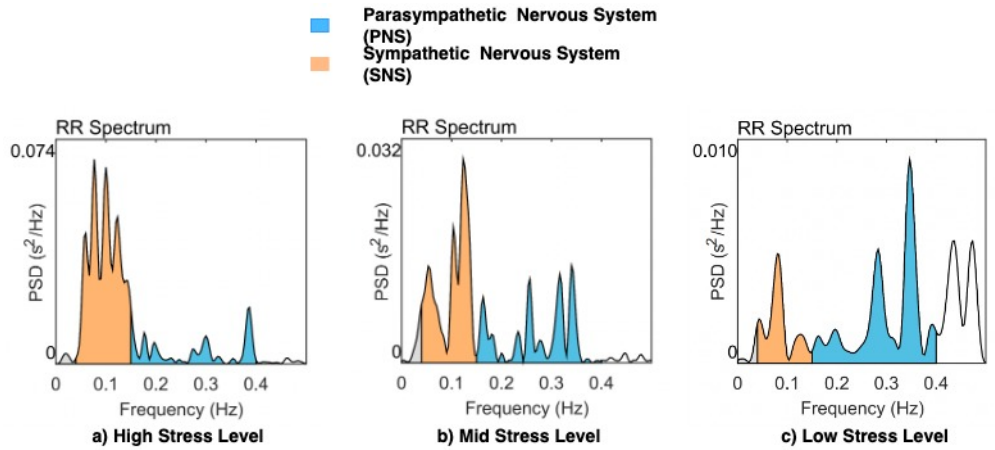


Figure 5.8: Fast Fourier Transform Spectrum of distinct stress levels derived from IBI data and visualized using Kubios software: a) High-stress level, b) Mid-stress level, c) Low-stress level.

level of parasympathetic function persists, potentially mitigating stress effects. In contrast, Figure 5.8 c) illustrates a pronounced peak in the HF range during low-stress conditions, signifying heightened PNS activity associated with relaxation and stress recovery.

Figure 5.9 presents the EDA graph, illustrating the analysis of phasic and tonic components. Multiple factors, including stress, arousal, and perspiration, influence skin conductance. Phasic EDA represents rapid fluctuations in skin conductance triggered by specific stimuli or events, typically lasting for a short duration and associated with emotional arousal. In contrast, tonic EDA reflects gradual and sustained variations in skin conductance over time, indicating the overall level of arousal and physiological activation.

Figure 5.9 EDA graph identifies various stress states by examining their phasic and tonic components. High-stress conditions show increased phasic EDA responses and elevated tonic EDA levels. Moderate stress exhibits moderate phasic EDA responses and a moderate tonic EDA level. The figure illustrates a progressive decline in phasic and tonic EDA levels during low-stress conditions, with a gradual reduction in EDA peaks.

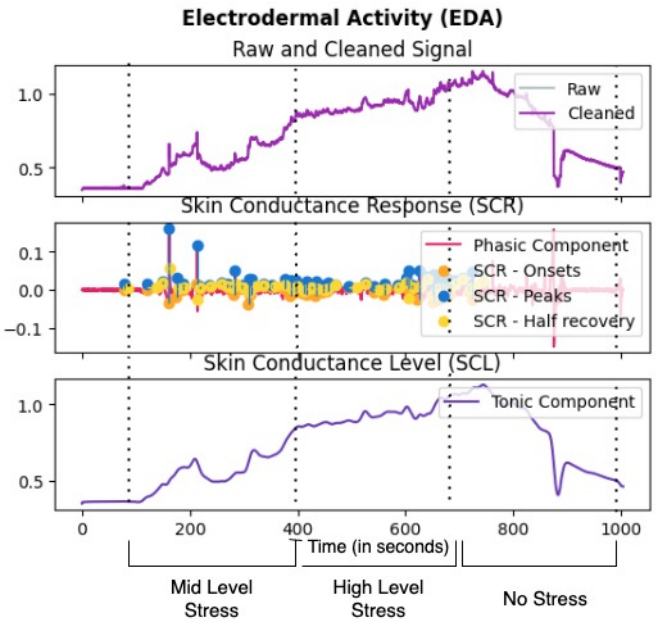


Figure 5.9: EDA graph to analyze phasic and tonic components of EDA during stress classification.

5.5.1 Comparison with Existing Studies

The ASD consists of both stress classification and stress alleviation phases. The stress classification model attains an accuracy of 98.28% for binary stress classification and 97.02% for three-level stress classification. During the stress alleviation phase, participants engage in meditation audio sessions. Additionally, the study utilizes Explainable AI (XAI) to determine the most influential physiological features contributing to stress classification. This work examines the role of EDA and IBI-derived HRV in stress classification, focusing on how stress and stress alleviation influence the sympathetic and parasympathetic branches of the ANS. The results demonstrate that meditation audio effectively reduces stress in an academic environment.

Table 5.9 presents a comparative analysis of previous MIST-based studies incorporating stress alleviation techniques. It outlines key aspects, including physiological signals, number of channels, participant details, stress alleviation methods, experiment duration, selected features, and classification performance. Limited studies have explored stress alleviation within the MIST framework. Zhang et al. [276] classified stress levels using FD features and applied the Fisher Ratio for feature selection. The study evaluates multiple stress relief techniques, including

Table 5.9: Performance comparison of the proposed ASD with existing MIST-based stress alleviation studies in the literature.

Author	Signals	Channels	Sub	Stress Alleviation Technique	Experiment Duration	Feature Selection	Features	Classifiers	Accuracy	F1-score	Other Metrics
Zhang et al. [276]	EEG	4	25	Watching a video, listening to alpha music, squeezing a stress ball, bubble wrap popping, and hugging a pillow.	16 min approx	Fisher Ratio	4	LDA	2 level: 85.6%, 3 level: 71.5%	–	–
Xia et al. [265]	EEG, ECG	128	22	Sitting in a relaxed position and focusing on a circle appearing on a computer screen	80 min	FDR, mRMR, Information Gain	–	SVM	79.45%	–	Recall: 81%, Specificity: 78%
Setz et al. [222]	EDA	–	32	Questionnaire and reading magazines	50 min	Not mentioned	16	LDA, SVM, NCC	82.80%	–	–
Han et al. [94]	ECG, RSP	–	39	Sitting in a relaxed position	18 minutes (excluding stress report filling time)	OOB	36	SVM, KNN, LDA, Adaboost	84%	83%	Recall: 84, Precision: 83
Proposed Work	ECG, EDA	–	30	Listening to meditation audio	22 min	GA + MI	9	RF, DT, SVM, KNN, GB, XGB	2 levels: 98.28%, 3 levels: 97.02%	2 levels: 97.44%, 3 levels: 96.53%	AUC: 99.78, AUC: 99.13

listening to alpha music, watching nature videos, popping bubble wrap, hugging a pillow, and squeezing a stress ball. The findings indicated that hugging a pillow is the most effective stress reduction method, as it demonstrates a clear reduction in stress levels compared to baseline measures. The study achieves an accuracy of 85.6% for binary stress classification and 71.5% for three-level stress classification.

Xia et al. [265] employed EEG and ECG signals for stress classification, utilizing Fisher's Discriminant Ratio, Minimum Redundancy, Maximum Relevance (mRmR), and Information Gain for feature selection. Their approach achieved a maximum accuracy of 79.5%. Further, participants follow instructions to sit in a relaxed position and focus on a circle displayed on a computer screen to alleviate stress induced by MIST. However, the study did not specify the final selected features used for classification.

Similarly, Han et al. [94] investigated ECG and RESP signals for stress classification and identified sitting in a relaxed position as a stress-relief technique. However, this method may be influenced by environmental factors, as external distractions or discomfort in the selected setting could reduce its effectiveness. Additionally, this approach is primarily a short-term stress alleviation method, as some individuals may struggle to recall stressful situations or experience genuine stress responses when attempting to relax.

Setz et al. [222] introduced a general questionnaire and reading magazines as methods for stress relief. However, the effectiveness of reading magazines as a stress reduction strategy may differ among individuals, as engagement levels vary based on personal preferences. Additionally, the passive nature of this activity may not provide an equally effective stress-relief mechanism for all participants. Moreover, these studies

lack a quantitative assessment of the proposed stress alleviation techniques, leaving their overall effectiveness unverified.

5.5.2 Limitations

This research collects the ASD dataset from 36 engineering students. However, noise and artifacts in physiological signals led to the exclusion of six participants, resulting in a final dataset of 30 students. It highlights the artifact and noise removal challenge in wearable devices, which is crucial for ensuring reliable data quality in real-time stress classification applications. Additionally, while the GA+MI feature selection method effectively reduces feature redundancy, some selected features may still be correlated. Future studies can explore advanced feature selection techniques to enhance interpretability and reduce redundancy. Moreover, this research primarily focuses on academic stress for engineering students. In the future, researchers can extend this approach to investigate automated workplace stress and anxiety classification, or sports performance stress, which enlarges the scope of physiological signals-based stress classification.

5.6 Summary

This chapter outlines the data collection methodology designed to assess automated stress classification for engineering students using wearable physiological sensors in an academic environment. The experiment collects EDA, IBI, and BVP signals from 30 engineering students. It uses MIST to induce progressive stress, followed by a stress alleviation phase, where participants listen to audio meditation to assess stress reduction. The experimental protocol is structured into five phases: Neutral, Rest, Control, Experimental, and Recovery, ensuring a controlled and progressive stress-inducing framework. The collected signals undergo preprocessing, including artifact removal, signal normalization, and resampling, to enhance data quality for further analysis. The study labels stress into two and three levels and extracts statistical, time-domain, and frequency-domain features for classification. The findings highlight that IBI-derived HRV and EDA are highly effective biomarkers for stress classification.

Furthermore, the stress alleviation phase demonstrates that listening to meditation audio effectively reduces stress, supporting its use as a practical intervention for students. The GA+MI hybrid feature selection approach combined with Bayesian-optimized machine learning classifiers improves classification performance. The results show a 2.61% improvement in two-level stress classification and a 1.86% improvement in three-level stress classification compared to models without feature selection.

In conclusion, this chapter emphasizes the potential of wearable technology for real-time automated stress classification in academic environments. The results prove that HRV and EDA signals can enhance stress classification accuracy, while meditation-based interventions offer an effective stress alleviation strategy. Future research can explore advanced feature selection techniques to optimize classification accuracy further and expand the application of this methodology to other domains, including workplace stress, healthcare, and sports performance stress.

This chapter is based on the following work:

- **J1: Shikha**, Divyashikha Sethia, and S. Indu. "Optimization of Wearable Biosensor Data for Stress Classification Using Machine Learning and Explainable AI" **IEEE Access** (2024): 1-17. **(SCIE, Impact factor: 3.4, Publisher: IEEE)**. Doi: `https://doi.org/10.1109/ACCESS.2024.3463742`. **(Published)**.

Chapter 6

Efficient Feature Selection Algorithm for Stress and Anxiety Classification

High-dimensional data from physiological signals such as EDA, HRV, and RESP often leads to overfitting, increased computation time, and reduced generalizability. This chapter proposes two efficient feature selection algorithms, the Correlation-Logistic Mutual Information Feature Selection Algorithm (CorLMI-FSA) and the Correlated Interactive Reinforcement Learning-based Feature Selection Algorithm (CIRL-FSA), to address these limitations for improving stress and anxiety classification using physiological signals. 1) CorLMI-FSA enhances mutual information feature selection by incorporating correlation-based preselection and logistic redundancy control. 2) CIRL-FSA formulates feature selection as a dynamic decision-making problem using interactive reinforcement learning to enable adaptive and context-aware feature selection strategies. Additionally, CIRL-FSA is further extended to STCIRL-FSA by integrating the SMOTETomek resampling approach to manage class imbalance in the Spiderphobic dataset [109]. The chapter also presents a cross-dataset evaluation to assess the generalizability of both algorithms across stress and anxiety domains.

6.1 Motivation

Physiological signals such as EDA, HRV, and RESP offer valuable information for stress and anxiety classification. These signals generate high-dimensional feature

spaces across time, frequency, and non-linear domains [272]. The resulting feature sets often contain redundancy and irrelevant features, which degrade classification performance and increase computational cost. These issues are problematic in real-time and resource-constrained systems, where model efficiency and interpretability are critical.

Feature selection is a crucial step in the machine learning pipeline to address the high dimensionality problem. Existing methods include filter-based algorithms like Mutual Information (MI) [197], statistical approaches such as ANOVA [244][277], and wrapper-based methods [259][24]. However, most traditional approaches rely on static ranking criteria or ignore feature dependencies, limiting their adaptability across datasets.

Contributions:

The key contributions of this chapter are as follows:

1. **Development of CorLMI-FSA:** Propose an improved filter-based approach that uses correlation-based preselection followed by a logistic redundancy control over mutual information scores to select highly relevant and non-redundant features.

2. **Development of STCIRL-FSA :** Formulate feature selection as an interactive reinforcement learning process to enable dynamic decision-making. Further, extend CIRL-FSA to STCIRL-FSA by incorporating SMOTETomek resampling to improve performance on Spiderphobic imbalanced dataset [109].

3. **Cross-Dataset Generalizability Evaluation:** Evaluate CorLMI-FSA and CIRL-FSA on both the Academic Stress Dataset (ASD) and the Spiderphobic anxiety dataset [109]. Apply each algorithm to both domains to test robustness and adaptability.

The following sections describe each algorithm in detail, followed by a comparative analysis highlighting their relative performance, strengths, and potential for real-world stress and anxiety classification systems.

6.2 Correlation-Logistic Mutual Information Feature Selection Algorithm (CorLMI-FSA)

Traditional feature selection methods struggle to balance relevance and redundancy, especially in complex physiological datasets where features often exhibit non-linear dependencies. Among these methods, MI is widely used to quantify the dependency between features and target variables [249]. However, it fails to effectively manage redundancy, which becomes more pronounced as the number of features increases. In physiological signal features, redundancy grows nonlinearly and limits the classification performance. Addressing this requires a feature selection approach that dynamically adjusts redundancy weighting to retain only the most informative features.

This section introduces an efficient Correlation-Logistic Mutual Information Feature Selection Algorithm (CorLMI-FSA). The novel feature selection algorithm optimizes the trade-off between relevance and redundancy. It incorporates a logistic function to adaptively weight redundancy, ensuring that selected features maximize classification performance. This work validates CorLMI-FSA using HRV and EDA signals recorded during the MIST in the self-collected ASD dataset, as proposed in Chapter 5.

Following are the main contributions of this section:

1. ***Developing CorLMI-FSA for Stress Classification:*** Proposes a novel feature selection algorithm, Correlation-Logistic Mutual Information Feature Selection Algorithm (CorLMI-FSA), for stress classification using EDA and IBI-derived HRV signals.

2. ***Comprehensive Evaluation with Machine Learning Models:*** Evaluates performance of CorLMI-FSA using machine learning models, including SVM, KNN, RF, GB, and EBM.

3. ***Feature Importance Analysis with Explainable Boosting Machine (EBM):*** Employ EBM to identify the most influential physiological features, providing interpretability in stress classification.

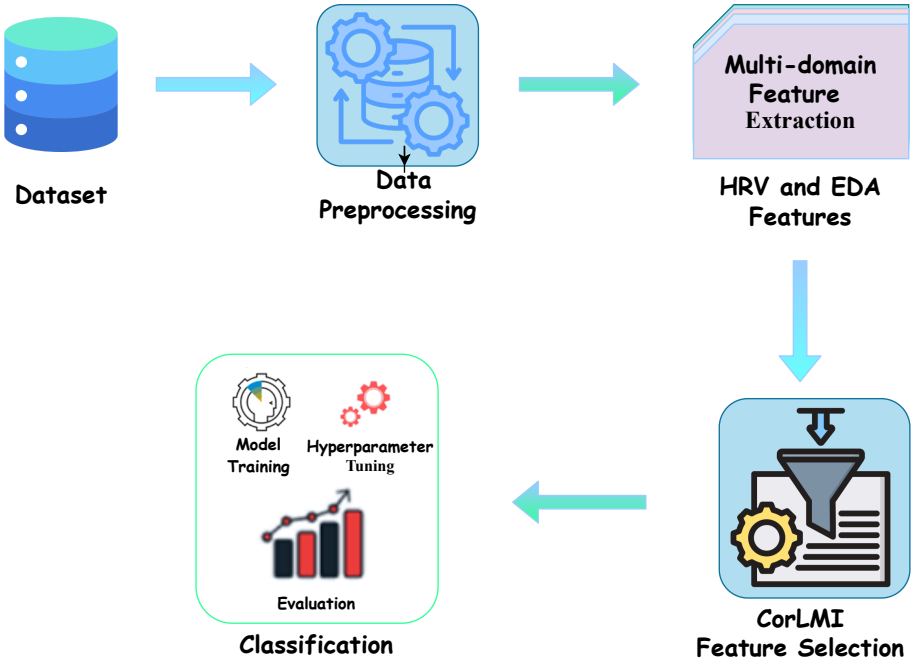


Figure 6.1: CorLMI-FSA proposed methodology for stress classification in an academic environment.

6.2.1 Experimental Methodology

This section describes the methodology as illustrated in Figure 6.1. It begins with data acquisition, outlining the process of collecting physiological signals. Data preprocessing is followed to enhance signal quality and eliminate noise. Next, feature extraction and selection identify the most relevant physiological features for stress classification. Finally, the selected features are input into various machine learning classifiers, which undergo further optimization using Bayesian optimization to improve predictive performance.

6.2.1.1 Dataset Description

This work utilizes the Academic Stress Dataset (ASD) collected in Chapter 5. Previous findings in Chapter 5 concluded that EDA and IBI-derived HRV are the most informative signals for stress classification. Therefore, this chapter focuses on extracting features from these signals. Section 5.1.1 presents detailed information about the dataset and collection process.

6.2.1.2 Data Preprocessing

This section utilizes the preprocessed dataset from Chapter 5. IBI data undergoes cleaning and processing using the hrvanalysis Python module, where interpolation methods correct outliers and the "malik" method handles ectopic beats. EDA signals are upsampled from 4 Hz to 64 Hz to match the IBI signal's sampling rate. Section 5.1.2 provides a detailed description of data preprocessing and labeling.

6.2.1.3 Feature Extraction

The experiment extracts HRV features from IBI data recorded by the Empatica E4 device. These features span multiple domains, including time, frequency, and non-linear measures, and are computed using the *hrvanalysis* Python module. Additionally, the study processes EDA signals using the *PyEDA* package to extract phasic and tonic component features. It derives a total of 58 features from HRV and EDA signals. Section 5.1.3 provides a detailed description of each feature category based on its respective domain.

6.2.1.4 CorLMI - Feature Selection Algorithm

Feature selection enhances stress classification by reducing dimensionality and improving computational efficiency. Mutual Information, a widely used filter-based method, evaluates feature relevance based on its statistical dependency on the target variable. However, it struggles to balance relevance and redundancy in high-dimensional physiological datasets, often retaining features with overlapping information. To address this limitation, this work proposes *Correlation-Logistic Mutual Information Feature Selection Algorithm* (**CorLMI-FSA**), which optimizes feature selection by dynamically adjusting redundancy weighting:

1. *Correlation-Based Pre-Selection:* The experiment first applies a correlation-based feature selection algorithm to remove highly correlated features. This process eliminates 20 features, ensuring that the remaining 31 provide unique and complementary information. By reducing redundancy, this pre-selection step refines the candidate feature set for further evaluation.

2. *Enhanced Mutual Information FSA:* The proposed CorLMI-FSA enhances the traditional mutual information FSA to overcome the limitations of traditional MI-based algorithms. Traditional approaches assume that feature redundancy increases linearly as more features are selected. However, this assumption does not hold for complex physiological datasets, where features often exhibit non-linear dependencies. CorLMI-FSA introduces a logistic function-based redundancy coefficient, which dynamically adjusts redundancy weighting as new features are added to the selected subset. This adaptive approach ensures that only the most informative and non-redundant features contribute to stress classification. Equation 6.1 defines the mutual information (MI) between a candidate feature X_k and the target variable Y:

$$I(X_k;\, Y) = H(X_k) - H(X_k \mid Y) \tag{6.1}$$

where $H(X_k)$ represents the entropy of the feature, and $H(X_k \mid Y)$ denotes its conditional entropy given the target class Y.

To incorporate a redundancy control mechanism, CorLMI-FSA introduces a logistic coefficient β, as defined in Equation 6.2:

$$\beta = \frac{1}{1 + e^{-\frac{\mathsf{S}}{\mathsf{F}}}} \tag{6.2}$$

where S represents the number of already selected features, and F denotes the total number of candidate features. This coefficient gradually increases as more features are selected, allowing a controlled balance between feature relevance and redundancy. Equation 6.3 presents the scoring function used by CorLMI-FSA to rank and select features:

$$J(X_k) = I(X_k;\, Y) - \beta \sum_{X_j \in S} I(X_j;\, X_k) \tag{6.3}$$

The CorLMI-FSA algorithm selects 10 features for the two-level stress classification and 11 features for the three-level stress classification, ensuring optimal feature

representation for each classification task.

6.2.1.5 Classification Algorithms

This work evaluates the performance of CorLMI-FSA using five machine-learning classifiers: SVM, KNN, RF, GBM, and EBM. The dataset is split into 80% training and 20% testing, ensuring a balanced evaluation. The experiment applies 10-fold cross-validation, which trains and tests the models on multiple subsets of the data to improve reliability.

Further, performance evaluation relies on accuracy (Acc) and F1-score (F1) to assess classification effectiveness. Additionally, the Receiver Operating Characteristic (ROC) curve provides insights into the models' ability to distinguish between stress levels. Chapter 3, Section 3.7 presents a detailed description of these performance measures. The study applies Bayesian optimization for hyperparameter tuning to optimize predictive performance and ensure efficient parameter selection while maintaining computational efficiency. Section 5.1.5 provides a detailed explanation of the optimization process.

6.2.2 Results and Discussion

Multi-domain features extracted from signals such as IBI and EDA often include redundant and irrelevant information, which can obscure critical patterns necessary for accurate stress classification. Identifying the most relevant features is essential to reduce dimensionality, enhance model performance, and improve interpretability. However, traditional feature selection methods struggle with high inter-feature correlation and non-linear dependencies in physiological data. These limitations often result in suboptimal feature subsets that may reduce classification accuracy.

This experiment utilizes the self-collected ASD introduced in Chapter 5. It focuses on IBI-derived HRV and EDA features, which effectively capture physiological responses to stress, as demonstrated in the previous chapter. After extracting features, the study applies the proposed CorLMI-FSA algorithm to manage redundancy. To assess classification performance, the study evaluates multiple machine learning models, including SVM, KNN, RF, GB, and EBM.

Table 6.1: Performance comparison of various classifiers for two-level stress classification with and without the proposed CorLMI-FSA.

	Before FSA (58)		**CorLMI-FSA (10)**	
	Accuracy	**F1 score**	**Accuracy**	**F1 score**
SVC	83.96	83.50	87.61	87.17
KNN	83.5	81.45	87.02	86.06
RF	93.35	92.54	94.39	94.39
GBM	**94.97**	95.35	**96.82**	96.81
EBM	90.23	89.41	93.43	93.43

Table 6.2: Performance comparison of various classifiers for three-level stress classification with and without the proposed CorLMI-FSA.

	CorLMI-FSA (11)		**Before FSA (58)**	
	Accuracy	**F1 score**	**Accuracy**	**F1-score**
SVC	81.93	81.22	75.04	75.78
KNN	76.9	75.11	70.28	70.9
RF	92.43	92.59	91.91	92.9
GBM	**95.84**	95.12	**93.87**	93.91
EBM	90.32	90.45	86.03	86.63

Table 6.1 presents the classification performance of machine learning models before and after applying the CorLMI-FSA algorithm for two-level stress classification. The algorithm reduces the original 58 features to 10, retaining only the most relevant ones. This reduction enhances classification accuracy across all models. SVM improves from 83.96% to 87.61%, while KNN increases from 83.5% to 87.02%, reflecting an accuracy gain of approximately 4% in both models. RF achieves the highest accuracy of 94.39%, while EBM shows significant improvement, reaching 93.43% accuracy. Notably, GB outperforms all models with an accuracy of 96.82% and an F1-score of 96.81%.

Table 6.2 presents the classification performance for three-level stress classification after applying the CorLMI-FSA algorithm. The algorithm reduces the original feature set to 11 features, enhancing model performance across all classifiers. SVM improves by 6.89%, while KNN shows a 6.62% increase. RF achieves an accuracy of 94.39%, while EBM improves significantly from 90.23% to 93.43%. GB demonstrates the highest performance after feature selection, achieving an accuracy of 95.84% and an F1-score of 95.12%. Figure 6.2 illustrates the ROC curves for different classifiers after applying feature selection for both two-level and three-level stress classification. The

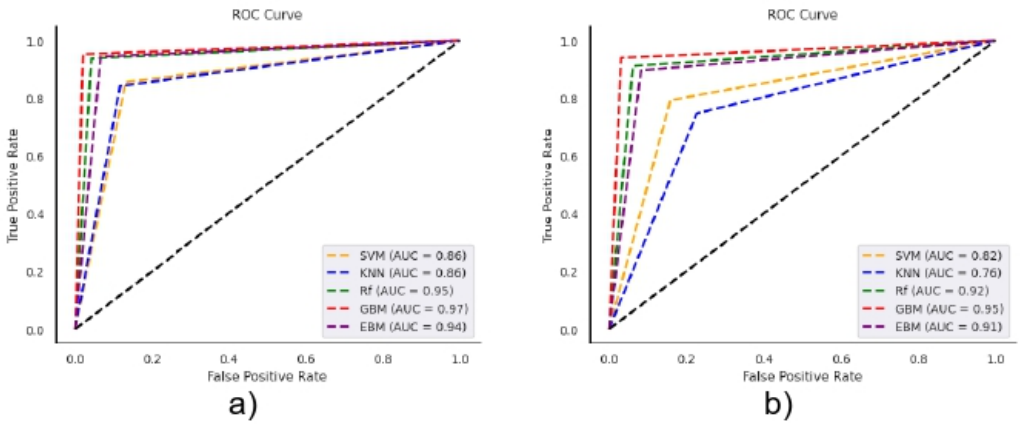


Figure 6.2: ROC curve for the distinct classifiers after CorLMI-FSA. a) Two-level stress classification, a) Three-level stress classification.

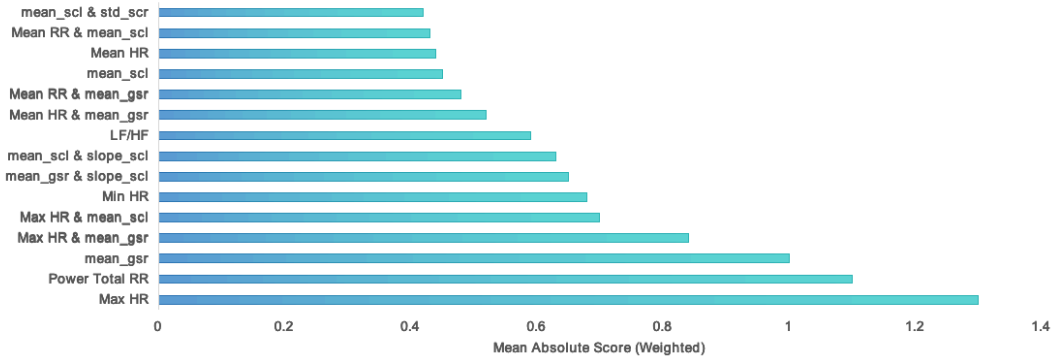


Figure 6.3: CorLMI-FSA: Feature importance plot generated by EBM for two-level stress classification, illustrating the importance of individual physiological features and notable feature interactions.

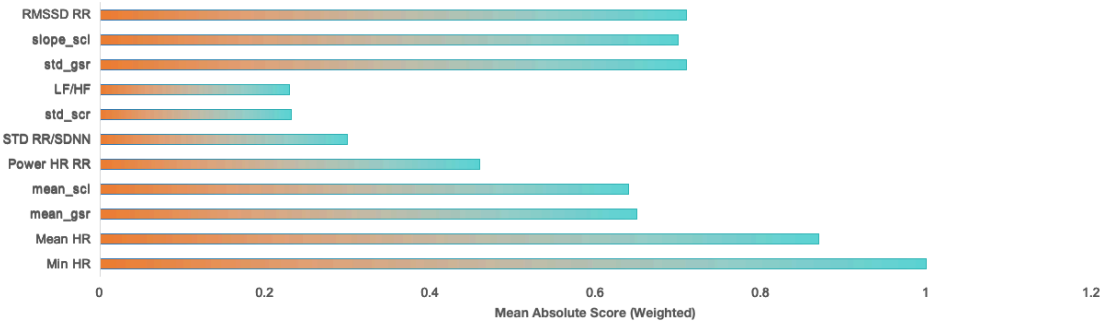


Figure 6.4: CorLMI-FSA: Feature importance plot generated by EBM for three-level stress classification, illustrating the importance of individual physiological features for stress classification.

curves highlight the models' ability to distinguish between stress levels, demonstrating

Table 6.3: CorLMI-FSA performance comparison with existing stress classification studies based on feature selection algorithm.

Author	Dataset	Signal	Subjects	Original Features	Feature Selection Algorithm (FSA)	Selected Features	Algorithms	Accuracy (%)
Nath et al. [176]	Private	EDA, BVP, Salivary cortisol	19	39	ANOVA F-value	10	RF, KNN, LR, SVM, LSTM	LSTM: 93 (macro F1-score)
Can et al. [42]	Private	HR, EDA, ACC	21	–	PCA	–	LDA, SVM, KNN, LR, RF, MLP	PCA + MLP: 92.15
Castaldo et al. [46]	Private	HRV	42	23	Wilcoxon signed-rank test + Spearman's correlation	6	SVM, MLP, IBK, C4.5, LDA	IBK: 94
Dalmeida et al. [51]	SRAD [99]	ECG, HRV	16+4	22	Pearson correlation	7	NB, KNN, SVM, MLP, RF, GB	GB: 85
Radhika et al. [190]	CLAS [150]	ECG, EDA	59	63	RFE	50	SVM, CNN	SVM: 88.9
Radhika et al. [191]	CLAS [150]	ECG, EDA	59	63	RFE	50	SVM (poly), CNN	SVM: 84.7
Proposed Work	ASD	IBI, EDA	30	58	CorLMI-FSA		SVM, KNN, RF, GB, EBM	GB: 96.82%(2 levels), GB: 95.84%(3 levels)

improved classification performance following the application of CorLMI-FSA.

The experiment applies EBM to identify the most influential features for two-level and three-level stress classification. Figure 6.3 shows that *Max HR*, *Power Total RR*, and *mean_gsr* contribute the most to classification accuracy for the two-level stress classification task. Additionally, the study observes notable feature interactions, such as *Max HR & mean_gsr* and *Max HR & mean_scl*, demonstrating how combining specific features enhances predictive performance. These results reinforce the importance of selecting optimal features to improve classification accuracy and model interpretability.

Figure 6.4 shows that EBM identifies *Min HR*, *Mean HR*, and *mean_gsr* as the most influential features for three-level stress classification. The results indicate that these individual features differentiate between low, mid, and high stress levels. Unlike the two-level classification, this analysis does not reveal strong feature interactions, suggesting that single features provide sufficient distinction in the multi-class setting. These findings highlight the importance of selecting features that independently contribute to stress classification accuracy.

6.2.2.1 Comparison with Existing Studies

Table 6.3 compares the proposed CorLMI-FSA with existing stress classification models to assess improvements in feature selection efficiency and classification accuracy.

In one study, Nath et al. [176] used EDA, BVP, and salivary cortisol for stress classification in older adults. The authors selected 10 statistical features using ANOVA F-value and achieved a macro F1-score of 93% with LSTM. However, the method does not explicitly reduce redundancy, which may affect performance on high-dimensional datasets.

Can et al. [42] employed PCA to reduce feature dimension for stress classification using Heart Rate (HR), EDA, and accelerometer (ACC) data. The authors achieved the highest accuracy of 92.15% with MLP. However, the PCA linear transformation may fail to capture complex feature dependencies, limiting its effectiveness in multi-class classification. Castaldo et al. [46] showed that ultra-short-term HRV can be used for stress classification. The authors achieved the highest accuracy of 94% with IBK. Furthermore, Dalmeida et al. [51] extracted ECG-derived HRV features and applied Pearson correlation for feature selection. The authors reduced the feature set from 23 to 6. The model achieved an accuracy of 95% using GB and RF. Radhika et al. [190] applied RFE with SVM and CNN for binary stress classification and achieved 88.9% accuracy with SVM. In another study, Radhika et al. [191] achieved the highest accuracy of 84.7 using SVM and RFE FSA.

In contrast, the proposed CorLMI-FSA reduces the feature set from 58 to 10 for two-level classification and from 58 to 11 for three-level classification. Further, it achieves accuracy improvement from 94.97% to 96.82% using 10 features for 2-level stress classification. Similarly, it achieves accuracy improvement from 93.87% to 95.84% using 11 features for 3-level stress classification. The results demonstrate performance comparable to that of existing studies. Unlike previous methods, CorLMI-FSA dynamically adjusts redundancy weighting using a logistic function, ensuring efficient feature selection without high computational costs. Additionally, Explainable Boosting Machine (EBM) improves interpretability, making the framework well-suited for real-time stress monitoring in wearable applications.

6.2.3 Significant Outcome

The CorLMI-FSA algorithm successfully reduces the high dimensionality of physiological signal features by identifying the most informative and non-redundant subset. It demonstrates that incorporating logistic control over mutual information enables better handling of redundancy, which often limits the performance of traditional filter-based methods. However, CorLMI-FSA does not account for complex feature interdependencies and lacks adaptability to dataset-specific variations. Therefore, the following section introduces a interactive reinforcement learning-based approach that supports adaptive and context-aware feature selection to address these limitations.

6.3 Correlated Interactive Reinforcement Learning based Feature Selection Algorithm (CIRL-FSA)

While the CorLMI-FSA algorithm effectively reduces feature redundancy and improves classification performance, it operates as a static filter-based approach. It does not capture complex dependencies among features or adapts to dataset-specific characteristics. Therefore, this section proposes a dynamic and context-aware feature selection method based on interactive reinforcement learning to address these limitations.

This section proposes the Correlated Interactive Reinforcement Learning Feature Selection Algorithm (CIRL-FSA). It extends the proposed CIRL-FSA by incorporating SMOTETomek resampling to address class imbalance present in the Spiderphobic anxiety dataset [109]. CIRL-FSA initially applies the Pearson Correlation Coefficient (PCC) to eliminate highly correlated features, reducing redundancy in the initial feature space. Then, interactive reinforcement learning is employed to iteratively select the most informative subset of features based on performance feedback. The algorithm processes multimodal physiological signals, including ECG, EDA, and RESP, to capture richer and more discriminative anxiety patterns.

Following are the main contributions of this section:

1. ***Designing the Correlated Interactive Reinforcement Learning Feature Selection Algorithm (CIRL-FSA):*** proposes a novel two-stage feature selection

approach that combines correlation-based filtering with interactive reinforcement learning. The algorithm uses meta-descriptive statistics to represent the state space during learning to enable dynamic and context-aware selection of the most informative features based on reward feedback.

2. ***Extending CIRL-FSA with SMOTETomek for Imbalanced Data (STCIRL-FSA):*** Addresses class imbalance in the Spiderphobic anxiety dataset [109]. This work extends CIRL-FSA by incorporating SMOTE and Tomek-link resampling techniques, resulting in the STCIRL-FSA.

3. ***Hyperparameter Optimization Using Optuna:*** Utilizes Optuna framework to optimize ML classifiers. This technique systematically fine-tunes hyperparameters to improve classification accuracy and computational efficiency.

4. ***Comparing CIRL-FSA with Standard Feature Selection Algorithms:*** Compares the performance of CIRL-FSA with distinct existing feature selection techniques.

6.3.1 Experimental Methodology

This section presents the methodology for the proposed STCIRL-FSA, as depicted in Fig. 6.5. The study designs and evaluates experiments using two distinct approaches. The first follows a traditional machine learning framework, while the second implements the enhanced STCIRL-FSA. This advanced approach integrates SMOTETomek for data balancing, interactive reinforcement learning for feature selection, and Optuna for hyperparameter optimization.

6.3.1.1 Dataset Description

This work employs the publicly available Spiderphobic dataset [109]. This dataset consists of physiological recordings from 57 participants aged 18 to 40, captured using the BITalino device with ECG, EDA, and RESP sensors. Participants watch 16 Spiderphobic video clips across two sessions, followed by a five-minute resting phase.

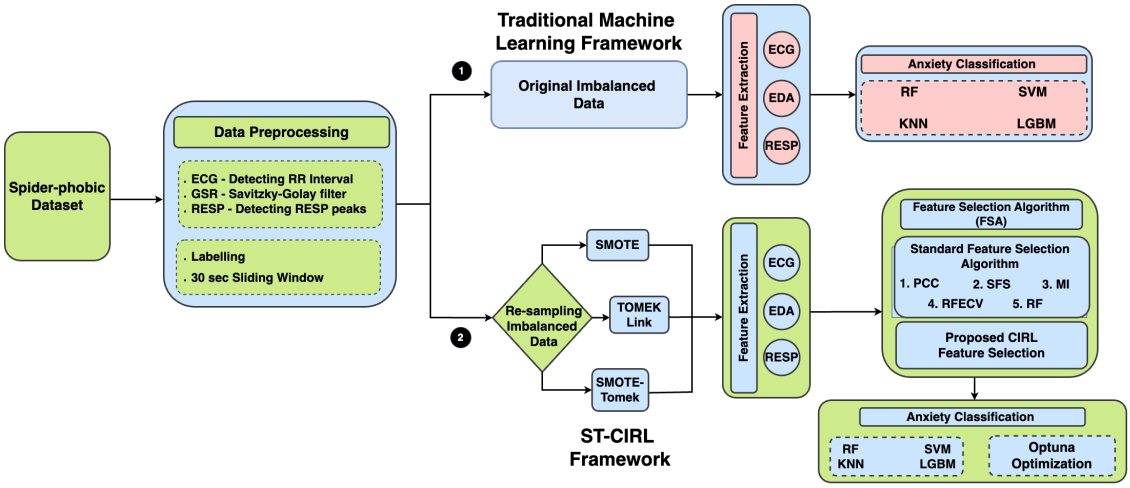


Figure 6.5: Proposed STCIRL-FSA for anxiety classification using physiological signals.

Table 6.4: Summary of Spiderphobic anxiety dataset [109] for STCIRL-FSA.

Dataset	Subjects	Task	Duration	Signals	Device	Sampling Frequency
Spiderphobic dataset [109]	57	Participants watch spiderphobic videos	Rest: 5 minutes Task: 16 videos × 1 minute	ECG, EDA, RESP	Recorded using BITalino device	100 Hz

6.3.1.2 Data Preprocessing

This work employs the same preprocessing pipeline described in Chapter 5 for ECG and EDA. Further, it preprocesses RESP signals with the *NeuroKit2* library[1], which provides filtering and feature extraction techniques for RESP data.

Labeling and Windowing: The Triggers.txt file in the dataset provides identifiers and timestamps that indicate the start and end of each video clip and resting phase. These timestamps align with the Signal folders to ensure precise labeling of segments as "anxiety" for video clips and "non-anxiety" for resting phases. Since no subjective survey data are publicly available, video clips are labeled as "anxiety," while resting phases are labeled as "non-anxiety." Further, the experiment employs a 30-second sliding window for signal feature extraction to overcome the difficulty of detecting RESP peaks in shorter windows.

[1] `https://neuropsychology.github.io/NeuroKit/functions/rsp.html`

6.3.1.3 Imbalanced Dataset Issue

The Spiderphobic dataset [109] used in this work exhibits a significant class imbalance, with 76% of samples labeled as "anxiety" and 24% as "rest." This uneven distribution can bias the classification model and may favor the majority class and misclassify underrepresented instances. Addressing this issue is essential to improve model generalizability and performance.

Previous studies have attempted to mitigate class imbalance using various labeling strategies. Ihmig et al. [109] categorized video clips based on physiological responses, where the authors label the eight clips with the highest mean HR and EDA as "high" anxiety, the remaining eight as "medium" anxiety, and the entire five-minute resting phase as "low" anxiety. Despite these efforts, the dataset remained imbalanced, with 38% classified as high, 38% as medium, and 24% as low. To improve classification performance, the authors applied feature selection using the SFS wrapper approach and achieved a maximum accuracy of 89.8% with bagged trees. Similarly, Gazi et al. [80] labeled the top five clips with the highest HR means as "anxiety" and assigned all resting data as "rest," implementing a 300-second window with 299 seconds of overlap. The classification approach using an RF model achieved an accuracy of 88%. Although these studies introduced feature selection and classification techniques, they did not fully resolve the class imbalance issue.

SMOTETomek Resampling Technique

This work applies SMOTETomek, a resampling method that integrates both over-sampling and under-sampling to create a more balanced dataset. Chawla et al. [49] introduced the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance. This advanced over-sampling method generates synthetic samples for the minority class using a k-nearest neighbor approach. Instead of duplicating existing data, it creates new instances within the feature space and minimizes the risk of overfitting [195]. Equation 6.4 presents the formula for generating synthetic samples.

$$S_{syn} = r(S_{kNN} - S_f) + S_f \tag{6.4}$$

Here, S_f represents the feature sample, S_{kNN} refers to the k-nearest neighbor, and r is a randomly generated value between 0 and 1. The classifier utilizes these synthetic

Table 6.5: Multi-domain features derived from EDA, IBI, and RESP signals for STCIRL-FSA.

Signal	Feature Domain	Feature Description	Features	Feature ID
EDA	Statistical	Captures central tendencies and variability, providing important information about overall arousal levels and autonomic activity.	var_eda, mean_eda, kurtosis_eda, skew_eda, std_eda	1-5
	Tonic (Skin Conductance Level)	Represent slowly varying baseline characteristics of the EDA signal, capturing continuous variations in skin conductance levels and reflecting an individual's overall arousal state and tonic sympathetic activity.	SCL_slope (tonic), var_scl, mean_scl, kurtosis_scl, skew_scl, slope_scl, std_scl	6-12
	Phasic (Skin Conductance Response)	Captures rapid and short-term fluctuations in EDA signals, represents momentary responses to stimuli or events, and reflects the phasic sympathetic activity.	scr_peaks (phasic), var_scr, mean_scr, kurtosis_scr, skew_scr, max_scr, std_scr	13-19
ECG	HRV-Time Domain	Statistical measures derived from the time intervals between consecutive heartbeats (RR intervals) in an HRV analysis.	median_nni, mean_nni, pnni_50, nni_50, nni_20, pnni_20, hrv_sdnn, hrv_rmssd, hrv_sdsd, range_nni, cvnni, cvsd, max_hr, mean_hr, std_hr, min_hr, tinn, triangular_index, cvi, sd1, csi, sd2, sampen, ratio_sd1_sd2, HR_max-HR_min	20-44
	HRV-Frequency Domain	Captures the power distribution across different frequency bands, representing the relative contributions of sympathetic and parasympathetic activities in heart rate regulation.	total_power, lf, vlf, hf, lf_hf_ratio, hfnu, lfnu	45-51
RESP	RRV Features	Numerical values that describe the RESP signal and its changes over time and frequency.	RRV_RMSSD, MeanBB, SDBB, SDSD, CVSD, MedianBB, MCVBB, HF, SD1, SD2, SD2/SD1, Apen	52-63

samples to define distinct regions for classification.

In contrast, Tomek-link operates on two distinct classes, typically a majority and a minority class. It identifies a link, known as the Tomek link, which represents the distance between these two classes, denoted as $d(y_a, y_b)$. A Tomek link is valid if no other class, y_z, is closer to either class than they are to each other. This condition holds if neither $d(y_a, y_z) < d(y_a, y_b)$ and $d(y_b, y_z) < d(y_a, y_b)$ is satisfied [183]. The Tomek-link method removes the majority class instances closest to the minority class by applying the nearest neighbor rule to refine class separation [212].

6.3.1.4 Feature Extraction

This section outlines the feature extraction process. Table 6.5 provides a detailed list of 63 features derived from ECG, EDA, and RESP signals. These features span multiple domains, including statistical analysis, time-domain, and frequency-domain methods.

This work extracts HRV features from ECG signals following the approach described in Chapter 5, incorporating time-domain, frequency-domain, and non-linear analyses to capture heart rate fluctuations. Similarly, the work extracts EDA features, including Skin Conductance Level (SCL) and Skin Conductance Responses (SCR), using the *PyEDA* package in Python, following the methodology presented in Chapter 5. For RRV feature extraction, it applies the *$rsp_rrv()* function from the *NeuroKit2* Python module.

6.3.1.5 Feature Selection

Feature selection plays a vital role in machine learning by enhancing model performance and reducing computational complexity by identifying the most relevant features from a dataset [241]. Traditional feature selection methods are categorized into filter [140][130][176][244], wrapper [109][24][165], and embedded [207][137] approaches. However, these algorithms often evaluate features independently or in conjunction with specific models, which may lead to overlooked feature interactions or overfitting to predefined assumptions [142]. Recent advancements have investigated the application of Reinforcement Learning (RL) for feature selection [112] [141]. This section applies RL-based feature selection to Human-Computer Interaction (HCI), focusing on multimodal physiological signal analysis.

Reinforcement Learning (RL) is a branch of ML that optimizes decision-making strategies in dynamic and uncertain environments. RL algorithms enable an agent to interact with its environment through trial and error. It refines its actions to maximize cumulative rewards over time. A Markov Decision Process (MDP) commonly models RL problems, representing sequential decision-making under uncertainty. RL adopts a single-agent approach in feature selection and formulates the selection process as an MDP. The framework defines the tuple (S, A, $P_a(s,s^{'})$, $R(s,a)$, γ), as follows:

1. s represents the set of states, each corresponding to a selected feature combination.

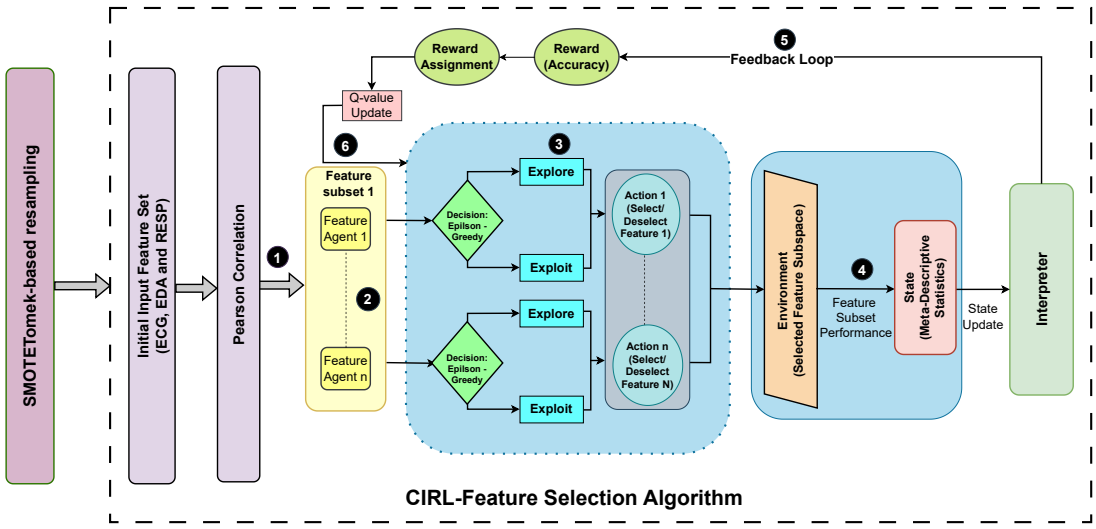


Figure 6.6: Proposed CIRL-FSA for anxiety classification.

2. A denotes the actions that involve selecting or removing a feature.

3. $P_a(s,s^{'})$ defines the probability of transitioning from state s to $s^{'}$ after performing action a.

4. $R(s,a)$ represents the reward obtained for executing action a in state s.

5. γ serves as the discount factor, determining the significance of future rewards in the decision-making process.

However, the single-agent approach struggles to explore and exploit the feature space effectively, as it fails to capture intricate interdependencies among features [73]. Additionally, relying on a single agent to select or remove all N features results in an exponentially growing action space of 2^N. As the number of features increases, this expansion significantly reduces the efficiency of the learning process. This approach resembles evolutionary algorithms, which attempt to identify optimal solutions within a constrained search space but often face scalability and computational efficiency challenges.

The Correlated Interactive Reinforcement Learning Feature Selection Algorithm (CIRL-FSA) enhances feature selection for anxiety classification by utilizing physiological signals, including ECG, EDA, and RESP. Figure 6.6 illustrates the proposed CIRL-FSA for anxiety classification, while Table 6.6 provides an overview of the parameters used in its implementation. The algorithm operates through multiple sequential phases:

Table 6.6: IRL parameters for implementing CIRL-FSA.

Parameter	Description	Value
epsilon	Exploration rate, balances exploration and exploitation.	0.7
alpha	Learning rate	0.4
epsilon decay rate	Decay rate for epsilon, decreases exploration over time.	0.999
alpha decay rate	Decay rate for an alpha, decreases learning rate over time.	0.999
num episodes	Number of episodes for the learning process.	150
num agents	Number of agents, each representing a feature.	24
Meta Descriptive Stat	mean, variance, standard deviation, median, Q1 (first quartile), Q2 (second quartile), Q3 (third quartile) and entropy	—

1. **Initial Feature Set and Pearson Correlation Analysis:**

   (a) **Feature Selection:** The algorithm starts with an initial feature set extracted from ECG, EDA, and RESP signals. The PCC evaluates the linear association between each feature and the target variable. This process helps to eliminate the redundant features by selecting those with strong correlations to the target, improving the selected subset's quality.

   (b) **Correlation Computation:** The relationship between a feature x and the target variable y is quantified using Equation 6.5:

$$\text{Correlation}(x,y) = \frac{\text{cov}(x,y)}{\text{std}(x) \cdot \text{std}(y)} \tag{6.5}$$

2. **Feature Agents and Actions:**

   (a) **Agent Allocation:** Once the initial feature selection is completed, each feature f_i is assigned to a distinct agent within the Interactive Reinforcement Learning (IRL) framework. Each agent operates within a binary action space, where 0 represents feature exclusion and 1 signifies feature selection.

   (b) **Formation of Feature Subset:** The selected agents collectively determine a feature subset, which is then evaluated based on its impact on classification accuracy. This approach minimizes the complexity of the action space,

facilitating a more effective and efficient exploration process.

3. **Exploration, Exploitation, and Reward Mechanism:**

   (a) **Epsilon-Greedy Exploration:** Agents utilize an epsilon-greedy strategy to maintain a balance between exploration, where they test new feature selections, and exploitation, where they rely on previously learned actions. This adaptive approach enables agents to refine feature selection based on its contribution to model performance.

   (b) **Reward Mechanism:** After agents finalize their selections, the updated feature subset is applied to the environment. The model uses this subset to assess its accuracy and assigns rewards accordingly. Performance improvements receive positive rewards, while accuracy declines result in negative rewards, guiding the learning process toward optimal feature selection.

4. **Environment and State Representation Using Meta-Descriptive Statistics:**

   (a) **Environment Update:** The environment, defined by the selected feature subspace, is updated continuously based on the actions taken by agents and the rewards they receive.

   (b) **State Representation:** At time t, the state s_t represents the configuration of the selected features. Meta-descriptive statistics enhance this representation by providing deeper insights into feature distributions. Once the process computes the initial set of statistics $Stat_i$ for each feature, it derives a second layer of statistical measures to refine feature characterization. The state s_t is defined by Equation 6.6:

$$s_t = Meta\_Descriptive\_Stats(f_1, f_2, ..., f_n) \tag{6.6}$$

   In this context, $Meta\_Descriptive\_Stats$ calculates higher-order statistical measures, including mean, variance, standard deviation, median, quartiles $(Q1, Q2, Q3)$, and entropy, based on the initially extracted feature statistics.

This method improves the decision-making process within the reinforcement learning framework by offering a deeper understanding of feature distributions.

5. **Interpreter and Feedback Mechanism:**

   (a) **Analysis and Iterative Refinement:** The interpreter evaluates the selected feature subset and updates the state representation using meta-descriptive statistics. This iterative process continues until agents converge on an optimized feature subset that enhances classification accuracy.

6. **Learning and Q-Value Updates:**

   (a) **Q-Value Adjustment:** Each agent refines its policy by incorporating both individual and collective rewards. The Q-value update mechanism, which guides the agent's decision-making, is defined by Equation 6.7:

$$Q_i(s,a_i) \leftarrow Q_i(s,a_i) + \alpha \left[R(s,a) + \gamma \max_{a_i^{'}} Q_i(s^{'},a_i^{'}) - Q_i(s,a_i) \right] \quad (6.7)$$

The agent updates its Q-value $Q_i(s,a_i)$ by considering the immediate reward $R(s,a)$ and the estimated maximum future reward $\max_{a_i^{'}} Q_i(s^{'},a_i^{'})$. The learning rate α controls the degree of Q-value adjustment, while the discount factor γ determines the importance of future rewards in the learning process.

6.3.1.6 Classification Algorithm and Hyperparameter Tuning

This work extracts features from three primary physiological signals: ECG-derived HRV, RESP-derived RRV, and EDA. It employs STCIRL-FSA to balance the data and minimize data dimensionality by removing redundant and less informative features. Following feature selection, the dataset is divided into training and testing sets using an 80:20 split.

Various classification models, including RF, SVM, KNN, and LGBM, are utilized for anxiety classification. Each model incorporates 10 to 15 hyperparameters, covering

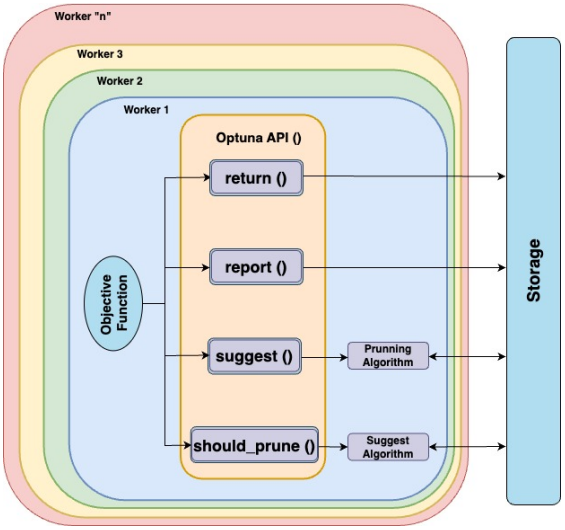


Figure 6.7: Optuna architecture hyperparameter tuning approach for STCIRL-FSA

categorical, continuous, conditional, and integer types, which are vital in optimizing performance. Given the computational complexity of hyperparameter tuning, the approach balances by keeping certain parameters at their default values while selectively optimizing others.

Further, this work employs the Optuna framework to optimize hyperparameters systematically. Researchers recognize Optuna for its efficient search and pruning algorithms. This approach enhances the tuning process while ensuring robust and reproducible classification outcomes [78]. Figure 6.7 presents the architecture of the Optuna framework, which utilizes a define-by-run strategy to construct the search space and determine optimal hyperparameters dynamically.

Optuna optimizes hyperparameters by leveraging advanced search *(SUGGEST ALGO)* and pruning mechanisms *(PRUNE ALGO)*, which refine the model by minimizing or maximizing the objective function to achieve an optimal validation score [6] [78]. The *SUGGEST ALGO* dynamically defines the search space, while the *PRUNE ALGO* eliminates underperforming trials early, ensuring computational efficiency and cost-effective hyperparameter tuning.

Multiple workers execute the objective function independently for each study. This function conducts trials using Optuna's APIs, retrieving relevant data from shared storage as required [237]. The *SUGGEST()* API dynamically generates hyperparameters, the *REPORT()* API tracks intermediate results, and the *SHOULD_PRUNE()* API terminates trials that fail to meet predefined conditions [237]. Further, each

Table 6.7: STCIRL-FSA confusion matrices for anxiety classification across distinct phases.

Initial											
RF			**SVM**			**KNN**			**LGBM**		
Class	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**
Low	0.044	0.065	**Low**	0.007	0.102	**Low**	0.035	0.074	**Low**	0.051	0.058
High	0.014	0.286	**High**	0.001	0.299	**High**	0.019	0.281	**High**	0.016	0.284
Post SMOTETomek											
RF			**SVM**			**KNN**			**LGBM**		
Low	0.848	0.124	**Low**	0.858	0.136	**Low**	0.819	0.153	**Low**	0.919	0.075
High	0.143	0.822	**High**	0.224	0.719	**High**	0.253	0.712	**High**	0.106	0.833
STCIRL-FSA: SMOTETomek + Post CIRL FSA											
RF			**SVM**			**KNN**			**LGBM**		
Low	0.869	0.103	**Low**	0.889	0.103	**Low**	0.940	0.132	**Low**	0.891	0.080
High	0.128	0.837	**High**	0.166	0.779	**High**	0.175	0.690	**High**	0.107	0.859
STCIRL-FSA + Optuna Optimization											
RF			**SVM**			**KNN**			**LGBM**		
Low	0.901	0.071	**Low**	0.929	0.074	**Low**	0.899	0.104	**Low**	0.952	0.029
High	0.106	0.859	**High**	0.094	0.840	**High**	0.153	0.779	**High**	0.061	0.895

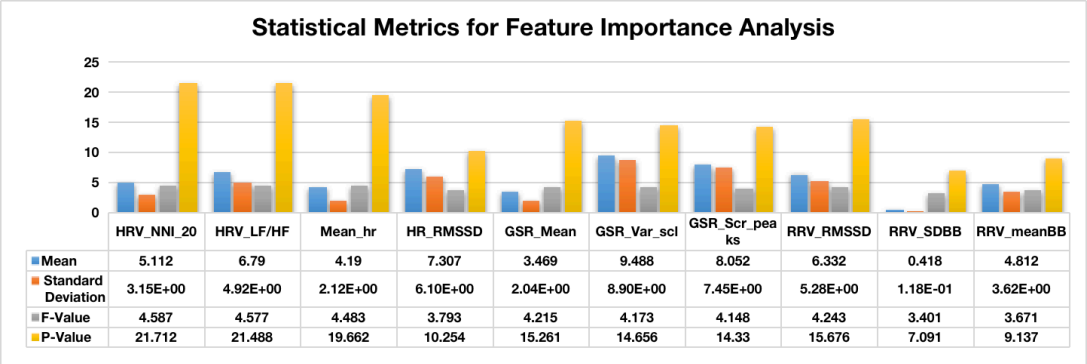
classification model undergoes evaluation using 10-fold cross-validation to ensure robust and generalizable results.

6.3.2 Experimental Results

This work utilizes accuracy (Acc), F1-score (F1), sensitivity (Sens), specificity (Spec), and the receiver operating characteristic (ROC) curve to measure the performance of the proposed STCIRL-FSA on the Spiderphobic dataset [109]. Table 6.7 details the confusion matrix, which provides the evaluation metrics.

6.3.2.1 Experimental setup

This work evaluates performance under two experimental setups: one employing a traditional machine learning framework and the other integrating the SMOTETomek data balancing technique, interactive reinforcement learning for feature selection, and Optuna for hyperparameter optimization.



	HRV_NNI_20	HRV_LF/HF	Mean_hr	HR_RMSSD	GSR_Mean	GSR_Var_scl	GSR_Scr_peaks	RRV_RMSSD	RRV_SDBB	RRV_meanBB
Mean	5.112	6.79	4.19	7.307	3.469	9.488	8.052	6.332	0.418	4.812
Standard Deviation	3.15E+00	4.92E+00	2.12E+00	6.10E+00	2.04E+00	8.90E+00	7.45E+00	5.28E+00	1.18E-01	3.62E+00
F-Value	4.587	4.577	4.483	3.793	4.215	4.173	4.148	4.243	3.401	3.671
P-Value	21.712	21.488	19.662	10.254	15.261	14.656	14.33	15.676	7.091	9.137

Figure 6.8: Statistical analysis for EDA+HRV+RRV feature subset obtained using CIRL-FSA Note: The p-values presented are transformed ($-\log_{10}$(p-value)).

6.3.2.2 Statistical Analysis

This work utilizes statistical metrics such as mean, standard deviation, F-value, and p-value to assess the significance and variability of features in the anxiety classification model. The mean represents the central tendency, while the standard deviation measures the dispersion, offering insights into the distribution and consistency of each feature. Section 3.7.2 provides a detailed description of the statistical analysis. Figure 6.8 illustrates the feature importance metrics to showcase the mean, standard deviation, F-value, and transformed p-values ($-log_{10}$) for each feature.

The analysis identifies HRV_NNI_20, HRV_LF/HF, Mean_hr, GSR_Mean, and RRV_RMSSD as key features for distinguishing anxiety levels, as indicated by their high F-values and p-values. These features exhibit significant variance between groups, highlighting their importance in the classification model. While features such as HR_RMSSD and RRV_SDBB show lower F-values and p-values, they still provide relevant information with a lesser impact.

6.3.2.3 Evaluation of classifier performance with different balancing approaches

This section investigates how a class imbalance in the Spiderphobic dataset [109] affects classifier performance and evaluates different techniques for data balancing. To improve class distribution, the data balancing approach includes SMOTE, Tomek Links, and the SMOTETomek combination. Before classification, the dataset undergoes normalization using the *standard_scaling* technique.

Table 6.8 imbalanced dataset results presents the baseline performance of classifiers

Table 6.8: STCIRL-FSA: Comparative performance of classifiers before and after data balancing techniques with 63 features.

	Imbalanced Dataset				Post SMOTE			
Classifier	**Acc.**	**F1**	**Sens**	**Spec**	**Acc.**	**F1**	**Sens**	**Spec**
RF	81.89	79.84	95.50	43.17	83.53	81.48	74.07	92.58
SVM	75.58	66.75	99.52	59.30	70.18	71.85	69.58	55.25
KNN	78.46	75.87	93.67	34.30	75.78	78.35	74.72	77.29
LGBM	81.90	88.47	94.66	46.78	88.47	87.27	83.62	92.81
	Post Tomek				**Post SMOTETomek**			
Classifier	**Acc.**	**F1**	**Sens**	**Spec**	**Acc.**	**F1**	**Sens**	**Spec**
RF	80.94	79.06	94.90	44.60	86.22	86.40	85.57	86.89
SVM	74.72	66.84	99.07	11.07	81.41	82.66	79.30	84.09
KNN	78.63	76.64	92.88	41.56	79.04	80.14	76.40	82.31
LGBM	83.52	82.14	76.16	90.82	90.64	91.04	89.66	91.74

on the original Spiderphobic dataset [109]. The results indicate high sensitivity but significantly lower specificity. The analysis reveals that the learning curve of the models tends to favor the majority class. Initially, RF achieves a sensitivity of 95.5% but has a lower specificity of 43.17%. In contrast, SVM achieves the highest sensitivity of 99.52% while maintaining a specificity of 59.3%. Similarly, KNN and LGBM show variations in sensitivity and specificity across different cases.

Further, Table 6.8 presents the results for Post SMOTE on distinct classifiers. The results indicate that applying SMOTE improves overall accuracy significantly. RF achieves an accuracy of 83.53%, while LGBM achieves the highest accuracy of 88.47%. These results highlight SMOTE's role in enhancing class balance and optimizing performance. However, accuracy slightly declines for SVM and KNN, with the accuracy of 70.18% and 75.78%, respectively. Despite this reduction, SMOTE improves the trade-off between sensitivity and specificity, demonstrating its effectiveness in achieving a more balanced class distribution.

Moreover, the Tomek resampling method improves the accuracy of SVM and KNN, reaching 74.72% and 78.63%, respectively. This enhancement results from TOMEK's ability to eliminate ambiguous data points to reduce misclassifications and increase model precision. However, its impact on balancing sensitivity and specificity remains limited, as it primarily addresses overlapping instances rather than fully resolving

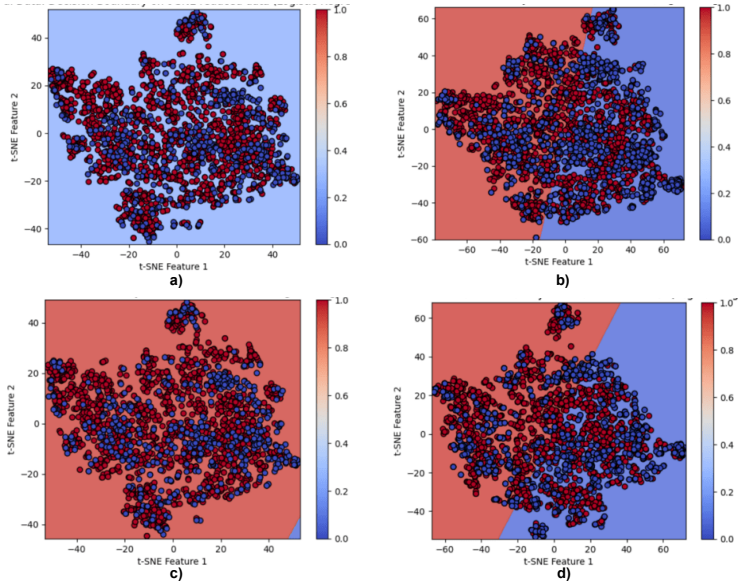


Figure 6.9: Effectiveness of different resampling techniques on the Spiderphobic dataset [109]. (a) Original imbalanced dataset. (b) Dataset after applying SMOTE. (c) Dataset after applying Tomek Links. (d) Dataset after applying SMOTETomek.

class imbalance. Hence, this experiment incorporates the SMOTETomek approach by leveraging the improved metric balance provided by SMOTE and the enhanced precision achieved through the Tomek method.

Table 6.8 indicates that using the SMOTETomek technique significantly improves the balance between sensitivity and specificity, essential for effective classification in imbalanced datasets. Among the classifiers, LGBM achieves the highest accuracy of 92.45% and an F1-score of 90.50%, demonstrating a balance with sensitivity at 91.14% and specificity at 93.84%. Additionally, Figure 6.9 compares the visualizations and decision boundaries of the dataset before and after applying various resampling techniques. The visualizations reveal how the dataset distribution changes and how these techniques influence the separability of the classes.

6.3.2.4 Evaluation of classifier performance with SMOTETomek and CIRL-FSA

Table 6.9 presents the impact of applying CIRL-FSA following the SMOTETomek resampling technique, which significantly enhances classifier performance. Using all 63 features initially allows the model to capture essential information from physiological signals for anxiety classification. However, reducing the feature set to 10 through CIRL-

Table 6.9: STCIRL-FSA before optimization: Evaluation of features selected using the CIRL-FSA algorithm after SMOTETomek resampling and before optimization.

STCIRL-FSA: Post SMOTETomek + CIRL-FSA (10)				
	Acc	**F1**	**Sens**	**Spec**
RF	88.07	88.27	87.16	89.04
SVM	86.11	86.86	84.27	88.32
KNN	84.15	85.96	84.30	83.94
LGBM	**90.35**	90.50	89.28	91.48

FSA further improves classifier performance, increasing efficiency and potentially lowering computational complexity without compromising predictive accuracy.

The SVM classifier demonstrates an accuracy increase from 84.41% to 86.11%, while its F1-score rises from 82.66% to 86.86%. KNN also shows notable improvement, with accuracy increasing from 79.04% to 84.15% and the F1-score advancing from 80.14% to 85.96%. Additionally, RF achieves an accuracy of 88.07%, and LGBM exhibits a slight enhancement, reaching 90.35%. These results highlight the effectiveness of CIRL-FSA in selecting the most influential features, strengthening classifier performance with improved precision.

6.3.2.5 STCIRL-FSA optimization using Optuna technique

Table 6.10 presents the impact of applying the Optuna optimization approach to classifiers previously refined through SMOTETomek resampling and CIRL-FSA. The accuracy of LGBM improved notably from 90.35% to 95.35%, with specificity increasing to 96.86%. Similarly, SVM achieved an accuracy of 91.33%, along with enhancements in both sensitivity and specificity. These findings demonstrate that integrating feature selection with hyperparameter tuning enhances classifier performance. The optimization technique allows models to capture complex data patterns better and effectively manage class imbalance.

The significant enhancement in classifier performance is mainly due to Optuna's adaptive hyperparameter tuning, which leverages Bayesian optimization to explore the hyperparameter space efficiently. This structured optimization process allows models to adjust their parameters based on the characteristics of the balanced dataset, leading to better generalization and improved overall accuracy.

Table 6.10: STCIRL-FSA after optimization: Enhancements in classifier performance using Optuna optimization approach.

STCIRL-FSA + Optuna Optimization (10 features)				
	Acc.	**F1**	**Sens**	**Spec**
RF	90.86	91.06	89.47	92.37
SVM	91.33	91.71	90.81	91.90
KNN	86.72	87.49	85.46	88.22
LGBM	**95.35**	95.49	93.98	96.86

Figure 6.10: STCIRL-FSA: Comparative performance metrics of all classifiers at different stages of the Optimization process.

Figure 6.10 illustrates the performance metrics of RF, SVM, KNN, and LGBM classifiers across different stages: initial, after SMOTETomek resampling, following CIRL-FSA feature selection, and with CIRL-FSA combined with Optuna optimization. Additionally, Figure 6.11 displays the ROC curves, demonstrating how each phase influences the true positive and false positive rates across the classifiers.

6.3.3 Discussion

Recent progress in anxiety classification has shown that combining physiological signals with ML enhances predictive accuracy. Signals like ECG, RESP, and EDA

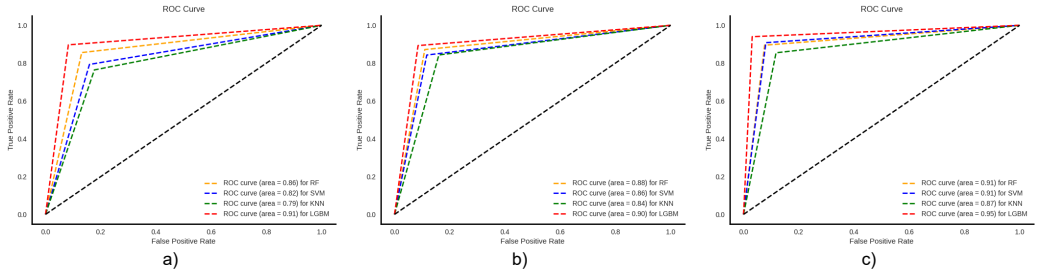


Figure 6.11: ROC curves illustrating the performance at different stages of the study: (a) Post SMOTETomek, (b) Post SMOTETomek + CIRL-FSA, and (c) Post SMOTETomek + CIRL-FSA + Optuna.

effectively capture physiological responses associated with anxiety. However, the Spiderphobic dataset [109] used in this study presents a significant challenge due to class imbalance, where 76% of the data corresponds to anxiety states and only 24% represents relaxed states. This uneven distribution can bias model predictions and may affect the classification performance and reduce overall reliability.

Hence, this study applies the SMOTETomek technique, which integrates oversampling for the minority class with data refinement to eliminate mislabeled instances that could affect model accuracy. This method effectively balances the dataset, allowing classifiers to learn more effectively from both classes. As a result, classification performance improves significantly, with the LGBM classifier achieving 90.64% accuracy, an F1-score of 91.04%, a sensitivity of 89.66%, and a specificity of 91.74% before optimization.

A major challenge in physiological data analysis is the high dimensionality of extracted features, including complex interactions and linear and non-linear correlations, often limiting predictive accuracy. Hence, the study introduces the CIRL-FSA algorithm, which utilizes agent-model interaction and prediction accuracy feedback to analyze feature relationships systematically. The agent continuously refines feature subset selection through dynamic interactions with the prediction models, resulting in improved classification accuracy for anxiety classification.

Table 6.11 presents a summary of the selected features, their physiological significance, and their relevance to anxiety classification. Machine learning models, including RF, SVM, KNN, and LGBM, use the selected features as inputs. The study applies hyperparameter tuning through the Optuna framework to enhance classification

Table 6.11: CIRL-FSA-based selected features, domains, and their impact on anxiety classification

Feature	Domain	Impact on Anxiety Disorder
HRV_NNI_20	HRV	Indicating adaptive responses to anxiety-inducing situations.
HRV_LF/HF	HRV	Increased LF/HF ratio signals sympathetic arousal.
Mean_hr	HRV	Elevates during anxiety due to sympathetic activation.
HR_RMSSD	HRV	Decrease in increased sympathetic activity.
GSR_Mean	EDA	Increases with emotional arousal.
GSR_Var_scl	EDA	Reflects dynamic emotional states.
GSR_Scr_peaks	EDA	Represents increased arousal during anxiety.
RRV_RMSSD	RRV	Indicates irregular breathing during anxiety.
RRV_SDBB	RRV	Highlighting disrupted patterns during stress.
RRV_meanBB	RRV	Represents average respiratory cycle length and alters during emotional distress.

performance.

Optuna optimization leads to a 5% increase in LGBM accuracy, while SVM improves by 5.22%, KNN by 2.75%, and RF by 2.79%. These findings highlight the importance of combining advanced feature selection with adaptive hyperparameter tuning. CIRL-FSA effectively selects the most relevant features, while Optuna optimization fine-tunes classifier parameters to align with dataset characteristics. This combined approach enhances model generalization, mitigates class imbalance and high-dimensional data challenges, and strengthens its applicability for anxiety classification.

This work further conducts a detailed comparative analysis against five conventional feature selection methods used in previous anxiety studies: PCC, SFS, MI, RFE, Permutation Importance (PI), and RF, to assess the effectiveness of the proposed CIRL-FSA algorithm. A standardized approach is applied to preprocessing, feature extraction, and classification, with the only variation being the feature selection technique. As shown in Table 6.12, CIRL-FSA surpasses all other methods, selecting just ten features while achieving the highest scores across key performance metrics, including accuracy, F1-score, precision, and recall.

Traditional feature selection techniques, including PCC, SFS, MI, RFE, PI, and RF, often involve a trade-off between the number of selected features and recall performance. For example, PCC, SFS, PI, and RF reduce the feature sets to 24, 16, 16, and 17

Table 6.12: Comparison of CIRL-FSA with traditional feature selection algorithms using LGBM.

Feature Selection Method	Acc	F1	Precision	Recall	No of Features
Baseline (All Features)	92.45	92.57	94.05	91.14	62
PCC	92.79	92.76	94.78	89.23	24
SFS	94.09	94.54	96.56	91	16
MI	94.94	94.40	97.37	90.95	12
RFE	91.73	90.22	93.27	87.47	12
PI	93.34	91.82	94.64	91.73	16
RF	94.17	94.63	96.18	90.85	17
CIRL-FSA	**95.35**	95.49	97.04	93.98	10

features, respectively, yet yield only slight improvements in recall. Although MI and RFE narrow the feature set to 12 features and achieve high precision, they show lower recall, suggesting that they misclassify several relevant instances. In contrast, CIRL-FSA maintains a balance between recall and precision, which is critical for anxiety classification, ensuring accurate identification of both anxiety and non-anxiety cases. These findings emphasize the importance of further exploring the broader applicability of the CIRL-FSA algorithm.

6.3.3.1 Comparison with Existing Studies

Several studies have explored anxiety classification using physiological signals, employing various feature selection techniques and classification models. Table 6.13 presents distinct methodologies and limitations in previous studies. It highlights the need for more advanced approaches to improve classification accuracy and generalizability.

In previous studies on the Spiderphobic dataset [109], Ihmig et al. [109] extracted statistical, linear, and time-domain features from ECG, EDA, and BR signals and employed a wrapper approach for feature selection. The study utilized Bagged Trees (BT), DT, and quadratic SVM classifiers and achieved the highest accuracy of 89.8% using BT. However, the lack of hyperparameter optimization limited the potential for further performance improvements. Vulpe-Grigoras et al. [251] employed a 1D-CNN model on ECG data and extracted time domain features. The authors achieved the highest accuracy of 83.29% without employing a feature selection algorithm. Further,

Table 6.13: STCIRL-FSA comparison with existing anxiety classification studies.

Author	Subjects	Signal	Features	FSA	Optimization	Algorithm	Result	Limitations
Vulpe et al. [251]	57	ECG	TD	–	–	1D-CNN	1D-CNN 2 level: 83.29%	No feature selection and optimization performed.
Gazi et al. [80]	57	ECG, RESP, EDA	ECG: FD and linear, EDA: phasic, tonic, RESP: RRV features	Permutation approach	Leave-One-Subject-Out Cross-Validation	RF	RF 2 level: 88%	Permutation method is efficient but may miss feature interactions.
Ihmig et al. [109]	57	ECG, EDA, BR	Statistical, linear, and time-domain features	Sequential Feature Selection	–	Bagged Trees, DT, Quadratic SVM	Bagged Trees 2 level: 89.8%	No optimization technique applied.
Vaz et al. [248]	15	ECG, EDA, EMG	HRV: TD, FD, and non-linear, EDA, SCR, SCL, EMG: EMG peaks	RFECV	Nested stratified cross-validation	LR, LDA, DT, SVM, ADB, RF, XGB	ADB: 92%	Small sample size limits generalization.
Tripathy et al. [244]	19	ECG	Entropy features	ANOVA	Grid Search with 10-fold CV	DT, LGBM, XGBoost, RF, ERT	XGBoost: 92.27%	Small dataset and focuses solely on entropy features, potentially missing other important features.
Henry et al. [101]	45	ECG, BVP	HR, HRV, FD	–	Grid Search with nested 5-fold CV	SVM, RF, XGBoost, MLP, ResNet	SVM: 81.1%	Deep learning models underperformed, suggesting inadequate feature representation.
Baygin et al. [35]	19	ECG	Proposed Probabilistic Binary Pattern (PBP)	NCA, Chi2	–	KNN+NCA, SVM+Chi2	SVM+Chi2: 99.94%	No validation on larger datasets, raising concerns about overfitting. No cross-validation was applied.
Salkevicius et al. [207]	30	BVP, EDA, ST	Statistical, Differential, Phasic, HR	RF	–	SVM	SVM: 86.3%	No optimization applied. A small dataset affects generalizability.
Rodriguez et al. [197]	21	ST, EDA, HR, BR, SPo2	Statistical	Correlation and MI	–	KNN, SVM, DT, LR	KNN: 95.56	No optimization technique applied. A small sample size reduces model generalizability.
Proposed Work	57	ECG, EDA, RESP	HRV, Phasic, Tonic, and RRV	Interactive Reinforcement Learning	Optuna + 10-fold CV	RF, SVM, KNN, LGBM	LGBM: 95.35%	Feature selection is automated but requires further validation on diverse datasets to ensure broad applicability.

Gazi et al. [80] utilized ECG, EDA, and RESP signals with a permutation-based feature selection approach. The authors utilized Leave-One-Subject-Out cross-validation and employed a Random Forest classifier. They achieved the highest accuracy of 88%. However, this method achieved an accuracy of 88%.

In another study, Vaz et al. [248] incorporated ECG, EDA, and EMG signals and employed RFECV with an unsupervised feature selection algorithm. The study used a nested stratified cross-validation approach with distinct classifiers, including LR, LDA, DT, SVM, ADB, RF, and XGB. The authors achieved the highest accuracy of 92% with 34 features using ADB. Tripathy et al. [244] utilized ECG-based entropy features and ANOVA for feature selection. The authors employed grid search with 10-fold cross-validation for optimization. The study achieved a maximum accuracy of 92.27% using XGBoost. Further, Salkevicius et al. [207] applied statistical, differential, and phasic feature extraction techniques on BVP, EDA, and ST signals. The authors employed RF for feature selection and reduced the features from 33 to 10. The study achieved the highest accuracy of 86.3% using SVM, but the absence of hyperparameter tuning and optimization limited the model's full potential. Rodriguez et al. [197] investigated statistical features from ST, EDA, HR, BR, and SpO2 signals, using correlation and

MI for feature selection. The study employed classifiers such as KNN, SVM, decision trees, and logistic regression. The authors achieved the highest accuracy of 95.56% with 13 features using KNN.

In a recent study, Baygin et al. [35] proposed a PBP approach for ECG feature extraction and utilized NCA and Chi-square methods for feature selection. The study achieved the highest classification accuracy, 98.81% for KNN+NCA and 99.94% for SVM+Chi2. Unlike previous studies, the proposed CIRL-FSA integrates interactive reinforcement learning-based feature selection with Optuna hyperparameter tuning to enhance classification performance. This experiment employs ECG, EDA, and RESP signals while addressing class imbalance through SMOTETomek resampling to ensure a more balanced dataset for training.

Moreover, while many existing studies apply grid search for hyperparameter tuning, this method exhaustively searches across predefined hyperparameter values, which can be computationally expensive and inefficient, particularly for high-dimensional datasets. Grid search also lacks adaptability, as it does not dynamically adjust the search space based on previous results, often leading to suboptimal tuning and increased training time.

In contrast, Optuna's Bayesian optimization intelligently explores the hyperparameter space by leveraging an adaptive search strategy. Unlike grid search, Optuna dynamically adjusts the search process based on prior evaluations, efficiently identifying the best-performing hyperparameter configurations while reducing computational costs. This adaptive approach significantly improves model optimization, allowing classifiers to align better with dataset characteristics. As a result, the LGBM classifier achieves 95.35% accuracy, outperforming conventional methods while reducing computational complexity by selecting only ten essential features.

6.3.4 Significant Outcome

The CIRL-FSA demonstrates the effectiveness of interactive reinforcement learning in adaptively selecting informative features from multimodal physiological data. Modeling feature selection as a dynamic decision-making process captures complex interdependencies that static methods often overlook. This work further extends CIRL-

FSA to the STCIRL-FSA by integrating the SMOTETomek resampling technique to enhance performance under class-imbalanced conditions. This extension significantly improves classification results while preserving the adaptive nature of the original algorithm. Although the reinforcement learning-based approach introduces higher computational costs, CIRL-FSA achieves high classification performance with a reduced feature set, confirming its potential for real-world anxiety detection in Human-Computer Interaction systems.

6.4 Comparative Evaluation of CorLMI-FSA and CIRL-FSA

This section presents a cross-dataset evaluation of CorLMI-FSA and CIRL-FSA to validate their performance and adaptability beyond their original application domains. Initially, CorLMI-FSA and CIRL-FSA are developed for stress and anxiety classification, respectively. Then, the evaluation applies each algorithm to the alternate dataset to assess their generalizability across related physiological contexts. Specifically, the evaluation applies CorLMI-FSA to the Spiderphobic anxiety dataset [109] and CIRL-FSA on the self-collected ASD. The comparison focuses on classification accuracy, feature reduction, computational efficiency, and adaptability to dataset-specific characteristics.

Table 6.14 presents the comparative evaluation of CorLMI-FSA and CIRL-FSA across the self-collected ASD and the publicly available Spiderphobic anxiety dataset [109]. On the ASD with two-level classification, CIRL-FSA achieves an accuracy of 95.35% and an F1-score of 95.49%, outperforms CorLMI-FSA, which attains 96.82% accuracy and 96.81% F1-score, but with substantially lower runtime. Similarly, CIRL-FSA achieves the highest accuracy of 96.94%, slightly outperforming CorLMI-FSA, with increased runtime. On the Spiderphobic dataset, CorLMI-FSA achieves 95.35% accuracy and 95.49% F1-score, outperforming CIRL-FSA, achieving an accuracy of 92.43% and F1-score of 91.53%, although its runtime is also higher in this case. These findings illustrate that while CIRL-FSA can model complex feature interactions and adapt to increased class granularity, CorLMI-FSA provides a better trade-off in simpler

Table 6.14: Comparative evaluation of CorLMI-FSA and CIRL-FSA on self-collected ASD and Spiderphobic anxiety [109] datasets

Datasets	Domain	Stress Class	Algorithm	Selected Features	Acc (%)	F1 (%)	Runtime (s)
Self-collected ASD	Stress	two	CorLMI-FSA	10	96.82	96.81	156
Self-collected ASD	Stress	two	CIRL-FSA	9	97.35	96.49	1074
Self-collected ASD	Stress	three	CorLMI-FSA	11	95.84	95.12	168
Self-collected ASD	Stress	three	CIRL-FSA	12	96.94	96.08	1179
Spiderphobic [109]	Anxiety	two	CorLMI-FSA	10	95.35	95.49	1015
Spiderphobic [109]	Anxiety	two	CIRL-FSA	11	92.43	91.53	226

or well-separated tasks due to its lower computational overhead and stable performance.

The cross-dataset comparison highlights the complementary strengths of the two proposed feature selection algorithms. CIRL-FSA demonstrates strong adaptability and achieves the highest accuracy in complex scenarios, such as multi-class classification, by dynamically selecting features through interactive reinforcement learning. However, this performance gain comes with increased runtime. In contrast, CorLMI-FSA offers faster execution with comparable or superior performance in binary classification settings. These results suggest that CorLMI-FSA is well suited for real-time or resource-constrained environments, whereas CIRL-FSA is better suited for tasks that require high adaptivity. Furthermore, Table 6.15 presents a qualitative comparison of CorLMI-FSA and CIRL-FSA across key algorithmic characteristics.

6.5 Summary

This chapter presents and evaluates two novel feature selection algorithms, CorLMI-FSA and ST-CIRL-FSA, for physiological signal-based stress and anxiety classification, respectively. CorLMI-FSA integrates correlation filtering with mutual information and logistic redundancy control to identify relevant and non-redundant features. ST-CIRL-FSA formulates feature selection as a sequential decision-making process using

Table 6.15: Qualitative comparison of CorLMI-FSA and CIRL-FSA across key algorithmic characteristics

Parameter	CorLMI-FSA	CIRL-FSA
Feature Selection	Filter-based	Reinforcement learning-based
Adaptability	Static selection	Dynamic selection based on environment feedback
Classification Performance	High accuracy in binary stress classification	High accuracy in multi-class and complex environment
Runtime Efficiency	Low computational overhead	High due to iterative exploration and feedback loop
Redundancy Handling	Logistic redundancy control in MI-based ranking	Implicitly handled through reward optimization
Interpretability	High	Moderate
Best Use Case	Real-time or resource-constrained applications	High-performance applications with flexible runtime

interactive reinforcement learning and incorporates SMOTETomek resampling to address class imbalance. It applies CorLMI-FSA to the Self-collected ASD and CIRL-FSA on the Spiderphobic anxiety dataset [109]. Further, it conducts cross-dataset evaluation by applying each algorithm to the alternate dataset to assess generalizability. Results indicate that CIRL-FSA achieves higher accuracy in multi-class classification tasks, while CorLMI-FSA exhibits better runtime efficiency and competitive performance in binary classification scenarios. Furthermore, qualitative comparisons show that CIRL-FSA offers superior adaptability and effectively captures complex feature interactions. In contrast, CorLMI-FSA provides faster execution, higher interpretability, and consistent feature selection.

This chapter is based on the following work:

- **J3: Shikha,** Divyashikha Sethia, and S. Indu. "ST-CIRL: A Reinforcement Learning-Based Feature Selection Approach for Enhanced Anxiety Classification" **Physiological Measurement.** (2025): 1-18. **SCIE, Impact factor: 2.3, Publisher: IOP Science**. Doi: `https://doi.org/10.1088/1361-6579/adb006`. **(Published)**.

- **C1: Shikha,** Divyashikha Sethia, and S. Indu. "CorLMI-FSA: An Efficient Feature Selection Approach for Stress Classification Using Physiological Signals." *In Proceedings of the 5th IEEE International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies (ICAECT), 2025.* Doi: `https://doi.org/10.1109/ICAECT63952.2025.10958862.` **(Published)**.

Chapter 7

Conclusion, Future Scope and Social Impact

With the growing demand for objective and real-time mental health monitoring, physiological signals have proven to be a reliable modality for detecting stress and anxiety [85]. These signals reflect dynamic changes in the autonomic nervous system, making them well-suited for capturing emotional fluctuations in everyday life. However, developing classification models that are both accurate and computationally efficient remains a significant challenge, particularly for wearable and real-world applications.

This thesis presents a comprehensive literature review to identify the current limitations in physiological signal-based stress and anxiety classification. It examines the relationship between physiological responses and mental states, highlights preprocessing techniques, feature extraction methods, and feature selection strategies, and evaluates machine learning and deep learning models, open datasets, and performance metrics. The review reveals critical gaps such as high-dimensional data, limited model interpretability, class imbalance, and computational inefficiency, emphasizing the need for scalable and real-time solutions.

To address these challenges, the thesis proposes **KRAFS-ANet**, a lightweight framework for EEG-based stress classification that integrates automated channel selection and ensemble stacking to enhance classification accuracy. This framework improves classification accuracy while significantly reducing computational load that makes it suitable for real-time implementation in wearable devices. Further, the

development and validation of this framework demonstrate its robustness across multiple datasets, confirming its adaptability and practical value.

Further, the thesis also introduces the **Academic Stress Dataset ASD**, a novel multimodal dataset collected in an academic environment using MIST to induce progressive stress in engineering students. The self-collected dataset fills a critical gap in existing research by enabling the analysis of stress responses for engineering students in the academic environment and highlighting the effectiveness of EDA and HRV as primary biomarkers.

In addition to enhance model interpretability and efficiency, the thesis proposes two novel feature selection algorithms for physiological signal-based stress and anxiety classification. The **CorLMI-FSA** employs correlation filtering and logistic redundancy control to reduce feature dimensionality without compromising accuracy. While, **CIRL-FSA** formulates feature selection as an interactive reinforcement learning process that dynamically explores and exploits feature dependencies across ECG, EDA, and RESP signals. To address class imbalance in the Spiderphobic dataset [109], CIRL-FSA is further extended to **STCIRL-FSA** by integrating the SMOTETomek resampling technique. Cross-dataset evaluations demonstrate that CIRL-FSA is more effective in multi-class scenarios, whereas CorLMI-FSA provides greater efficiency and stability in binary classification tasks. Chapter 6 covers this work.

In conclusion, this thesis establishes an effective and scalable pipeline for physiological signal-based stress and anxiety classification. The proposed methods advance the field by offering interpretable, efficient, and high-performing solutions, with strong potential for real-time and wearable applications in mental health monitoring and human–computer interaction.

7.1 Key Insights from This Research

Based on the literature review, experiments performed, and findings, the following are the key insights:

1. **How can physiological signals be used to classify stress and anxiety?** Physiological signals provide objective and quantifiable measures of stress and

anxiety by capturing the body's autonomic and central nervous system responses. These signals reflect emotional, cognitive, and physiological changes associated with stress and anxiety.

(a) EEG captures electrical activity in the brain, reflecting mental workload and cognitive stress responses. Stress can influence overall brain activity, causing changes in signal amplitude and frequency bands that help differentiate stress and anxiety levels.

(b) EDA measures sweat gland activity and reflects autonomic nervous system arousal, making it a key indicator of emotional and psychological stress. Increased skin conductance levels indicate heightened stress and anxiety responses.

(c) HRV reflects autonomic nervous system balance, where decreased HRV indicates higher sympathetic activity, stress, and anxiety, while increased HRV suggests a relaxed state.

(d) RESP analyzes breathing patterns, which change under stress and anxiety. Shallow, rapid breathing is often associated with heightened stress and anxiety, while deep, slow breathing corresponds to a relaxed state.

(e) BVP tracks vascular changes linked to stress-induced blood flow and heart rate fluctuations. Stress and anxiety can cause vasoconstriction or variations in pulse amplitude, which provide additional markers for classification.

Machine learning models can differentiate between varying stress and anxiety levels by analyzing patterns in these signals. The multimodal combination of these physiological markers improves classification accuracy, providing a reliable and objective assessment of stress and anxiety states.

2. **Why is multimodal integration important for stress and anxiety classification?**

   A single physiological signal may not always provide a complete and reliable representation of stress and anxiety responses due to noise, artifacts, and external influences. Many factors, such as motion artifacts, sensor inconsistencies, or

environmental interference, can distort signals, which may reduce classification accuracy. Additionally, physical exertion can cause fluctuations in heart rate, skin conductance, and respiratory patterns, which may not necessarily indicate stress and anxiety but rather physiological adaptation. The classification system becomes more robust and adaptable by integrating multiple signals, such as EEG and other peripheral signals, including ECG, EDA, BVP, and RESP. EEG provides direct insights into cognitive and emotional stress, while peripheral signals capture autonomic nervous system responses. If one signal is affected by noise or external conditions, other signals can compensate, improving reliability.

3. **Why is EEG integration important for stress classification?**

   While peripheral physiological signals such as ECG, EDA, and BVP capture autonomic nervous system responses to stress, EEG provides direct insights into brain activity, making it crucial for distinguishing between cognitive and physiological stress responses. EEG signals can help identify stress-induced neural changes that may not be reflected in peripheral signals alone. By integrating EEG with other physiological signals, stress classification models become more robust and adaptable, improving accuracy in detecting mental workload, cognitive stress, and emotional states. EEG-based stress assessment is particularly valuable in cases where physical exertion or environmental factors may alter peripheral physiological signals, ensuring a more comprehensive and reliable classification framework.

4. **How does automated channel selection enhance EEG-based stress and anxiety classification?**

   Traditional EEG-based stress classification relies on predefined channel selection, where researchers manually choose brain regions based on neuroscience knowledge. For example, the prefrontal cortex is commonly associated with cognitive stress, while other regions, such as the central and parietal lobes, are relevant for different cognitive and motor functions. However, this manual selection process lacks adaptability across different datasets and experimental conditions. Automated channel selection dynamically identifies the most informative channels based on data characteristics rather than relying

on fixed, task-specific assumptions. Data-driven selection retains only the most relevant EEG channels, reducing dimensionality, computational load, and sensor redundancy while maintaining high classification accuracy. Furthermore, minimizing the required EEG channels simplifies experimental setups, improves user comfort, and enhances the feasibility of real-time, wearable stress monitoring applications. This approach supports the development of more efficient and scalable EEG-based stress classification systems, making stress detection more adaptable across different datasets and real-world scenarios.

5. **What challenges arise from motion artifacts and physiological signal noise?**

   Physiological signals are prone to motion artifacts, sensor inconsistencies, and powerline interference, which can introduce noise and missing values, affecting classification performance. This issue is particularly significant in real-world environments where users may be in motion, leading to electrode displacement, muscle activity interference, and external disturbances. EEG, ECG, EDA, BVP, and RESP signals can all be affected, making distinguishing stress-induced changes from noise-related variations challenging. Preprocessing techniques such as filtering, Independent Component Analysis (ICA), and Empirical Mode Decomposition (EMD) help improve signal quality by reducing noise and artifacts. However, manual preprocessing is time-consuming and may not be effective in all cases. Further research is needed to develop automated artifact detection and removal techniques to enhance stress classification models' robustness and real-world applicability.

6. **How does feature selection impact stress classification performance?**

   Feature selection improves stress and anxiety classification performance by reducing data dimensionality, enhancing model interpretability, and improving generalization across datasets. Since the analysis of physiological signals spans multiple domains, such as time, frequency, and time-frequency, the extracted features create high-dimensional data. Although this provides comprehensive information, it also increases computational complexity and may introduce redundant or irrelevant features that negatively impact model performance. Some features may not contribute to stress and anxiety classification or could

even introduce noise, leading to overfitting and reduced model robustness. Additionally, certain features may have a negative impact on classification accuracy. Therefore, a well-optimized feature selection process removes redundant and less relevant features, ensuring better dataset generalization and improving classification efficiency. By selecting only the most discriminative and relevant features, models become computationally efficient and effective, less prone to overfitting, and more adaptable to real-world applications.

7. **How can machine learning be used for automated stress and anxiety classification?**

   Physiological signals such as EEG, ECG, EDA, BVP, and RESP are collected from wearable sensors and then preprocessed to remove artifacts and noise, ensuring cleaner data for analysis. After preprocessing, feature extraction techniques obtain relevant statistical, time-domain, and frequency-domain features, which serve as inputs for machine learning models. Supervised learning algorithms such as SVM, KNN, RF, and GB are commonly employed to classify stress levels, as they effectively map extracted features to different stress categories. In addition, deep learning models like ANN, MLP, and CNN are utilized, allowing the system to learn complex temporal and spatial patterns directly from raw physiological signals. High-dimensional feature sets often introduce redundancy and affect model performance, making feature selection essential to retain the most relevant features and improve classification accuracy and computational efficiency. The overall effectiveness of these models depends on multiple factors, including the quality and diversity of training data, feature engineering strategies, and hyperparameter optimization techniques. Methods such as Optuna-based tuning enhance performance further, ensuring that the models generalize well across different datasets and provide reliable stress classification.

7.1.1 Limitations

One of the key challenges in real-world stress classification is handling motion artifacts and noise caused by body movements, power interference, and sensor inconsistencies.

Unlike controlled laboratory environments, real-world physiological signals are prone to missing values, noise, and outliers, which can degrade classification performance. Physiological signals may vary due to muscle activity, electrode displacement, baseline drift, and environmental interference. To address these issues, researchers should explore advanced filtering techniques such as low-pass, high-pass, and Butterworth filters, along with Independent Component Analysis (ICA) and Empirical Mode Decomposition (EMD) for artifact removal and signal enhancement.

Although this study employs a multimodal approach, it does not integrate EEG with other physiological signals such as ECG and EDA. EEG plays a crucial role in stress processing, as brain activity directly reflects emotional and cognitive stress responses. Combining EEG with peripheral signals could enhance classification accuracy by distinguishing stress-induced changes from physiological fluctuations due to physical exertion or external factors. Future research should explore EEG-based multimodal fusion to improve the robustness and reliability of stress detection.

Additionally, while machine learning models have demonstrated high performance, incorporating advanced deep learning approaches and self-supervised learning techniques could further enhance feature representation and classification accuracy in stress detection.

7.1.2 Future work

This study provides a strong foundation for stress classification using physiological signals, but several areas remain open for further exploration. Future researchers can extend this work in the following directions:

1. Conducting real-time data collection in natural environments to evaluate model performance outside controlled laboratory conditions.

2. Integrating the proposed system into a mobile application to enable real-time stress monitoring and adaptive intervention strategies for mental health management.

3. Further, designing a real-time stress intervention framework that not only detects stress but also provides personalized interventions, such as biofeedback, breathing

exercises, or meditation recommendations, based on physiological responses.

4. Expanding stress classification by integrating Physiological signals with additional bodily expressions of emotion, such as eye tracking, speech, and gestures, to develop a more comprehensive stress detection system.

5. Future research can explore advanced deep learning models such as transformers and contrastive learning for better feature extraction and classification accuracy. Additionally, investigating self-supervised learning techniques can help models generalize better across different datasets.

7.2 Social Impact & Applications

This research on stress and anxiety classification using physiological signals has significant social implications, contributing to mental health awareness, workplace well-being, academic performance, and healthcare advancements. Integrating machine learning-based stress detection models into wearables, mobile applications, and human-computer interaction (HCI) systems enables real-time stress monitoring and early intervention strategies, ensuring a proactive approach to stress management.

1. **Smart Mental Health Awareness and Early Intervention:** Stress and anxiety disorders often go undiagnosed, leading to long-term mental health consequences such as burnout, depression, and reduced cognitive performance. By leveraging machine learning models and physiological signals, this research provides an objective and non-invasive approach to detecting stress early. Integrating real-time stress monitoring into wearables and mobile applications allows individuals to track their stress levels and take preventive actions such as guided breathing, meditation, and lifestyle adjustments. Future AI-driven advancements can improve personalized mental health interventions, reducing the societal burden of stress-related disorders, as shown in Figure 7.1.

2. **Enhancing Workplace Productivity and Employee Well-being:** Chronic stress in workplaces leads to decreased efficiency, burnout, and health complications, affecting both employees and organizations. Implementing real-time stress

Figure 7.1: Smart mental health monitoring through mobile applications and wearable devices.

Figure 7.2: AI-driven workplace stress monitoring systems integrate wearable sensors, allowing employees to track their stress levels and real-time dashboards to show personalized interventions.

detection systems in workplaces can help monitor employees' stress levels, suggest personalized interventions, and optimize workload distribution. Figure 7.2 illustrates an AI-driven workplace stress monitoring that integrates wearable sensors with real-time dashboards. It can track employees' stress levels while receiving personalized stress-relief recommendations such as guided breathing exercises and workload adjustments. Employers can integrate these models into corporate wellness programs to promote a healthier work environment. Organizations can offer stress management resources by identifying high-stress periods, reducing employee turnover, and improving overall productivity.

3. **Stress Detection in Academic Environments:** Students face significant academic stress, which affects learning outcomes, concentration, and emotional well-being. The data collected in this research highlights how physiological signals reflect stress responses in an academic environment. Figure 7.3 illustrates how AI-assisted stress detection systems can be integrated into classrooms to monitor physiological responses in real-time. Wearable sensors and adaptive

Figure 7.3: AI-driven stress monitoring in academic environment

learning platforms help track cognitive load, enabling personalized interventions that enhance student well-being and academic performance.

4. **Advancing Healthcare and Remote Monitoring for Elderly and Chronic Patients:** Chronic stress significantly impacts cardiovascular health, hypertension, and mental well-being, particularly among elderly individuals and individuals with long-term health conditions. The integration of wearable devices in stress monitoring enables continuous physiological data collection, as illustrated in Figure 7.4. Additionally, AI-driven virtual assistants enhance patient engagement in telemedicine applications, shown in Figure 7.5. By embedding stress classification models into wearable medical devices, clinicians can assess stress levels in real time, leading to improved mental health diagnostics. This research has the potential to enhance mental healthcare accessibility, ensuring that stress management becomes an integral component of preventive healthcare systems.

5. **Human-Computer Interaction (HCI) and Smart Environments:** Integrating stress detection models into HCI applications enables adaptive and personalized interactions across various domains. Smart home environments can adjust lighting, temperature, and background music based on detected stress levels

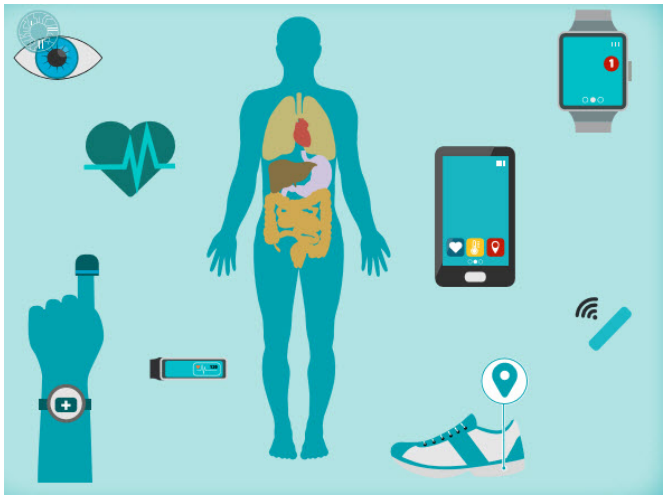


Figure 7.4: Wearable technology and remote patient monitoring for AI-driven stress detection.

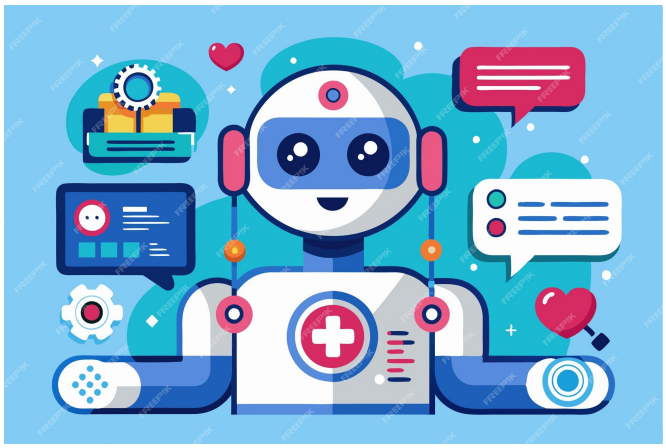


Figure 7.5: AI-driven virtual assistants in telemedicine: Intelligent chatbots and virtual assistants provide real-time support for stress management and mental health monitoring.

to create a calming atmosphere. Similarly, virtual reality (VR) and gaming applications can modify gameplay difficulty in real time to prevent frustration and optimize user experience. Additionally, AI-driven virtual assistants can recommend stress-relief activities such as guided breathing and mindfulness exercises to help users manage their emotional states effectively. Figure 7.6a) illustrates the application of VR-based stress relief techniques, where users engage in immersive relaxation environments to reduce stress levels. Additionally, Figure 7.6b) demonstrates how AI-powered smart assistants can provide personalized stress management solutions, such as music therapy or meditation suggestions, enhancing human-computer interactions in everyday life.

Figure 7.6: AI-Driven Stress Management in HCI

Figure 7.7: Real-time workplace stress monitoring using smart wearables and AI-driven notifications.

6. **Smart Wearables and Personalized Stress Management:** With the growing use of wearable technology, stress classification models can be embedded into smartwatches, fitness bands, and biosensors to provide real-time stress tracking and adaptive interventions. Features may include:

   - Automated stress alerts when stress levels exceed normal thresholds.
   - AI-driven recommendations for relaxation techniques such as guided breathing and meditation.
   - Daily stress pattern analysis to help users understand their stress triggers.

   These advancements contribute to personalized stress management, enabling individuals to actively monitor and control their stress levels through real-time feedback and intervention strategies, as shown in Figure 7.7.

Bibliography

[1] A. K. Abbas et al. Effective Connectivity in Brain Networks Estimated Using EEG Signals Is Altered in Children with ADHD. *Computers in Biology and Medicine*, 134:104515, 2021.

[2] J. Adu. *Perceived Impact of Stress on the Academic Performance and Health of Students in Colleges of Education in Ghana*. Ph.D. Thesis, University of Cape Coast, 2023.

[3] K. Ahammed and M. U. Ahmed. Quantification of Mental Stress Using Complexity Analysis of EEG Signals. *Biomedical Engineering: Applications, Basis and Communications*, 32(02):2050011, 2020.

[4] Z. Ahmad and N. Khan. A Survey on Physiological Signal-Based Emotion Recognition. *Bioengineering*, 9(11):688, 2022.

[5] J. W. Ahn, Y. Ku, and H. C. Kim. A Novel Wearable EEG and ECG Recording System for Stress Assessment. *Sensors*, 19(9):1991, 2019.

[6] T. Akiba et al. Optuna: A Next-Generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 2623–2631, 2019.

[7] A. Al-Ezzi et al. EEG Frontal Theta-Beta Ratio and Frontal Midline Theta for the Assessment of Social Anxiety Disorder. In *Proceedings of the 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, pages 107–112, 2020.

[8] A. Al-Ezzi et al. Review of EEG, ERP, and Brain Connectivity Estimators as

Predictive Biomarkers of Social Anxiety Disorder. *Frontiers in Psychology*, 11:730, 2020.

[9] A. Al-Ezzi et al. Severity Assessment of Social Anxiety Disorder Using Deep Learning Models on Brain Effective Connectivity. *IEEE Access*, 9:86899–86913, 2021.

[10] U. M. Al-Saggaf et al. Performance Evaluation of EEG-Based Mental Stress Assessment Approaches for Wearable Devices. *Frontiers in Neurorobotics*, 15:819448, 2022.

[11] F. Al-Shargie. Prefrontal Cortex Functional Connectivity Based on Simultaneous Record of Electrical and Hemodynamic Responses Associated With Mental Stress. *arXiv preprint arXiv:2103.04636*, 2021.

[12] F. Al-Shargie et al. Simultaneous Measurement of EEG-fNIRS in Classifying and Localizing Brain Activation to Mental Stress. In *Proceedings of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 282–286, 2015.

[13] A. Alberdi, A. Aztiria, and A. Basarab. Towards an Automatic Early Stress Recognition System for Office Environments Based on Multimodal Measurements: A Review. *Journal of Biomedical Informatics*, 59:49–75, 2016.

[14] F. Albertetti, A. Simalastar, and A. Rizzotti-Kaddouri. Stress Detection With Deep Learning Approaches Using Physiological Signals. In *Proceedings of the International Conference on IoT Technologies for HealthCare*, pages 95–111, 2020.

[15] M. Aldayel and A. Al-Nafjan. A Comprehensive Exploration of Machine Learning Techniques for EEG-Based Anxiety Detection. *PeerJ Computer Science*, 10:e1829, 2024.

[16] O. AlShorman et al. Frontal Lobe Real-Time EEG Analysis Using Machine Learning Techniques for Mental Stress Detection. *Journal of Integrative Neuroscience*, 21(1):20, 2022.

[17] F. Amzica and M. Steriade. Electrophysiological Correlates of Sleep Delta Waves. *Electroencephalography and Clinical Neurophysiology*, 107(2):69–83, 1998.

[18] L. Ancillon, M. Elgendi, and C. Menon. Machine Learning for Anxiety Detection Using Biosignals: A Review. *Diagnostics*, 12(8):1794, 2022.

[19] S. Andradóttir. A Review of Random Search Methods. *Handbook of Simulation Optimization*, pages 277–292, 2014.

[20] T. Anwar and S. Zakir. Machine Learning Based Real-Time Diagnosis of Mental Stress Using Photoplethysmography. *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, 55:154–167, 2022.

[21] L. S. Aquino and L. dos Santos Coelho. Stacking Ensemble Learning Approaches Applied to Emotional State Classification, 2021.

[22] S. Aristizabal et al. The Feasibility of Wearable and Self-Report Stress Detection Measures in a Semi-Controlled Lab Environment. *IEEE Access*, 9:102053–102068, 2021.

[23] A. Arsalan et al. Classification of Perceived Mental Stress Using a Commercially Available EEG Headband. *IEEE Journal of Biomedical and Health Informatics*, 23(6):2257–2264, 2019.

[24] A. Arsalan and M. Majid. A Study on Multi-Class Anxiety Detection Using Wearable EEG Headband. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–11, 2021.

[25] A. Asif et al. Human Stress Classification Using EEG Signals in Response to Music Tracks. *Computers in Biology and Medicine*, 107:182–196, 2019.

[26] O. Attallah. An Effective Mental Stress State Detection and Evaluation System Using Minimum Number of Frontal Brain Electrodes. *Diagnostics*, 10(5):292, 2020.

[27] E. T. Attar. Review of Electroencephalography Signals Approaches for Mental Stress Assessment. *Neurosciences Journal*, 27(4):209–215, 2022.

[28] AUC-ROC. GeeksforGeeks Website. https://www.geeksforgeeks.org/auc-roc-curve/, 2025.

[29] Y. Badr et al. A Review on Evaluating Mental Stress by Deep Learning Using EEG Signals. *Neural Computing and Applications*, pages 1–26, 2024.

[30] A. Baghdadi et al. DASPS: A Database for Anxious States Based on a Psychological Stimulation. *arXiv preprint arXiv:1901.02942*, 2019.

[31] A. Baghdadi et al. Psychological Stimulation for Anxious States Detection Based on EEG-Related Features. *Journal of Ambient Intelligence and Humanized Computing*, 12(8):8519–8533, 2021.

[32] A. Baird et al. An Evaluation of Speech-Based Recognition of Emotional and Physiological Markers of Stress. *Frontiers in Computer Science*, 3:750284, 2021.

[33] I. N. Baranov-Krylov and V. T. Shuvaev. Effects of Selective Visual Attention in the Parietal and Temporal Areas of the Human Cortex Using Evoked Potential Data. *Neuroscience and Behavioral Physiology*, 35:159–164, 2005.

[34] L. Barker. How to Build Occipital Lobes. In *How to Build a Human Brain*, pages 165–208. Springer, 2024.

[35] M. Baygin et al. Automated Anxiety Detection Using Probabilistic Binary Pattern With ECG Signals. *Computer Methods and Programs in Biomedicine*, 247:108076, 2024.

[36] O. M. Bazanova and D. Vernon. Interpreting EEG Alpha Activity. *Neuroscience & Biobehavioral Reviews*, 44:94–110, 2014.

[37] M. Benchekroun. *Continuous Stress Detection From Physiological Signals*. Ph.D. Thesis, Université de Technologie de Compiègne, 2024.

[38] S. Betti et al. Evaluation of an Integrated System of Wearable Physiological Sensors for Stress Monitoring in Working Environments by Using Biological Markers. *IEEE Transactions on Biomedical Engineering*, 65(8):1748–1758, 2017.

[39] S. Bhatnagar et al. A Deep Learning Approach for Assessing Stress Levels in Patients Using Electroencephalogram Signals. *Decision Analytics Journal*, 7:100211, 2023.

[40] P. J. Bota et al. A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE Access*, 7:140990–141020, 2019.

[41] S. Campanella et al. A Method for Stress Detection Using Empatica E4 Bracelet and Machine-Learning Techniques. *Sensors*, 23(7):3565, 2023.

[42] Y. S. Can et al. Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. *Sensors*, 19(8):1849, 2019.

[43] Y. S. Can et al. Stress Detection in Daily Life Scenarios Using Smart Phones and Wearable Sensors: A Survey. *Journal of Biomedical Informatics*, 92:103139, 2019.

[44] S. Carreiro et al. Wearable Sensor-Based Detection of Stress and Craving in Patients During Treatment for Substance Use Disorder: A Mixed Methods Pilot Study. *Drug and Alcohol Dependence*, 209:107929, 2020.

[45] R. Castaldo et al. Detection of Mental Stress Due to Oral Academic Examination via Ultra-Short-Term HRV Analysis. In *Proceedings of the 38th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, pages 3805–3808, 2016.

[46] R. Castaldo et al. Ultra-Short Term HRV Features as Surrogates of Short Term HRV: A Case Study on Mental Stress Detection in Real Life. *BMC Medical Informatics and Decision Making*, 19:1–13, 2019.

[47] M. N. Chaitanya et al. A Wearable, EEG-Based Massage Headband for Anxiety Alleviation. In *Proceedings of the 39th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, pages 3557–3560, 2017.

[48] S. Chatterjee and Y. Byun. EEG-Based Emotion Classification Using Stacking Ensemble Approach. *Sensors*, 22(21):8550, 2022.

[49] N. V. Chawla et al. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[50] W. Chen. Electrocardiogram. In *Seamless Healthcare Monitoring: Advancements in Wearable, Attachable, and Invisible Devices*, pages 3–44. Springer, 2018.

[51] K. M. Dalmeida and G. L. Masala. HRV Features as Viable Physiological Markers for Stress Detection Using Wearable Devices. *Sensors*, 21(8):2873, 2021.

[52] F. Daneshmand-Bahman and A. Goshvarpour. Anxiety Recognition Using a New EEG Signal Analysis Approach Based on Sample Density in a Chebyshev Chaotic Map. In *Affective Computing in Healthcare: Applications Based on Biosignals and Artificial Intelligence*, pages 1–1. IOP Publishing Bristol, UK, 2023.

[53] Y. Dasari, L. R. Chebolu, and V. Balasubramanian. Electroencephalogram Analysis on Alpha/Beta and Theta/Beta Ratios Due to Shirodhara. *Journal of Ayurveda and Integrative Medicine*, 16(2):101094, 2025.

[54] C. Daudelin-Peltier et al. The Effect of Acute Social Stress on the Recognition of Facial Expression of Emotions. *Scientific Reports*, 7(1):1036, 2017.

[55] R. J. Davidson et al. While a Phobic Waits: Regional Brain Electrical and Autonomic Activity in Social Phobics During Anticipation of Public Speaking. *Biological Psychiatry*, 47(2):85–95, 2000.

[56] R. J. S. G. Davidson. Frontal Versus Parietal EEG Asymmetry During Positive and Negative Affect. *Psychophysiology*, 16(2):202–203, 1979.

[57] G. Debard et al. Making Wearable Technology Available for Mental Healthcare Through an Online Platform With Stress Detection Algorithms: The Carewear Project. *Journal of Sensors*, 2020(1):8846077, 2020.

[58] K. Dedovic et al. The Montreal Imaging Stress Task: Using Functional Imaging to Investigate the Effects of Perceiving and Processing Psychosocial Stress in the Human Brain. *Journal of Psychiatry and Neuroscience*, 30(5):319–325, 2005.

[59] A. Demerdzieva. EEG Characteristics of Generalized Anxiety Disorder in Childhood. *Acta Informatica Medica*, 19(1):9, 2011.

[60] Y. Deng et al. Evaluating Feature Selection for Stress Identification. In *Proceedings of the 13th IEEE International Conference on Information Reuse and Integration (IRI)*, pages 584–591, 2012.

[61] H. Díaz et al. EEG Beta Band Frequency Domain Evaluation for Assessing Stress and Anxiety in Resting, Eyes Closed, Basal Conditions. *Procedia Computer Science*, 162:974–981, 2019.

[62] G. Dogan and F. P. Akbulut. Multi-Modal Fusion Learning Through Biosignal, Audio, and Visual Content for Detection of Mental Stress. *Neural Computing and Applications*, 35(34):24435–24454, 2023.

[63] V. Doma and M. Pirouz. A Comparative Analysis of Machine Learning Methods for Emotion Recognition Using EEG and Peripheral Physiological Signals. *Journal of Big Data*, 7(1):18, 2020.

[64] A. K. Dubey et al. Improved Method for Analyzing Electrical Data Obtained From EEG for Better Diagnosis of Brain Related Disorders. *Multimedia Tools and Applications*, 81(24):35223–35244, 2022.

[65] B. Egilmez et al. Ustress: Understanding College Student Subjective Stress Using Wrist-Based Passive Sensing. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 673–678, 2017.

[66] M. Elgendi et al. Dataset of Psychological Scales and Physiological Signals Collected for Anxiety Assessment Using a Portable Device. *Data*, 7(9):132, 2022.

[67] M. Elsadek, B. Liu, and Z. Lian. Green FaCCAdes: Their Contribution to Stress Recovery and Well-Being in High-Density Cities. *Urban Forestry & Urban Greening*, 46:126446, 2019.

[68] EMPATICA E4. Empatica E4 Website. `https://www.empatica.com/en-gb/`, 2022. Last accessed: August 7, 2022.

[69] E. Eren and T. S. Navruz. Stress Detection with Deep Learning Using BVP and EDA Signals. In *Proceedings of the 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–7, 2022.

[70] E. Ertin et al. Autosense: Unobtrusively Wearable Sensor Suite for Inferring the Onset, Causality, and Consequences of Stress in the Field. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems (SenSys)*, pages 274–287, 2011.

[71] I. I. Esener. Subspace-Based Feature Extraction on Multi-Physiological Measurements of Automobile Drivers for Distress Recognition. *Biomedical Signal Processing and Control*, 66:102504, 2021.

[72] E. M. Falconer et al. Developing an Integrated Brain, Behavior and Biological Response Profile in Posttraumatic Stress Disorder (PTSD). *Journal of Integrative Neuroscience*, 7(3):439–456, 2008.

[73] S. M. H. Fard, A. Hamzeh, and S. Hashemi. Using Reinforcement Learning to Find an Optimal Set of Features. *Computers & Mathematics with Applications*, 66(10):1892–1904, 2013.

[74] F. Fatih, Z. En-Naimani, and K. Haddouch. Comparative Study of Bayesian Optimization Process for the Best Machine Learning Hyperparameters. In *Proceedings of the Springer International Conference on Big Data and Internet of Things (BDIOT)*, pages 239–249, 2022.

[75] B. Fatimah et al. Mental Arithmetic Task Classification Using Fourier Decomposition Method. In *Proceedings of the IEEE International Conference on Communication and Signal Processing (ICCSP)*, pages 46–50, 2020.

[76] B. Fatimah, D. Pramanick, and P. Shivashankaran. Automatic Detection of Mental Arithmetic Task and Its Difficulty Level Using EEG Signals. In *Proceedings of the 11th IEEE International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–6, 2020.

[77] A. Fernandes et al. Determination of Stress Using Blood Pressure and Galvanic Skin Response. In *Proceedings of the IEEE International Conference on Communication and Network Technologies*, pages 165–168, 2014.

[78] M. Feurer and F. Hutter. Hyperparameter Optimization. In *Automated Machine Learning: Methods, Systems, Challenges*, pages 3–33. Springer, 2019.

[79] P. Garg et al. Stress Detection by Machine Learning and Wearable Sensors. In *Proceedings of the 26th ACM International Conference on Intelligent User Interfaces - Companion (IUI Companion)*, pages 43–45, 2021.

[80] A. H. Gazi et al. Respiratory Markers Significantly Enhance Anxiety Detection Using Multimodal Physiological Sensing. In *Proceedings of the IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–4, 2021.

[81] S. Gedam and S. Paul. A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques. *IEEE Access*, 2021.

[82] A. S. George et al. Predicting Autonomic Dysfunction in Anxiety Disorder From ECG and Respiratory Signals Using Machine Learning Models. *International Journal of Online & Biomedical Engineering*, 17(7), 2021.

[83] A. Ghaderi, J. Frounchi, and A. Farnam. Machine Learning-Based Signal Processing Using Physiological Signals for Stress Detection. In *Proceedings of the 22nd IEEE Iranian Conference on Biomedical Engineering (ICBME)*, pages 93–98, 2015.

[84] R. Ghosh et al. SAM 40: Dataset of 40 Subject EEG Recordings to Monitor the Induced-Stress While Performing Stroop Color-Word Test, Arithmetic Task, and Mirror Image Recognition Task. *Data in Brief*, 40:107772, 2022.

[85] G. Giannakakis et al. Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing*, 2019.

[86] G. Giannakakis, D. Grigoriadis, and M. Tsiknakis. Detection of Stress/Anxiety State From EEG Features During Video Watching. In *Proceedings of the 37th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, pages 6034–6037, 2015.

[87] N. D. Giardino, S. D. Friedman, and S. R. Dager. Anxiety, Respiration, and Cerebral Blood Flow: Implications for Functional Brain Imaging. *Comprehensive Psychiatry*, 48(2):103–112, 2007.

[88] M. Gil-Martin et al. Human Stress Detection With Wearable Sensors Using Convolutional Neural Networks. *IEEE Aerospace and Electronic Systems Magazine*, 37(1):60–70, 2022.

[89] E. Gkintoni and I. Dimakos. An Overview of Cognitive Neuroscience in Education. *EDULEARN22 Proceedings*, pages 5698–5707, 2022.

[90] S. Greene, H. Thapliyal, and A. Caban-Holt. A Survey of Affective Computing for Stress Detection: Evaluating Technologies in Stress Detection for Better Health. *IEEE Consumer Electronics Magazine*, 5(4):44–56, 2016.

[91] M. A. Hafeez and S. Shakil. EEG-Based Stress Identification and Classification Using Deep Learning. *Multimedia Tools and Applications*, 83(14):42703–42719, 2024.

[92] M. A. Hafeez, S. Shakil, and S. Jangsher. Stress Effects on Exam Performance Using EEG. In *Proceedings of the 14th IEEE International Conference on Emerging Technologies (ICET)*, pages 1–4, 2018.

[93] Z. Halim and M. Rehan. On Identification of Driving-Induced Stress Using Electroencephalogram Signals: A Framework Based on Wearable Safety-Critical Scheme and Machine Learning. *Information Fusion*, 53:66–79, 2020.

[94] L. Han et al. Detecting Work-Related Stress With a Wearable Device. *Computers in Industry*, 90:42–49, 2017.

[95] N. E. Haouij et al. Affectiveroad System and Database to Assess Driver's Attention. In *Proceedings of the 33rd ACM Annual Symposium on Applied Computing (SAC)*, pages 800–803, 2018.

[96] M. J. Hasan and J. Kim. A Hybrid Feature Pool-Based Emotional Stress State Detection Algorithm Using EEG Signals. *Brain Sciences*, 9(12):376, 2019.

[97] T. Hastie et al. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer, 2009.

[98] T. Hayashi et al. Beta Activities in EEG Associated with Emotional Stress. *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, 3(1):57–68, 2009.

[99] J. A Healey and R. W Picard. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, 2005.

[100] A. Hemakom, D. Atiwiwat, and P. Israsena. ECG and EEG Based Detection and Multilevel Classification of Stress Using Machine Learning for Specified Genders: A Preliminary Study. *Plos One*, 18(9):e0291070, 2023.

[101] J. Henry et al. On the Robustness of Machine Learning Models for Stress and Anxiety Recognition From Heart Activity Signals. *IEEE Sensors*, 2023.

[102] C. S. Herrmann. Gamma Activity in the Human EEG. In *Detection of Change: Event-Related Potential and FMRI Findings*, pages 167–183. Springer, 2003.

[103] B. A. Hickey et al. Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review. *Sensors*, 21(10):3461, 2021.

[104] A. Holzinger et al. Explainable AI Methods: A Brief Overview. In *Proceedings of the Springer International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers (XAI Workshop)*, pages 13–38, 2020.

[105] M. R. Hossain and D. Timmer. Machine Learning Model Optimization With Hyperparameter Tuning Approach. *Global Journal of Computer Science and Technology: Neural and Artificial Intelligence*, 21(2), 2021.

[106] C. Hsieh et al. Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection. In *Proceedings of the IEEE International Workshop on Signal Processing Systems (SiPS)*, pages 330–335, 2019.

[107] B. Hu et al. Signal Quality Assessment Model for Wearable EEG Sensor on Prediction of Mental Stress. *IEEE Transactions on Nanobioscience*, 14(5):553–561, 2015.

[108] J. Huang, Y. Liu, and X. Peng. Recognition of Driver's Mental Workload Based on Physiological Signals: A Comparative Study. *Biomedical Signal Processing and Control*, 71:103094, 2022.

[109] F. R. Ihmig et al. On-Line Anxiety Level Detection From Biosignals: Machine Learning Based on a Randomized Controlled Trial With Spider-Fearful Individuals. *Plos One*, 15(6):e0231517, 2020.

[110] F. I. Indikawati and S. Winiarti. Stress Detection From Multimodal Wearable Sensor Data. In *Proceedings of the IOP Conference Series: Materials Science and Engineering (ICSMSE)*, volume 771, page 012028, 2020.

[111] V. L. Ives-Deliperi and J. T. Butler. Relationship Between EEG Electrode and Functional Cortex in the International 10 to 20 System. *Journal of Clinical Neurophysiology*, 35(6):504–509, 2018.

[112] J. Jabri et al. Reinforcement Learning-Based Feature Selection for Improving the Performance of the Brain–Computer Interface System. *Signal, Image and Video Processing*, 17(4):1383–1389, 2023.

[113] M. Jaiswal et al. MUSE: A Multimodal Dataset of Stressed Emotion. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*, pages 1499–1510, 2020.

[114] E. H. Jang et al. Automated Detection of Panic Disorder Based on Multimodal Physiological Signals Using Machine Learning. *ETRI Journal*, 45(1):105–118, 2023.

[115] H. Jebelli, S. Hwang, and S. Lee. EEG-Based Workers' Stress Recognition at Construction Sites. *Automation in Construction*, 93:315–324, 2018.

[116] V. Jeyhani et al. Comparison of HRV Parameters Derived From Photoplethysmography and Electrocardiography Signals. In *Proceedings of the 37th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, pages 5952–5955, 2015.

[117] D. D. Jobson et al. The Role of the Medial Prefrontal Cortex in Cognition, Ageing and Dementia. *Brain Communications*, 3(3):fcab125, 2021.

[118] G. Jun and K. G. Smitha. EEG-Based Stress Level Identification. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3270–3274, 2016.

[119] E. E. Kaczor et al. Objective Measurement of Physician Stress in the Emergency Department Using a Wearable Sensor. In *Proceedings of the Annual Hawaii International Conference on System Sciences (HICSS)*, page 3729, 2020.

[120] P. Kalra and V. Sharma. Mental Stress Assessment Using PPG Signal: A Deep Neural Network Approach. *IETE Journal of Research*, pages 1–7, 2020.

[121] P. Kalra and V. Sharma. Mental Stress Assessment Using PPG Signal: A Deep Neural Network Approach. *IETE Journal of Research*, 69(2):879–885, 2023.

[122] D. Kamińska, K. Smółka, and G. Zwoliński. Detection of Mental Stress Through EEG Signal in Virtual Reality Environment. *Electronics*, 10(22):2840, 2021.

[123] O. Kara and O. Polo. Autonomic and Central Stress-Regulation Disintegration in Stress-Related Anxiety Disorders. *Acta Neuropsychologica*, 12(1), 2014.

[124] R. Katmah et al. A Review on Mental Stress Assessment Methods Using EEG Signals. *Sensors*, 21(15):5043, 2021.

[125] A. Katsarou et al. Validation of a Greek Version of PSS-14: A Global Measure of Perceived Stress. *Central European Journal of Public Health*, 20(2):104–109, 2012.

[126] C. D. Katsis et al. An Integrated System Based on Physiological Signals for the Assessment of Affective States in Patients With Anxiety Disorders. *Biomedical Signal Processing and Control*, 6(3):261–268, 2011.

[127] A. P. Kennedy et al. Continuous In-The-Field Measurement of Heart Rate: Correlates of Drug Use, Craving, Stress, and Mood in Polydrug Users. *Drug and Alcohol Dependence*, 151:159–166, 2015.

[128] H. Kim et al. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, 15(3):235, 2018.

[129] C. Kirschbaum, K. Pirke, and D. Hellhammer. The 'Trier Social Stress Test'–A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, 28(1-2):76–81, 1993.

[130] M. A. Klados et al. An Automatic EEG-Based System for the Recognition of Math Anxiety. In *Proceedings of the 30th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pages 409–412, 2017.

[131] P. Koch et al. Autotune: A Derivative-Free Optimization Framework for Hyperparameter Tuning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 443–452, 2018.

[132] S. Koelstra et al. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2011.

[133] S. Krishnan and Y. Athavale. Trends in Biomedical Signal Feature Extraction. *Biomedical Signal Processing and Control*, 43:41–63, 2018.

[134] A. Kumar, K. Sharma, and A. Sharma. Hierarchical Deep Neural Network for Mental Stress State Detection Using IoT-Based Biomarkers. *Pattern Recognition Letters*, 145:81–87, 2021.

[135] J. S. Kumar and P. Bhuvaneswari. Analysis of Electroencephalography (EEG) Signals and Its Categorization—A Study. *Procedia Engineering*, 38:2525–2536, 2012.

[136] S. Lagrosen and Y. Lagrosen. Workplace Stress and Health—The Connection to Quality Management. *Total Quality Management & Business Excellence*, 33(1-2):113–126, 2022.

[137] S. Lee et al. Detection of Drivers' Anxiety Invoked by Driving Situations Using Multimodal Biosignals. *Processes*, 8(2):155, 2020.

[138] S. Lee and D. Park. Abnormal Beat Detection From Unreconstructed Compressed Signals Based on Linear Approximation in ECG Signals Suitable for Embedded IoT Devices. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–13, 2022.

[139] P. Lehrer et al. Heart Rate Variability Biofeedback Improves Emotional and Physical Health and Performance: A Systematic Review and Meta-Analysis. *Applied Psychophysiology and Biofeedback*, 45:109–129, 2020.

[140] Z. Li et al. The Recognition of Multiple Anxiety Levels Based on Electroencephalograph. *IEEE Transactions on Affective Computing*, 2019.

[141] K. Liu et al. Automated Feature Selection: A Reinforcement Learning Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):2272–2284, 2021.

[142] L. Liu et al. CorrDQN-FS: A Two-Stage Feature Selection Method for Energy Consumption Prediction via Deep Reinforcement Learning. *Journal of Building Engineering*, 80:108044, 2023.

[143] W. Liu et al. Enhancing Generalized Anxiety Disorder Diagnosis Precision: MSTCNN Model Utilizing High-Frequency EEG Signals. *Frontiers in Psychiatry*, 14:1310323, 2023.

[144] D. Lopez-Martinez, N. El-Haouij, and R. Picard. Detection of Real-World Driving-Induced Affective State Using Physiological Signals and Multi-View

Multi-Task Machine Learning. In *Proceedings of the 8th IEEE International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 356–361, 2019.

[145] U. Lueken et al. Neurobiological Markers Predicting Treatment Response in Anxiety Disorders: A Systematic Review and Implications for Clinical Application. *Neuroscience & Biobehavioral Reviews*, 66:143–162, 2016.

[146] W. Maier et al. The Hamilton Anxiety Scale: Reliability, Validity and Sensitivity to Change in Anxiety and Depressive Disorders. *Journal of Affective Disorders*, 14(1):61–68, 1988.

[147] D. Makowski et al. NEUROKIT2: A Python Toolbox for Neurophysiological Signal Processing. *Behavior Research Methods*, pages 1–8, 2021.

[148] L. Malviya and S. Mal. A Novel Technique for Stress Detection From EEG Signal Using Hybrid Deep Learning Model. *Neural Computing and Applications*, 34(22):19819–19830, 2022.

[149] R. Markiewicz, A. Markiewicz-Gospodarek, and B. Dobrowolska. Galvanic Skin Response Features in Psychiatry and Mental Disorders: A Narrative Review. *International Journal of Environmental Research and Public Health*, 19(20):13428, 2022.

[150] V. Markova, T. Ganchev, and K. Kalinkov. CLAS: A Database for Cognitive Load, Affect and Stress Recognition. In *Proceedings of the IEEE International Conference on Biomedical Innovations and Applications (BIA)*, pages 1–4, 2019.

[151] A. J. Marthinsen et al. Psychological Stress Detection With Optimally Selected EEG Channel Using Machine Learning Techniques. 2023.

[152] K. Masood and M. A. Alghamdi. Modeling Mental Stress Using a Deep Learning Framework. *IEEE Access*, 7:68446–68454, 2019.

[153] P. Mathur, S. Kaistha, and V. K. Chakka. Mental Task-Induced Stress Detection Using Multivariate Weighted Visibility Graph (MV-WVG) From EEG Signals.

In *Proceedings of the 20th IEEE India Council International Conference (INDICON)*, pages 1265–1270, 2023.

[154] R. M. Mehmood and H. J. Lee. A Novel Feature Extraction Method Based on Late Positive Potential for Emotion Recognition in Human Brain Signal Patterns. *Computers & Electrical Engineering*, 53:444–457, 2016.

[155] M. Memar and A. Mokaribolhassan. Stress Level Classification Using Statistical Analysis of Skin Conductance Signal While Driving. *SN Applied Sciences*, 3(1):1–9, 2021.

[156] N. Milstein and I. Gordon. Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. *Frontiers in Behavioral Neuroscience*, 14:148, 2020.

[157] J. Minguillon et al. Blue Lighting Accelerates Post-Stress Relaxation: Results of a Preliminary Study. *PLOS ONE*, 12(10):e0186399, 2017.

[158] J. Minguillon et al. Portable System for Real-Time Detection of Stress Level. *Sensors*, 18(8):2504, 2018.

[159] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo. Stress Assessment by Prefrontal Relative Gamma. *Frontiers in Computational Neuroscience*, 10:101, 2016.

[160] D. Miranda et al. Naturalistic Enactment to Elicit and Recognize Caregiver State Anxiety. *Journal of Medical Systems*, 40(9):1–7, 2016.

[161] S. Mittal et al. How Can Machine Learning Be Used in Stress Management: A Systematic Literature Review of Applications in Workplaces and Education. *International Journal of Information Management Data Insights*, 2(2):100110, 2022.

[162] H. Mo et al. A Multimodal Data-Driven Framework for Anxiety Screening. *IEEE Transactions on Instrumentation and Measurement*, 2024.

[163] R. Mohan and S. Perumal. Classification and Detection of Cognitive Disorders Like Depression and Anxiety Utilizing Deep Convolutional Neural Network (CNN) Centered on EEG Signal. *Traitement du Signal*, 40(3), 2023.

[164] L. S. Mokatren et al. EEG Classification Based on Image Configuration in Social Anxiety Disorder. In *Proceedings of the 9th IEEE International EMBS Conference on Neural Engineering (NER)*, pages 577–580, 2019.

[165] F. Muhammad and S. Al-Ahmadi. Human State Anxiety Classification Framework Using EEG Signals in Response to Exposure Therapy. *Plos One*, 17(3):e0265679, 2022.

[166] C. Mühl, C. Jeunet, and F. Lotte. EEG-Based Workload Estimation Across Affective Contexts. *Frontiers in Neuroscience*, 8:114, 2014.

[167] S. B. Mulkey and A. J. du Plessis. Autonomic Nervous System Development and Its Impact on Neuropsychiatric Outcome. *Pediatric Research*, 85(2):120–126, 2019.

[168] G. R. Müller-Putz. Electroencephalography. In *Handbook of Clinical Neurology*, volume 168, pages 249–262. Elsevier, 2020.

[169] M. U. Mustafa et al. Machine Learning Approach for Multi-Class Stress Assessment With Electroencephalography Signals. *Journal of Population Therapeutics and Clinical Pharmacology*, pages 1464–1478, 2023.

[170] K. Na, S. Cho, and S. Cho. Machine Learning-Based Discrimination of Panic Disorder From Other Anxiety Disorders. *Journal of Affective Disorders*, 278:1–4, 2021.

[171] M. Naeem et al. Wearable ECG Systems for Accurate Mental Stress Detection: A Scoping Review. *Journal of Public Health*, pages 1–17, 2023.

[172] P. Nagar and D. Sethia. Brain Mapping-Based Stress Identification Using Portable EEG-Based Device. In *Proceedings of the 11th IEEE International Conference on Communication Systems & Networks (COMSNETS)*, pages 601–606, 2019.

[173] E. Najafpour et al. Can Galvanic Skin Conductance Be Used as an Objective Indicator of Children's Anxiety in the Dental Setting? *Journal of Clinical and Experimental Dentistry*, 9(3):e377, 2017.

[174] M. Nakajima et al. Using Novel Mobile Sensors to Assess Stress and Smoking Lapse. *International Journal of Psychophysiology*, 158:411–418, 2020.

[175] R. K. Nath and H. Thapliyal. Machine Learning-Based Anxiety Detection in Older Adults Using Wristband Sensors and Context Feature. *SN Computer Science*, 2(5):1–12, 2021.

[176] R. K. Nath, H. Thapliyal, and A. Caban-Holt. Machine Learning-Based Stress Monitoring in Older Adults Using Wearable Sensors and Cortisol as Stress Biomarker. *Journal of Signal Processing Systems*, pages 1–13, 2022.

[177] J. J. Newson and T. C. Thiagarajan. EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies. *Frontiers in Human Neuroscience*, 12:521, 2019.

[178] V. Nguyen. Bayesian Optimization for Accelerating Hyperparameter Tuning. In *Proceedings of the 2nd IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 302–305, 2019.

[179] P. Ogden. The Different Impact of Trauma and Relational Stress on Physiology, Posture, and Movement: Implications for Treatment. *European Journal of Trauma & Dissociation*, 5(4):100172, 2021.

[180] P. B. Pankajavalli and G. S. Karthick. A Unified Framework for Stress Forecasting Using Machine Learning Algorithms. In *Proceedings of the Springer International Conference on Advances in Computational Intelligence and Informatics (ACI)*, pages 199–210, 2020.

[181] M. Patanè et al. Prevalence of Mental Disorders in Refugees and Asylum Seekers: A Systematic Review and Meta-Analysis. *Global Mental Health*, pages 1–14, 2022.

[182] G. Pei et al. Neural Evidence of Face Processing in Social Anxiety Disorder: A Systematic Review With Meta-Analysis. *Neuroscience & Biobehavioral Reviews*, page 105283, 2023.

[183] R. M. Pereira, Y. M. G. Costa, and N. S. J. Carlos. MLTL: A Multi-Label Approach for the Tomek Link Undersampling Algorithm. *Neurocomputing*, 383:95–105, 2020.

[184] S. L. Perrin et al. Waking qEEG to Assess Psychophysiological Stress and Alertness During Simulated On-Call Conditions. *International Journal of Psychophysiology*, 141:93–100, 2019.

[185] N. Phutela et al. Effectiveness of Higuchi Fractal Dimension in Differentiating Subgroups of Stressed and Non-Stressed Individuals. *Multimedia Tools and Applications*, 83(17):52433–52450, 2024.

[186] M. Plechawska-Wójcik et al. A Three-Class Classification of Cognitive Workload Based on EEG Spectral Data. *Applied sciences*, 9(24):5340, 2019.

[187] T. H. Priya et al. Stress Detection From EEG Using Power Ratio. In *Proceedings of the IEEE International Conference on Emerging Trends in Information Technology and Engineering (IC-ETITE)*, pages 1–6, 2020.

[188] A. Puli and A. Kushki. Toward Automatic Anxiety Detection in Autism: A Real-Time Algorithm for Detecting Physiological Arousal in the Presence of Motion. *IEEE Transactions on Biomedical Engineering*, 67(3):646–657, 2019.

[189] P. Putman et al. EEG Theta/Beta Ratio in Relation to Fear-Modulated Response-Inhibition, Attentional Control, and Affective Traits. *Biological Psychology*, 83(2):73–78, 2010.

[190] K. Radhika and V. R. M. Oruganti. Deep Multimodal Fusion for Subject-Independent Stress Detection. In *Proceedings of the 11th IEEE International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 105–109, 2021.

[191] K. Radhika and V. R. M. Oruganti. Stress Detection Using CNN Fusion. In *Proceedings of the TENCON 2021 IEEE Region 10 Conference (TENCON)*, pages 492–497, 2021.

[192] A. A. Rahman et al. Detection of Mental State From EEG Signal Data: An Investigation With Machine Learning Classifiers. In *Proceedings of the 14th IEEE International Conference on Knowledge and Smart Technology (KST)*, pages 152–156, 2022.

[193] T. Rahman et al. QCovSML: A Reliable COVID-19 Detection System Using CBC Biomarkers by a Stacking Machine Learning Model. *Computers in Biology and Medicine*, 143:105284, 2022.

[194] V. G. Rajendran, S. Jayalalitha, and K. Adalarasu. EEG-Based Evaluation of Examination Stress and Test Anxiety Among College Students. *IRBM*, 2021.

[195] V. V. N. Raju et al. Enhancing Emotion Prediction Using Deep Learning and Distributed Federated Systems With SMOTE Oversampling Technique. *Alexandria Engineering Journal*, 108:498–508, 2024.

[196] A. Riener, A. Ferscha, and M. Aly. Heart on the Road: HRV Analysis for Monitoring a Driver's Affective State. In *Proceedings of the 1st ACM International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 99–106, 2009.

[197] J. Rodríguez-Arce et al. Towards an Anxiety and Stress Recognition System for Academic Environments Based on Physiological Features. *Computer Methods and Programs in Biomedicine*, 190:105408, 2020.

[198] M. Rose and J. Devine. Assessment of Patient-Reported Symptoms of Anxiety. *Dialogues in Clinical Neuroscience*, 2022.

[199] B. Roy et al. Hybrid Deep Learning Approach for Stress Detection Using Decomposed EEG Signals. *Diagnostics*, 13(11):1936, 2023.

[200] S. Sadruddin, V. D. Khairnar, and D. R. Vora. Issues and Challenges in Detecting

Mental Stress From Multimodal Data Using Machine Intelligence. *SN Computer Science*, 5(4):358, 2024.

[201] M. F. Safdar, R. M. Nowak, and P. Pałka. Pre-Processing Techniques and Artificial Intelligence Algorithms for Electrocardiogram (ECG) Signal Analysis: A Comprehensive Review. *Computers in Biology and Medicine*, 170:107908, 2024.

[202] S. Saganowski et al. Emotion Recognition for Everyday Life Using Physiological Signals From Wearables: A Systematic Literature Review. *IEEE Transactions on Affective Computing*, 14(3):1876–1897, 2022.

[203] M. Salai, I. Vassányi, and I. Kósa. Stress Detection Using Low-Cost Heart Rate Sensors. *Journal of Healthcare Engineering*, 2016, 2016.

[204] N. Salankar et al. Impact of Music in Males and Females for Relief From Neurodegenerative Disorder Stress. *Contrast Media & Molecular Imaging*, 2022(1):3080437, 2022.

[205] N. Salankar, D. Koundal, and S. M. Qaisar. Stress Classification by Multimodal Physiological Signals Using Variational Mode Decomposition and Machine Learning. *Journal of Healthcare Engineering*, 2021(1):2146369, 2021.

[206] N. Salankar and S. M. Qaisar. EEG Based Stress Classification by Using Difference Plots of Variational Modes and Machine Learning. *Journal of Ambient Intelligence and Humanized Computing*, 14(12):16347–16360, 2023.

[207] J. Šalkevicius et al. Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals. *Electronics*, 8(9):1039, 2019.

[208] C. Samson and A. Koh. Stress Monitoring and Recent Advancements in Wearable Biosensors. *Frontiers in Bioengineering and Biotechnology*, 8:1037, 2020.

[209] R. Sánchez-Reolid, M. T. López, and A. Fernández-Caballero. Machine Learning for Stress Detection From Electrodermal Activity: A Scoping Review. *Preprints*, 2020.

[210] V. Sandulescu et al. Stress Detection Using Wearable Physiological Sensors. In *Proceedings of the Springer International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC)*, pages 526–532, 2015.

[211] M. M. Sani et al. Support Vector Machine for Classification of Stress Subjects Using EEG Signals. In *Proceedings of the IEEE Conference on Systems, Process and Control (ICSPC)*, pages 127–131, 2014.

[212] S. Sawangarreerak and P. Thanathamathee. Random Forest With Sampling Techniques for Handling Imbalanced Prediction of University Student Depression. *Information*, 11(11):519, 2020.

[213] F. Scarpina and S. Tagini. The Stroop Color and Word Test. *Frontiers in Psychology*, 8:557, 2017.

[214] D. L. Schacter. EEG Theta Waves and Psychological Phenomena: A Review and Analysis. *Biological Psychology*, 5(1):47–82, 1977.

[215] E. Schleiger et al. Frontal EEG Delta/Alpha Ratio and Screening for Post-stroke Cognitive Deficits: The Power of Four Electrodes. *International Journal of Psychophysiology*, 94(1):19–24, 2014.

[216] P. Schmidt et al. Introducing WESAD: A Multimodal Dataset for Wearable Stress and Affect Detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI)*, pages 400–408, 2018.

[217] N. Schneiderman, G. Ironson, and S. D. Siegel. Stress and Health: Psychological, Behavioral, and Biological Determinants. *Annual Review of Clinical Psychology*, 1:607–628, 2005.

[218] R. Sebastião. Classification of Anxiety Based on EDA and HR. In *Proceedings of the Springer International Conference on IoT Technologies for HealthCare (IoT HealthCare)*, pages 112–123, 2020.

[219] H. Selye. The Stress Syndrome. *The American Journal of Nursing*, pages 97–99, 1965.

[220] S. Seo, Y. Gil, and J. Lee. The Relation Between Affective Style of Stressor on EEG Asymmetry and Stress Scale During Multimodal Task. In *Proceedings of the 3rd IEEE International Conference on Convergence and Hybrid Information Technology (ICHIT)*, pages 461–466, 2008.

[221] W. Seo et al. Deep Learning Approach for Detecting Work-Related Stress Using Multimodal Signals. *IEEE Sensors*, 22(12):11892–11902, 2022.

[222] C. Setz et al. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):410–417, 2009.

[223] L. D. Sharma et al. Evolutionary Inspired Approach for Mental Stress Detection Using EEG Signal. *Expert Systems with Applications*, 197:116634, 2022.

[224] N. Sharma and T. Gedeon. Objective Measures, Sensors and Computational Techniques for Stress Recognition and Classification: A Survey. *Computer Methods and Programs in Biomedicine*, 108(3):1287–1301, 2012.

[225] R. Sharma and K. Chopra. EEG Signal Analysis and Detection of Stress Using Classification Techniques. *Journal of Information and Optimization Sciences*, 41(1):229–238, 2020.

[226] M. Sharpe. New Oxford Textbook of Psychiatry, Vols 1 and 2, 2004.

[227] M. Shim et al. Machine-Learning-Based Classification Between Post-traumatic Stress Disorder and Major Depressive Disorder Using P300 Features. *NeuroImage: Clinical*, 24:102001, 2019.

[228] L. Shu et al. A Review of Emotion Recognition Using Physiological Signals. *Sensors*, 18(7):2074, 2018.

[229] A. I. Siam, S. A. Gamel, and F. M. Talaat. Automatic Stress Detection in Car Drivers Based on Non-invasive Physiological Signals Using Machine Learning Techniques. *Neural Computing and Applications*, 35(17):12891–12904, 2023.

[230] S. H. Siddiqi et al. Causal Mapping of Human Brain Function. *Nature Reviews Neuroscience*, 23(6):361–375, 2022.

[231] M. H. F. Siddiqui, D. Inkpen, and A. Gelbukh. Towards Interpretable Emotion Classification: Evaluating LIME, SHAP, and Generative AI for Decision Explanations. In *Proceedings of the 28th IEEE International Conference Information Visualisation (IV)*, pages 1–6, 2024.

[232] D. Y. Y. Sim and C. K. Chong. Effects of Dimension Reduction Methods on Boosting Algorithms for Better Prediction Accuracies on Classifications of Stress EEGs. In *Proceedings of the 6th IEEE International Conference on Electronics and Electrical Engineering Technology (EEET)*, pages 49–54, 2023.

[233] S. Simske. Introduction, Overview, and Applications. *Meta-Analytics*, pages 1–98, 2019.

[234] J. Singh and M. A. Hamid. Cognitive Computing in Mental Healthcare: A Review of Methods and Technologies for Detection of Mental Disorders. *Cognitive Computation*, pages 1–18, 2022.

[235] M. Singh and A. B. Queyam. Correlation Between Physiological Parameters of Automobile Drivers and Traffic Conditions. *International Journal of Electronics Engineering*, 5(2):6–12, 2013.

[236] C. D. Spielberger et al. The State-Trait Anxiety Inventory. *Revista Interamericana de Psicologia / Interamerican Journal of Psychology*, 5(3 & 4), 1971.

[237] P. Srinivas and R. Katarya. hyOPTXg: OPTUNA Hyperparameter Optimization Framework for Predicting Cardiovascular Disease Using XGBoost. *Biomedical Signal Processing and Control*, 73:103456, 2022.

[238] F. Sun et al. Activity-Aware Mental Stress Detection Using Physiological Sensors. In *Proceedings of the Springer International Conference on Mobile Computing, Applications, and Services (MobiCASE)*, pages 282–301, 2010.

[239] R. Suryawanshi and S. Vanjale. Brain Activity Monitoring for Stress Analysis Through EEG Dataset Using Machine Learning. *International Journal of Intelligent Systems and Applications in Engineering*, 11(1s):236–240, 2023.

[240] D. Talia, P. Trunfio, and F. Marozzo. Introduction to Data Mining. *Data Analysis in the Cloud*, pages 1–25, 2016.

[241] J. Tang, S. Alelyani, and H. Liu. Feature Selection for Classification: A Review. *Data classification: Algorithms and Applications*, page 37, 2014.

[242] J. Tejedor et al. Multiple Physiological Signals Fusion Techniques for Improving Heartbeat Detection: A Review. *Sensors*, 19(21):4708, 2019.

[243] H. Tiwari et al. Temporal Lobe Structures and Their Role in Psychiatric Manifestation. *Santosh University Journal of Health Sciences*, 10(1):64–69, 2024.

[244] R. K. Tripathy et al. Detection of Different Stages of Anxiety From Single-Channel Wearable ECG Sensor Signal Using Fourier-Bessel Domain Adaptive Wavelet Transform. *IEEE Sensors Letters*, 2023.

[245] R. Turner et al. Bayesian Optimization Is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020. In *Proceedings of the NeurIPS 2020 Competition and Demonstration Track*, pages 3–26, 2021.

[246] N. V. D. Vinne et al. Frontal Alpha Asymmetry As A Diagnostic Marker in Depression: Fact or Fiction? A Meta-Analysis. *Neuroimage: Clinical*, 16:79–87, 2017.

[247] A. Varshney et al. Automated Classification of Mental Arithmetic Tasks Using Recurrent Neural Network and Entropy Features Obtained From Multi-Channel EEG Signals. *Electronics*, 10(9):1079, 2021.

[248] M. Vaz et al. Multimodal Classification of Anxiety Based on Physiological Signals. *Applied Sciences*, 13(11):6368, 2023.

[249] J. R. Vergara and P. A. Estévez. A Review of Feature Selection Methods Based on Mutual Information. *Neural Computing and Applications*, 24:175–186, 2014.

[250] L. Vitetta et al. Mind-Body Medicine: Stress and Its Impact on Overall Health and Longevity. *Annals of the New York Academy of Sciences*, 1057(1):492–505, 2005.

[251] A. Vulpe-Grigoraşi and O. Grigore. A Neural Network Approach for Anxiety Detection Based on ECG. In *Proceedings of the IEEE International Conference on E-Health and Bioengineering (EHB)*, pages 1–4, 2021.

[252] Z. Wan et al. A Review on Transfer Learning in EEG Signal Analysis. *Neurocomputing*, 421:1–14, 2021.

[253] Q. Wang and O. Sourina. Real-Time Mental Arithmetic Task Recognition From EEG Signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(2):225–232, 2013.

[254] S. Wang et al. Advances in Data Preprocessing for Biomedical Data Fusion: An Overview of the Methods, Challenges, and Prospects. *Information Fusion*, 76:376–421, 2021.

[255] Z. Wang, S. Hu, and H. Song. Channel Selection Method for EEG Emotion Recognition Using Normalized Mutual Information. *IEEE Access*, 7:143303–143311, 2019.

[256] H. Wei, L. Chang, Q. Huang, and R. Zhou. Relation Between Spontaneous Electroencephalographic Theta/Beta Power Ratio and Test Anxiety. *Neuroscience Letters*, 737:135323, 2020.

[257] T. Y. Wen and S. A. M. Aris. Hybrid Approach of EEG Stress Level Classification Using K-Means Clustering and Support Vector Machine. *IEEE Access*, 10:18370–18379, 2022.

[258] T. Y. Wen and S. M. Aris. Electroencephalogram (EEG) Stress Analysis on Alpha/Beta Ratio and Theta/Beta Ratio. *Indonesian Journal of Electrical Engineering and Computer Science*, 17(1):175–182, 2020.

[259] W. Wen et al. Toward Constructing a Real-Time Social Anxiety Evaluation

System: Exploring Effective Heart Rate Features. *IEEE Transactions on Affective Computing*, 11(1):100–110, 2018.

[260] A. Wessam, Y. Li, and P. Wen. K-Complexes Detection in EEG Signals Using Fractal and Frequency Features Coupled With an Ensemble Classification Model. *Neuroscience*, 422:119–133, 2019.

[261] J. Wijsman et al. Towards Mental Stress Detection Using Wearable Physiological Sensors. In *Proceedings of the IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, pages 1798–1801, 2011.

[262] I. H. Witten et al. Ensemble Learning. In *Encyclopedia of Machine Learning and Data Mining*, pages 479–501. Springer, 2017.

[263] D. H. Wolpert. Stacked Generalization. *Neural Networks*, 5(2):241–259, 1992.

[264] World Health Organization. COVID-19 Pandemic Triggers 25% Increase in Prevalence of Anxiety and Depression Worldwide. `https://shorturl.at/P9Cd6`, 2022. Last accessed: March 2, 2022.

[265] L. Xia, A. S. Malik, and A. R. Subhani. A Physiological Signal-Based Method for Early Mental-Stress Detection. *Biomedical Signal Processing and Control*, 46:18–32, 2018.

[266] O. Xie, Z. Liu, and X. Ding. Electroencephalogram Emotion Recognition Based On A Stacking Classification Model. In *Proceedings of the 37th IEEE Chinese Control Conference (CCC)*, pages 5544–5548, 2018.

[267] Y. Xie et al. Anxiety and Depression Diagnosis Method Based on Brain Networks and Convolutional Neural Networks. In *Proceedings of the 42nd IEEE Annual International Conference on Engineering in Medicine & Biology Society (EMBC)*, pages 1503–1506, 2020.

[268] R. Xiong et al. Pattern Recognition of Cognitive Load Using EEG and ECG Signals. *Sensors*, 20(18):5122, 2020.

[269] H. Xu et al. Research on EEG Channel Selection Method for Emotion Recognition. In *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 2528–2535, 2019.

[270] K. Yamada et al. Detection of Anxiety Expression From EEG Analysis Using Support Vector Machine. *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, 17(1):87–94, 2019.

[271] J. Yedukondalu and L. D. Sharma. Cognitive Load Detection Using Circulant Singular Spectrum Analysis and Binary Harris Hawks Optimization Based Feature Selection. *Biomedical Signal Processing and Control*, 79:104006, 2023.

[272] R. Zebari et al. A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1(1):56–70, 2020.

[273] B. Zhang et al. Reaction Time and Physiological Signals For Stress Recognition. *Biomedical Signal Processing and Control*, 38:100–107, 2017.

[274] J. Zhang et al. ReliefF-Based EEG Sensor Selection Methods For Emotion Recognition. *Sensors*, 16(10):1558, 2016.

[275] Q. Zhang et al. Respiration-Based Emotion Recognition With Deep Learning. *Computers in Industry*, 92:84–90, 2017.

[276] Y. Zhang et al. Investigating Different Stress-Relief Methods Using Electroencephalogram (EEG). In *Proceedings of the 42nd IEEE Annual International Conference on Engineering in Medicine & Biology Society (EMBC)*, pages 2999–3002, 2020.

[277] Y. Zheng et al. Unobtrusive and Multimodal Wearable Sensing to Quantify Anxiety. *IEEE Sensors*, 16(10):3689–3696, 2016.

[278] L. Zhu et al. Stress Detection Through Wrist-Based Electrodermal Activity Monitoring and Machine Learning. *IEEE Journal of Biomedical and Health Informatics*, 27(5):2155–2165, 2023.

[279] L. Zhu, P. Spachos, and S. Gregori. Multimodal Physiological Signals and Machine Learning for Stress Detection by Wearable Devices. In *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2022.

[280] I. Zyma et al. Electroencephalograms During Mental Arithmetic Task Performance. *Data*, 4(1):14, 2019.