"INTEGRATIVE COMPUTATIONAL APPROACHES FOR RECEPTOR-BASED DRUG DISCOVERY AND BIOMARKER IDENTIFICATION IN NDDs"

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

MASTERS OF TECHNOLOGY
in
BIOINFORMATICS

by
RISHI MRINAL
23/BIO/14

Under the Supervision of
Prof. PRAVIR KUMAR
Professor and Dean of Office of International Affairs (OIA), Delhi Technological University, Delhi

Department of Biotechnology

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

May, 2025

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

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

CANDIDATE'S DECLARATION

I **Rishi Mrinal** hereby certify that the work which is being presented in the thesis entitled **"Integrative Computational Approaches for Receptor-based Drug Discovery and Biomarker Identification in NDDs"** in partial fulfillment of the requirements for the award of the Degree of **Master of Technology**, submitted in the Department of Biotechnology, Delhi Technological University is an authentic record of my own work carried out during the period from January 2025 to May 2025 under the supervision of **Prof. Pravir Kumar**.

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.

Candidate's Signature

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

20/05/2025

Signature of Supervisor

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

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

CERTIFICATE BY THE SUPERVISOR

Certified that **Rishi Mrinal (23/BIO/14)** has carried out their search work presented in this thesis entitled **"Integrative Computational Approaches for Receptor-based Drug Discovery and Biomarker Identification in NDDs"** for the award of **Master of Technology** from Department of Biotechnology, 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.

20/05/2025

Prof. Pravir Kumar

Supervisor, Dean of OIA

Department of Biotechnology

Delhi Technological University

23.05.25

Prof. Yasha Hasija

Head of Department

Department of Biotechnology

Delhi Technological University

Head of the Department
Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engg.)
Bawana Road, Delhi-110042

Date: 20.05.2025

INTEGRATIVE COMPUTATIONAL APPROACHES FOR RECEPTOR-BASED DRUG DISCOVERY AND BIOMARKER IDENTIFICATION IN NDDs

RISHI MRINAL

ABSTRACT

Neurodegenerative disorders (NDDs) represent a significant global health burden, with overlapping clinical and molecular features that often complicate early diagnosis and treatment. This thesis presents an integrative computational approach to address key challenges in the therapeutic and diagnostic landscape of NDDs by bridging receptor-based drug discovery and gene expression-driven biomarker identification. The first objective focuses on targeting the GABRA2 receptor, a subunit of the GABA-A complex implicated in both anxiety and neurodegeneration. Using homology modelling, ligand-based virtual screening, molecular docking, and ADME profiling, the study identifies novel FDA-approved compounds with improved binding affinity and pharmacokinetic properties over the benchmark drug Diazepam. Zolmitriptan emerged as a promising candidate with high BBB permeability, favourable bioavailability, and minimized toxicity. The second objective employs machine learning models on high-throughput transcriptomic data (GSE140830) to classify dementia subtypes and identify key biomarkers. Random Forest, Support Vector Classifier, and other models achieved robust classification performance, while feature importance and pathway enrichment analyses revealed subtype-specific gene signatures linked to neuroinflammatory and synaptic pathways. By unifying molecular pharmacology with ML-driven omics analytics, this study provides a dual-framework for stratified therapeutic targeting and early diagnosis in NDDs, offering translational value for precision medicine.

Keywords: NDDs, Anxiety and Dementia, GABRA2, Molecular Docking, ADME Profiling, Virtual Screening, ML, Gene Expression Analysis. Subtype Classification, Biomarker Identification, Homology Modeling, Precision Medicine

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor, Prof. Pravir Kumar for his constant support, encouragement, and invaluable guidance throughout this research. I am sincerely thankful for the direction and insights he has provided, which significantly shaped the course of this research.

I also extend my gratitude to the Department of Biotechnology at Delhi Technological University for providing all the essential facilities and resources. Their support was crucial for the experimental work and overall progress of this study.

A special thanks goes to the PhD scholars Dr. Mehar Sahu, Dr. Neetu Rani, Mr. Rahul Tripathi, Ms. Shefali Kardam and Ms. Shrutikirti Vashishth, who supported throughout this journey and guided with everyday work. Their expertise and willingness to help were truly invaluable, and I am deeply grateful for their mentorship.

Lastly, I would like to thank my family and friends for their unwavering support and for making this journey more enjoyable. Their encouragement and camaraderie were essential in helping us support each other through this process.

TABLE OF CONTENTS

Title **Page no.**

Candidate's Declaration…….…….. (ii)

Certificate…… (iii)

Abstract…… (iv)

Acknowledgement…… (v)

Contents…… (vi-viii)

List of Figures…… (ix)

List of Tables……. (x)

List of Abbreviations…… (xi-xii)

CHAPTER – 1

INTRODUCTION…… 1-2

CHAPTER – 2

LITERATURE REVIEW……. 3-19

2.1. Introduction to NDDs…….. 3

2.2. Anxiety and its Intersection with Neurodegeneration……. 3-4

2.2.1. Anxiety and its Types…….. 4

2.2.2. Anxiety as a comorbidity and prodromal symptom in Dementia.... 5

2.2.3. Impact of Anxiety in Life……..... 5

2.3. Bridging Dementia Subtypes with Neurodegenerative Mechanisms…. 5

2.3.1. Dementia and its Subtypes……. 5-6

2.3.2. Dementia Subtypes and Its integration with Neurodegeneration.... 6-7

2.4. Risk Factors and Symptoms of NDDs…….. 7-8

2.5. GABAergic Signalling and the Role of GABRA2…….. 8-9

2.5.1. Structure and Physiological Function of the GABA-A Receptor... 9

2.5.2. Role of GABRA2…….. 9-11

2.5.3. Therapeutic Potential of GABRA2…….. 11
2.6. Receptor-Based Drug Discovery in NDDs……..... 11-12
2.6.1. Homology modelling of GABRA2: Rationale and methods…... 12
2.6.2. Molecular Docking and ADME profiling……. 12-13
2.6.3. BBB Permeability and Drug-Repurposing……... 13-15
2.7. Machine Learning in NDD Diagnosis and Biomarker Discovery…….. 15
2.7.1. High-throughput Transcriptomics in Dementia…… 15
2.7.2. ML Models for Subtype Classification……..... 15-16
2.7.3. Feature Selection and Model Interpretability…… 16-17
2.7.4. Pathway Enrichment and Functional Relevance of Biomarkers... 17-18
2.8. Integration of Computational Drug Design and Biomarker Discovery… 18-19

CHAPTER – 3

METHODOLOGY…… 20-28
3.1. Computational Resources and Outline…… 20-22
3.2. Objective 1: Receptor-based Drug Discovery…… 22-24
3.2.1. Data Extraction……. 22
3.2.2. Homology Modeling…… 22-23
3.2.3. Target Receptor Preparation……..... 23
3.2.4. Ligand Molecules Preparation……... 23
3.2.5. Molecular Docking Studies……... 23-24
3.2.6. ADME Analysis……..... 24
3.2.7. Analysis and Visualisation……..... 24
3.2.8. Prediction of Biological Activity for Substances (PASS Analysis)…. 24
3.3. Objective 2: Gene-expression based Biomarker Discovery……. 25-28
3.3.1. Data Acquisition…… 25
3.3.2. Data Preprocessing…… 25
3.3.3. Feature Engineering and Dimensionality Reduction Methods….. 25
3.3.4. Machine Learning and Model Training……. 26

3.3.5. Hyperparameter Tuning and Model Optimisation…… 26

3.3.6. Model Evaluation and Cross-validation Strategy…….. 26-27

3.3.7. Biomarker Identification……..... 27

3.3.8. Pathway Analysis Integration…….. 28

CHAPTER – 4

RESULTS AND DISCUSSION…….. 29-50

4.1. Homology Modeling of GABRA2 gene…… 29-30

4.2. Ligand-based Virtual Screening……. 31-35

4.3. Interaction Analysis……. 36-41

4.4. Pharmacokinetic (ADMET) Profiling……. 41-42

DISCUSSION (OBJECTIVE – 1)…….. 43

4.5. Data Collection…… 44

4.6. Data Preprocessing…….... 44

4.7. ML Implementation and Model Evaluation……... 44-47

4.8. Biomarker Identification Analysis……. 47-48

4.9. Pathway Enrichment Analysis of Identified Biomarkers……... 48-49

DISCUSSION (OBJECTIVE – 2)…… 50

CHAPTER – 5

CONCLUSION AND FUTURE PERSPECTIVES……. 51-52

CHAPTER – 6

REFERENCES…… 53-66

List of Publications and their proofs……. 67-71

Plagiarism Verification……... 72

Plagiarism Report……... 73

LIST OF FIGURES

FIGURE NO.	TITLE OF FIGURES	PAGE NO.
Fig 2.1.	Classification of Anxiety disorder	4
Fig 2.2.	Four different subtypes of dementia	6
Fig 2.3.	Risk factors including, age, genetic mutations, environmental exposures, and lifestyle-related elements.	8
Fig 2.4.	Displaying the GABA domains of GABA-A Receptor, representing spatial arrangement of α-helices, β-sheets, and loop regions, emphasizing the conformational folding critical to its function.	10
Fig 3.3.1.	Pipeline of the methodology	27
Fig 4.1.1.	3-D model retrieved for GABRA2 by Homology Modeling	29
Fig 4.1.2.	Per-residue local QMEAN score plot	30
Fig 4.1.3.	Global QMEAN Z-score comparison with PDB reference structures	30
Fig 4.1.4.	Ramachandran Plot showing 91.98% with favoured regions	30
Fig 4.2.1.	Molecular Docked Complex of GABRA2-Diazepam (Control)	34
Fig 4.2.2.	Molecular Docked Complex of GABRA2-Zolmitriptan	34
Fig 4.2.3.	Molecular Docked Complex of GABRA2-Riluzole	35
Fig 4.2.4.	Molecular Docked Complex of GABRA2-Perampanel	35
Fig 4.3.	2-D structural representation of GABRA2 residues interacting to the (a) Diazepam, a reference drug (b) Zolmitriptan (c) Riluzole (d) Perampanel (e) Lormetazepam (f) Tranylcypromine	38-41
Fig 4.6.	Comparison of data distributions before and after quantile normalization	45
Fig 4.7.	Comparative analysis of confusion matrices across 5 different machine learning classifiers: Logistic Regression, SVC (Linear Kernel), Naive Bayes, MLP Classifier, and Random Forest. Each matrix displays the classification performance across five classes (CES, Control, PSP, bvFTD, nfvPPA).	46
Fig 4.8.	Top 20 feature importance scores derived from ML analysis	48
Fig 4.9.	KEGG Pathway Enrichment Analysis showing significantly enriched biological pathways ranked by adjusted p-values	49

LIST OF TABLES

TABLE NO.	TITLE OF TABLES	PAGE NO.
Table 4.1.	Binding Affinity of Top 15 Ligands in accordance of the Reference Drug	31-33
Table 4.2.	Interaction Networks of GABRA2 and Drugs	36-37
Table 4.4.	Physiochemical Properties of the top-five ligand molecules with highest binding affinity	41-42
Table 5.1.	Classification Performance of Machine Learning Models	47

LIST OF ABBREVIATIONS

NDD	Neurodegenerative Disorder
GABA-A	Gamma Amino Butyric Acid - A
GABRA2	Gamma Amino Butyric Acid – α2
BBB	Blood Barin Barrier
RF	Random Forest
SVC	Support Vector Classifier
LR	Logistic Regression
MLP	Multi-Layer Perceptron
NB	Naïve Bayes’
ML	Machine Learning
AD	Alzheimer’s Disease
PD	Parkinson’s Disease
HD	Huntington’s Disease
ALS	Amyotrophic Lateral Sclerosis
FTD	Frontotemporal Dementia
FTLD	Frontotemporal lobal degeneration
CBS	Corticobasal Syndrome
PSP	Progressive Supranuclear Palsy
PPA	Primary Progressive Aphasia
GWAS	Genome-wide Association Studies
MDS	Molecular Dynamics Simulation
MD	Molecular Docking
CNS	Central Nervous System
MAPT	Microtubule-associated tau protein
LGICs	Ligand-gated Cl- channels
SNP	Single Nucleotide Polymorphism
GAD	Generalised Anxiety Disorder

DEGs	Differential Expressed Genes
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
CADD	Computer-aided drug Design
DR	Drug Repurposing
HTS	High-throughput Screening
VS	Virtual Screening
SBDD	Structure-based Drug Design
RBF	Radial Basis Function
PK	Polynomial Kernels
UPS	Ubiquitin-proteasome system
DL	Deep Learning
DHA	Docosahexanoic Acid
SDF	Structure Data File
GI	Gastrointestinal Absorption

CHAPTER – 1

INTRODUCTION

Neurodegenerative disorders (NDDs) are considered as the prominent class of progressive and debilitating neurological conditions leading to the gradual loss of neuronal function i.e., cognitive, motor, and/or sensory functions in the specific region of brain and spinal cord [1]. With the increasing breakthroughs in genetic, behavioural, and neurobiological conditions, NDDs are the global health concern including prominent examples like Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), and Amyotrophic Lateral Sclerosis (ALS), each disease contributing significantly to mortality and morbidity worldwide. According to the WHO factsheets, as the global burden of these disorders are rising dramatically affecting 55 million people, understanding the exact cause is far more crucial for the improvement of more effective diagnostic tools, prevention, treatment and therapeutic strategies [2] [3].

There are several representative spectrums of NDDs including Frontotemporal Dementia (FTD), Progressive Supranuclear Palsy (PSP), Corticobasal Syndrome (CBS), Primary Progressive Aphasia (PSA), Anxiety, Depression, Epilepsy and more, placing increased pressure on healthcare systems worldwide. By 2050, dementia alone is predicted to affect more than 150 million people globally, emphasizing the need for innovative therapeutic strategies [4]. Central to the pathology of many NDDs, is the dysfunction and dysregulation of neurotransmitter signalling and synaptic loss of neuronal receptors and aberrant gene regulation [5]. Receptors such as, Nicotinic Acetylcholine Receptors (nAChRs), Gamma-amino Butyric Acid type A receptor (GABA-A), and Dopamine Receptors have been implicated in mechanism of diseases, influencing neuroplasticity, synaptic transmissions, and cognitive functions [6] [7]. Among these, the α-2 subunit of GABA-A, GABRA2 acts as a dual- action therapeutic target which addresses both the symptoms of anxiety and neurodegenerative processes, playing a central role in the inhibitory signalling within the brain. Alterations in GABAergic neurotransmission have been implicated in early-stage cognitive deficits, making GABRA2 a compelling target [8] [9]. Simultaneously, alterations in the gene expression within affected regions of brain provides molecular signatures aiding in early diagnosis and therapeutic target identification. At the molecular level, NDDs commonly exhibit the accumulation of misfolded proteins, mitochondrial dysfunction, oxidative stress, impaired axonal transport, and neuroinflammation. Genetic factors significantly contribute to the pathogenesis of several NDDs, either through genetic susceptibility loci identified in Genome-wide Association studies (GWAS) or through inherited mutations [10] [11] [12].

Recent advancements in computational biology, combined with Machine Learning (ML) approaches, have revolutionized new approaches to drug discovery, biomarker identification, and receptor targeting for different NDDs to develop more targeted, effective, and personalized therapeutic strategies. In silico techniques such as, Homology Modeling, Virtual Screening, Molecular Docking, Molecular Docking Simulations (MDS), and ML-integration with the gene expression analyses enables rapid identification of, high-affinity compounds, promising drug candidates with higher accuracy, classification of disease subtypes with remarkable precision, novel disease-specific biomarkers and stratify subtypes. The integration of omics with ML techniques provides a framework for uncovering molecular and gene-regulatory pathways involved in NDDs [13]. Simultaneously, advancements in MD are enhancing the accuracy of binding predictions (affinity), and validating receptor-ligand, ligand-receptor, receptor-receptor interactions, supported by molecular dynamic simulation (MDS) tools, GROMACS [14].

Early diagnoses of NDDs are critical for improving the patient outcomes based on the global reliability, facilitating the development of disease-modifying therapies. However, traditional methods have a large concern on clinical evaluations, neuroimaging and neuropsychological testing, often detect disease only after significant neuropathological changes have occurred. Current therapeutics combined with sophisticated computational methods, accelerates to identify biomarkers in the clinical management of NDDs [15] [16] [17]. They can aid in differential diagnosis, stratify patient populations for clinical trials, and serve as surrogate endpoints for the therapeutic efficacies [18]. Early diagnosis through biomarker identification improves patient care, accelerating the development of targeted therapies [19]. Current methodologies not only accelerate drug discovery but also offer insights into structural biology, binding energetics, and target specificity [20].

This thesis aims to explore an integrative computational approach combining receptor-based drug discovery and ML-driven biomarker identification to address the challenges in the management of NDDs. The thesis specifically targets two major objectives:

Objective 1

Receptor-based drug discovery: Modelling the GABRA2 receptor using Homology modelling, identifying the novel inhibitors through ligand-based Virtual screening, Molecular docking, followed by ADME profiling contextualising anxiety.

Objective 2

Gene expression-based biomarker discovery: Classification of dementia subtypes using ML-models applied to high-throughput gene expression datasets and identifying key biomarkers and the biological relevance through feature importance, pathway enrichment analysis.

CHAPTER -2

LITERATURE REVIEW

2.1. Introduction to NDDs

NDDs encompasses a group of chronic neurological conditions which are marked by the integral degeneration of both functional and structural components of the Central Nervous System (CNS). It is considered as an "umbrella" term for neurological diseases like AD, PD, HD, which ultimately have dominated much of the research landscape. Meanwhile, the significant attention is shifting towards non-AD dementias including bvFTD, PSP, CBS and PPA, relying under a group of Frontotemporal lobar degeneration (FTLD), with overlapping symptoms that complicate early and accurate diagnosis [21] [22].

Concurrently, non-cognitive symptoms such as anxiety disorders highlights another major feature of neurodegeneration, characterized by hyperarousal, excessive worry, and behavioural changes. Although, anxiety usually considered as a neuropsychiatric disorder rather than neurodegenerative, it has been increasingly recognised as both a prodromal symptom as well as a comorbidity in multiple NDDs [23]. These affective changes are often under-recognized but may reflect underlying neurochemical disruptions, including GABAergic dysfunction, which could offer novel molecular targets for intervention. In many patients, anxiety symptoms appear years before cognitive decline becomes clinically detectable, suggesting a possible shared neurobiological substrate. Furthermore, persistent anxiety in patients with dementia has been associated with faster cognitive deterioration, greater caregiver burden, and reduced treatment efficacy [24].

2.2. Anxiety and Its Intersection with Neurodegeneration

Anxiety disorders traditionally represent itself as a neuropsychiatric disorder, but is increasingly recognised for its neurobiological underpinnings and its complex interplay with neurodegeneration [25]. Recently, it has emerged not just as a co-occurring condition but as a potential prodromal indicator of several NDDs. The overlapping of anxiety and NDDs is evident both at the clinical level, where anxiety often forgoes cognitive symptoms, at the molecular level, particularly within the GABAergic system [26]. This section pinpoints anxiety from two perspectives, a neurobiological condition with specific molecular targets, and as a comorbid or early-stage feature of dementia-related disorders such as bvFTD, PSP, CBS, and PPA. Understanding this dual role is essential for developing integrative therapeutic

strategies that address both behavioural and molecular aspects of neurodegeneration [27] [28].

2.2.1. Anxiety and Its Types

Anxiety is the most occasional mental disorder with a feeling of fear that occurs when one's facing a stressful or in a threatening situation [29]. It is quite a normal response when confronted with danger, continuous fear, unrealistic worry about something to happen, and/or even unpleasant feelings of imminent death [30]. They are most prevalent neuropsychiatric conditions, affecting nearly about 301 million people and is to be expected affecting 500 Million people by 2050 worldwide. The life-time prevalence rate of anxiety disorder for adolescents aged between 13-17 is 7.7%, while it is 6.6% in adults aged between 18-64 years [31] [32]. Women are more prevalent to develop anxiety disorders and approximately twice as high as in male. The prevalence of anxiety disorders is as follows: 10.3% for specific phobias, 6.0% for panic disorder, 2.7% - Social Anxiety Disorder (phobia), and 2.2% - Generalized Anxiety Disorder (GAD) [33] [34].

Anxiety disorders are often accompanied by neuromuscular tension, restlessness, fatigue, and concentration deficit, resulting in significantly interfere with daily activities. Moreover, scientific evidence suggests that prolonged anxiety result in the development of more serious and detrimental health consequence, which lowered overall life expectancy of the individual [35]. From a neurobiological view, anxiety is in close proximity to dysfunction in various neurotransmitter systems, significantly GABA, serotonin, and norepinephrine that are modulated by the brain regions such as, amygdala, prefrontal cortex, and hippocampus [36] [37].

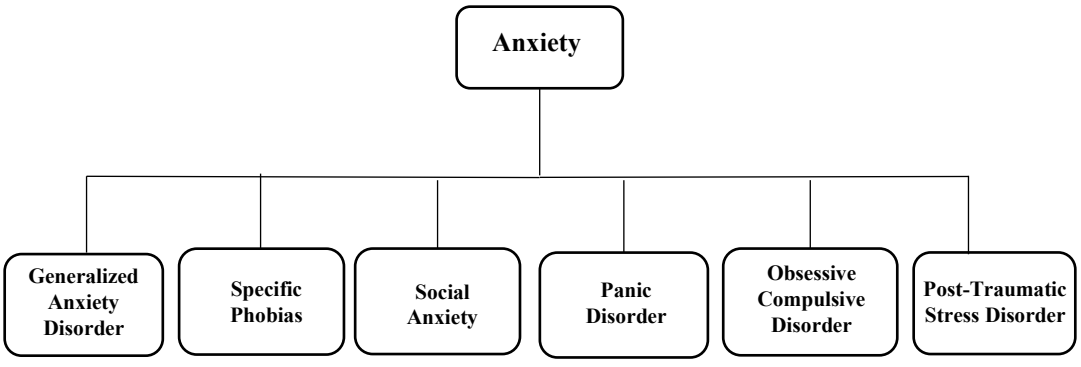


Fig 2.1. Classification of Anxiety disorder

2.2.2. Anxiety as a comorbidity and prodromal symptom in dementia

Anxiety is acknowledged as both the comorbid, a co-existing neuropsychiatric condition and prodromal (before full onset) symptom in several subjects of dementia. In conditions such as, bvFTD, PSP, PPA, and CBS, anxiety is often observed as a common predictive factor of the acceleration of the disease [38]. In the case of bvFTD, it is observed that anxiety often emerges during early transitional stages reflecting degeneration in prefrontal-limbic circuits that mediate emotional control where behavioural disinhibition and emotional blunting are the main hallmark of the disease [39] [40]. Consequently, in the case of PSP and CBS, often mistaken as a PD, expresses frequent anxiety-related symptoms connected to subcortical and brainstem atrophy. Meanwhile, in PPA, anxiety may result from social withdrawal and the impairment of communicating with other individual, magnifying cognitive decline and emotional distress [41]. Researchers suggest that the presence of anxiety usually correlates with rapid cognitive deterioration, poor quality of life, and higher caregiver burden. Addressing the symptoms of anxiety, it may provide for early interventions and modify the progression of dementia [42].

2.2.3. Impact of anxiety in disease progression and quality of life

The presence of anxiety in people with NDDs notably worsen the disease adversely affecting to the progression of the disease, response to the treatment, and the quality of life. It leads to significantly reduced cognitive performance, increment of the caregiver burden, also the heightened disabilities and more [43] [44]. Researchers highlight how these symptoms like anxiety are manifested earlier in prodromal stages of NDDs and can disrupt multiple domains of functioning in CNS [45]. The distress caused by anxiety often paves to misdiagnosis, not performing accurate identification of the condition, limiting the effectiveness of cognition abilities. Understanding the interdependence of the neurodegeneration and anxiety imparts a new way for strategic outcomes that would combine GABAergic modulations, cognitive therapy, and pharmacological management in the long-term effects [46].

2.3. Bridging Dementia Subtypes with Neurodegenerative Mechanisms

To understand dementia as a clinical syndrome caused by various NDD mechanisms is important for the treatment, diagnosis and the biomarker development. Dementia is not a single disease but it is a collection or say, cluster of symptoms arising from several underlying pathologies. Researchers, nowadays have increasedly focused on bridging the gaps between dementia subtypes and their cellular and molecular underpinnings.

2.3.1. Dementia and Its Subtypes

Dementia, also called a complex neurological disorder is characterized by a decline in cognitive functions such as memory, reasoning, and communication [47]. It primarily affects the parts of the brain associated with memory and thinking, including the hippocampal region and cerebral cortex [48]. According to WHO, currently over 50 million individuals live with dementia, with numbers expected to near double by 2050 due to the ageing populations [4]. Early identification and accurate diagnosis of the cause of dementia is important for a number of reasons, including personalized therapeutic interventions and to improve patient outcomes [49].

Among the diverse dementia subtypes, frontotemporal dementia (FTD) comprises of a significant clinical category characterized by progressive changes, like decrease in size in frontal and temporal lobes [50]. A prominent variant of neurodegenerative condition, bvFTD is marked by rapidly progressing behavioural, emotional and personality changes often underdiagnosed due to the overlapping symptoms of the psychiatric conditions [51]. PSP, is another subtype of neurodegenerative disease, commonly observed with postural inability, or motor dysfunction say, movement disorder, linked to tau protein aggregation [52]. CBS, presents asymmetrically with motor symptoms and cognitive impairments, complicating its differentiation from other neurological conditions [53]. PPA affecting the language capabilities, often referring as difficulty in 'word finding', highlights the heterogeneity within dementia subtypes necessitating different diagnostic approaches [54].

2.3.2. Dementia Subtypes and Its integration with Neurodegeneration

Each dementia subtype reflects a specific neurodegenerative process defined by unique proteinopathies, brain region involvement, and molecular pathways [55]. In proteinopathies, taupathies are central to all the four dementia subtypes and AD, though tau isoform composition and topography differentiates across the subtypes [56]. The inclusion of TDP-43 is quite common in bvFTD, and some PPA variants. Meanwhile, α-synuclein aggregates are typically observed in Lewy body dementia but may show overlapping symptoms in other syndromes as well [57].

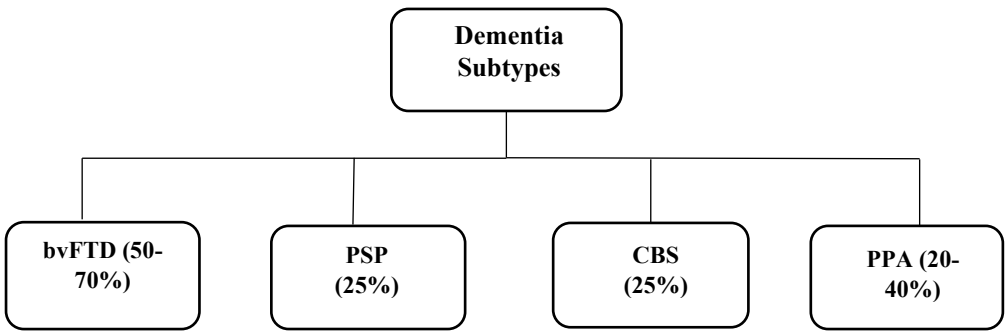


Fig 2.2. Four different subtypes of dementia

Transcriptomic analyses have revealed different molecular patterns across dementia subtypes [56]. For instance, bvFTD is enriched for immune response genes, while AD shows prominent synaptic and metabolic dysfunction. Misdiagnosis is quite common due top shared behavioural and cognitive symptoms [58]. For example, CBS can be misconnected with PD, bvFTD may be misclassified as later-stage symptoms for AD [59]. Recognizing these pathways improves the diagnosis precision and identifying biomarkers [60]. These insights are critical when studying comorbidities like anxiety, which may stem from shared or divergent neurobiological mechanisms depending on the subtype [61].

2.4. Risk Factors and Symptoms of NDDs

A complex spectrum of genetic, environmental, and lifestyle-related factors contributes to the influence of the onset of the disease [62]. These factors not only affect symptoms to the severity, also affects the progression rate and their therapeutic outcomes [63]. Among which, a wide range of genetic and environmental factors, such as, age, sex, genetic susceptibility, lifestyle are the most prominent unmodified risk factors acting as a central role in NDDs [64]. In FTD, mutations in genes, like MAPT (microtubule-associated tau protein), GRN (progranulin), and C9orf72 show a strong implication because of the genetic susceptibilities [65]. As these mutations are well-explained, C9orf72 have not only been linked to FTD, but also making an effect in ALS, stating the genetic overlap between NDDs. PSP and CBS are sporadic, associated with MAPT haplotypes.

Environmental and lifestyle factors, including the exposure to inflammation and chronic psychological stress, heavy metals, pesticides, head trauma, poor cardiovascular health, low cognition, physical inactivity are linked to the accelerated neurodegeneration and synaptic loss. Accordingly, anxiety and chronic psychologic stress may act as both neurodegeneration and independent risk factor promoting oxidative damage or by the disruption of the BBB [65]. Moreover, environmental factors, such as, age and sex is the most significant non modified risk factors, associated with reduced synaptic plasticity, genomic instability and mitochondrial dysfunction that usually predispose neurons to degeneration [66] [67].

NDDs presents a wide range of cognitive, behavioural, motor, and language deficits depending on the regions of the brain affected. For instance, patients with bvFTD, often exhibits impaired personality, difficulty with decision-making, emotional blunting, in despite of the preserved memory in early stages. PPA presents with progressive deterioration in language comprehension (language difficulty), including word-finding issues, grammatical errors and relatively preserved memory and cognition. CBS and PSP involve in slow processing, characterised by axial rigidity, postural instability, and supranuclear gaze palsy, asymmetric limb rigidity, dystonia, apraxia, and cortical sensory loss [68].

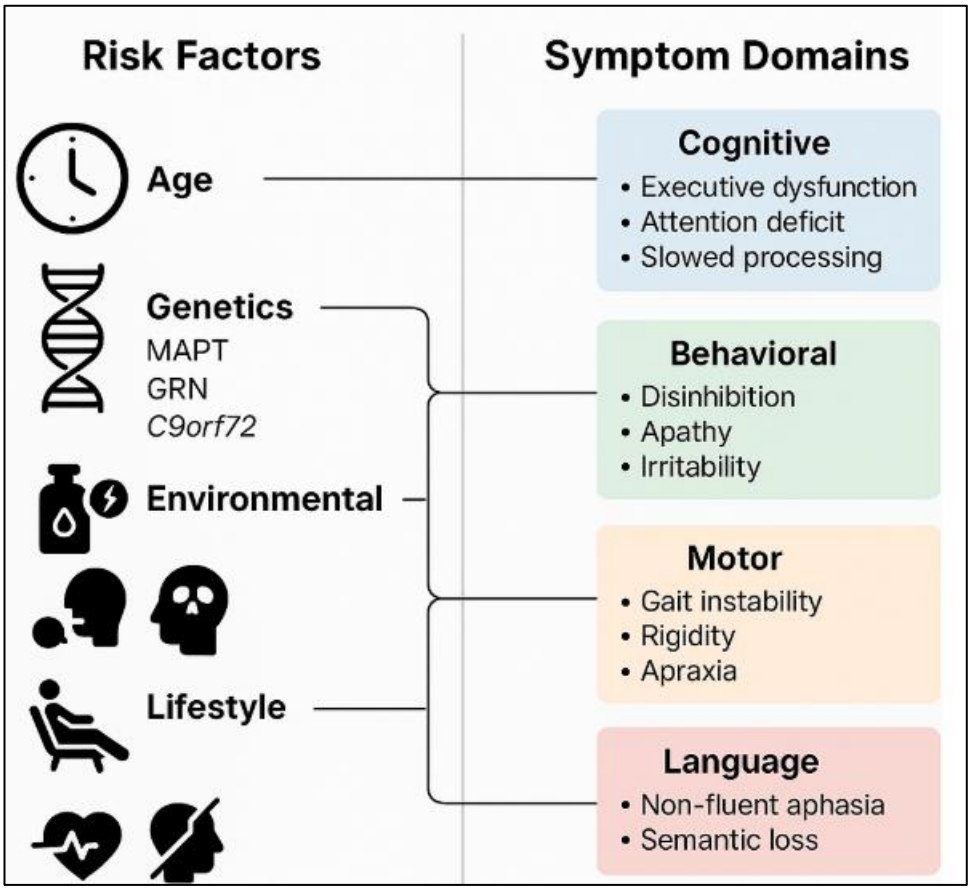


Fig 2.3. Risk factors including, age, genetic mutations, environmental exposures, and lifestyle-related elements.

2.5. GABAergic Signalling and the Role of GABRA2

Gamma-aminobutyric acid (GABA) is the principal inhibitory neurotransmitter in the adult mammalian CNS, acting primarily through GABA-A receptors. In the context of NDDs, alterations in GABAergic signalling contribute to the disruption of the excitatory-inhibitory balance in neural circuits [69]. In AD and FTD, a reduction in GABAergic interneurons and alterations in GABA-A receptor subunit composition have been observed, potentially contributing to network hyperexcitability. This hyperexcitability may exacerbate Aβ production and tau phosphorylation, further driving disease progression [70]. In HD, the preferential loss of GABAergic medium spiny neurons in the striatum leads to a profound disruption of basal ganglia circuitry. This loss of inhibitory control contributes to the characteristic motor symptoms [71]. Additionally, alterations in GABA-A receptor subunit expression have been observed in various brain regions in NDD models, suggesting a more widespread dysregulation of inhibitory signalling [72] [73]. PD also involves alterations in GABAergic transmission, particularly within the basal ganglia circuitry [74]. The loss of

dopaminergic input to the striatum leads to changes in the activity of GABAergic projection neurons, contributing to the motor symptoms of PD [75].

2.5.1. Structure and Physiological Function of the GABA-A Receptor

The GABA-A receptor is essential for maintaining neuronal excitability and preventing excessive neural activity that could lead to disorders such as epilepsy, anxiety, and insomnia [76]. These receptors are pentameric ligand-gated Cl^- channels (LGICs) composed of various combinations of α, β, γ, δ, ε, θ, and π subunits, with the most common subtype being α1: β2: γ2 [77]. Belonging to the Cys-loop receptor family, GABA-A receptors are structurally and functionally homologous to other LGICs such as nicotinic acetylcholine receptors (nAChRs), glycine receptors, and serotonin type-3 ($5\text{-}HT_3$) receptors [78]. This group of receptors is unified by a signature extracellular disulfide-bonded loop formed by a conserved pair of cysteine residues, which plays a critical role in ligand recognition and gating mechanisms. Binding of GABA to these receptors allows chloride influx into the neuron, hyperpolarizing the membrane and thereby reducing neuronal excitability [79].

The subunit composition of GABA-A receptors further underlines their complexity and versatility. Subunits are diversified by a set of genes, and their differential expression contributes to the functional heterogeneity observed across different brain regions and developmental stages. This heterogeneity is essential for tuning the pharmacological and biophysical properties of the receptor in response to various physiological conditions. These receptors are distributed widely across brain regions, including the hippocampus, amygdala, and prefrontal cortex central to emotion regulation, memory, and executive functioning [80] [81]. Specific subunits confer distinct pharmacological properties; for instance, α1 is associated with sedative effects, α2 and α3 with anxiolytic effects, and α5 with cognitive modulation [82].

2.5.2. Role of GABRA2

GABRA2 encodes alpha-2 (α2) subunit of the GABA-A receptor, which functions as a ligand-gated ion channel is a promising therapeutic target for the central nervous system that inhibits neurotransmitters. Diazepam, a widely prescribed benzodiazepine, works by modulating the GABAergic system especially GABRA2 where the therapeutic actions mediate through its binding to the binding site of the receptor. The GABRA2 is underexplored largely compared to other subunits. It plays a key role in cognition and anxiety, and its selective targeting could minimize sedative effect [83]. Studies have shown that GABRA2 mediate the anxiolytic effects of Benzodiazepines without causing significant sedation, a distinction from α1-containing receptors. This subunit is also involved in the modulation of working memory and decision-making, as observed in prefrontal cortical circuits. Genetic variation within the GABRA2 gene, particularly the single nucleotide polymorphism (SNP) rs27985 has been linked to a

height-end susceptibility to anxiety related traits, which is an elevated risk of developing alcohol dependence, also an amplified physiological response to stress [84]. These associations not only underscores the critical contribution of GABRA2 polymorphisms to individual differences in neuropsychiatric vulnerability but emotional regulation.

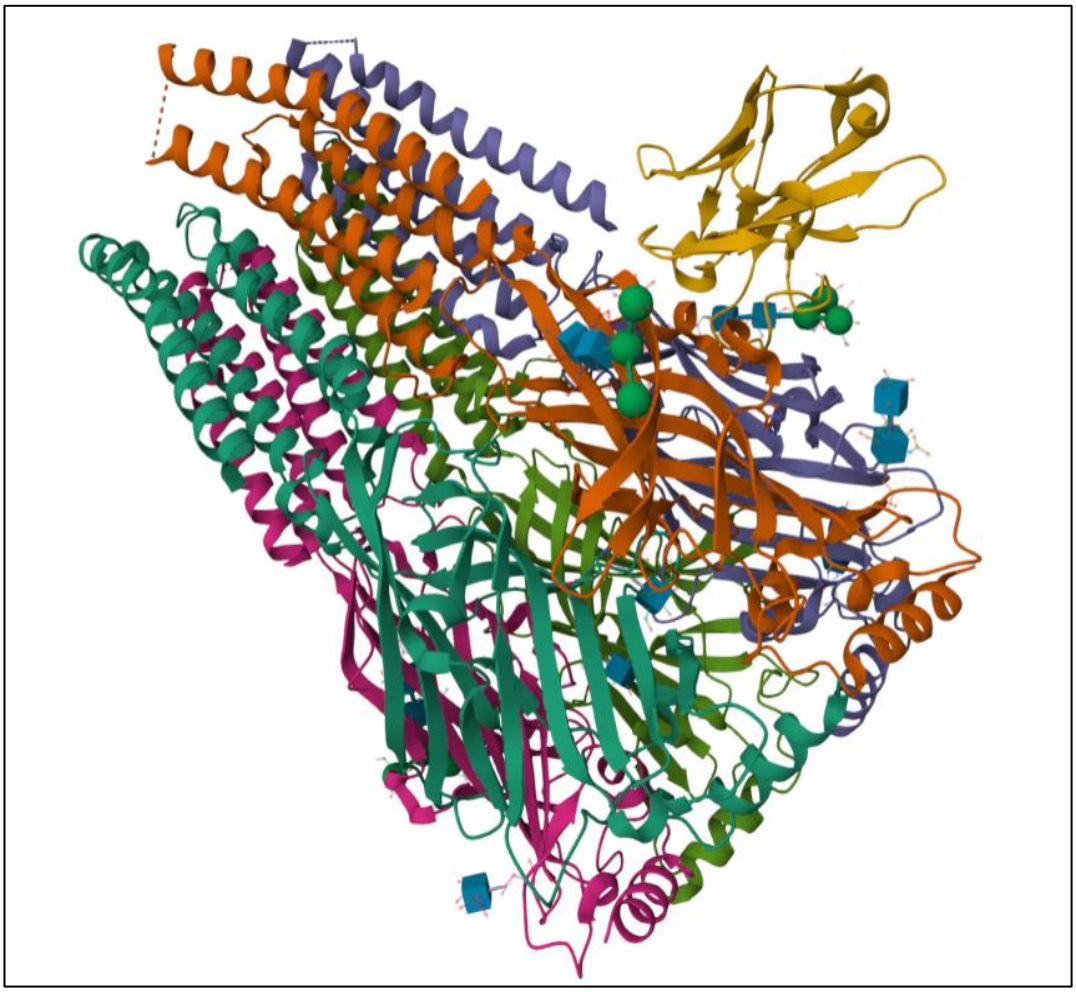


Fig 2.4. Displaying the GABA domains of GABA-A Receptor, representing spatial arrangement of α-helices, β-sheets, and loop regions, emphasizing the conformational folding critical to its function.

The evidence of GABRA2 Dysfunction in NDDs and Anxiety Disorders emerge as linked GABAergic dysfunction, including alterations in GABRA2 expression, to the pathophysiology of several NDDs and anxiety-related conditions. Postmortem studies have revealed altered expression of GABA-A receptor subunits in the hippocampus and temporal cortex, with decreased α-2 subunit levels correlating with cognitive decline and anxiety symptoms [85]. Reductions in GABRA2 mRNA and receptor

density have been observed in the frontal cortex of FTD patients, potentially underlying behavioural dysregulation and apathy. In generalized anxiety disorder (GAD) and panic disorder, neuroimaging studies show reduced GABA concentration in key brain areas, with altered GABRA2 expression linked to hyperactivity of the amygdala and prefrontal regions. Animal models of PD and HD also demonstrate compensatory changes in GABRA2 expression, likely due to loss of dopaminergic and glutamatergic input [86].

2.5.3. Therapeutic Potential of GABRA2

GABRA2 expense a dual role in anxiolytics and cognitive function, representing a promising target for selective GABAergic therapies. Unlike classical BDZs, which uniformly modulate all GABA-A receptor subtypes, α2-selective positive allosteric modulators (PAMs) represent a novel approach to treat anxiety and cognitive symptoms with fewer sedative and addictive liabilities. Compounds such as TPA023 and L-838417 have shown efficacy in preclinical models of anxiety by selectively enhancing α2 and α3 receptor activity. Genetic screening for GABRA2 variants may guide the development of personalized therapeutic interventions, especially for individuals with neuropsychiatric symptoms in NDDs. GABRA-2 targeting agents modulates the efficacy of current treatments when used in conjunction with anti-tau therapy [87] [88].

One intriguing therapeutic approach for reducing the detrimental aftereffects of NDDs is the combinatorial therapies of GABRA2, which enhances the efficiency of current treatments controlling neurotoxic proteins. As predicted, BDZs, has minimal significance in targeting neurological diseases due to its weak BBB permeability. It has also been demonstrated that GABRA2 can be pharmacologically inhibited by BDZs, like Diazepam and others, which can hinder the ability of different tumour types to proliferate, differentiate, and invade. GABRA2 is now a desirable therapeutic target for neurological illnesses as a result of these findings. However, a critical unmet need remains in the development of GABRA2 inhibitors with more BBB permeability than Diazepam. With the ultimate goal of determining their efficacy as medications to relieve the corresponding pathogenic cascades and decelerate the advancement of the disease, the synthesis of these ligands may make it possible to investigate GABRA2 inhibition as a novel therapeutic approach for neurological disorders [89] [90] .

2.6. Receptor-Based Drug Discovery in NDDs

With the use of computer-aided drug design (CADD) methodologies, the drug development process has become efficacious and economical. Through the prudent guidance of experimental endeavours toward viable molecular candidates. CADD approach has made it possible to lessen the cost and temporal constraints that come with traditional drug development pipelines. Receptor modelling enables the

prediction of 3-dimensional conformations with experimental structures which are unavailable. This allows experimenters to model different subunits, that is crucial for the anxiolytic designing that create a sedation. Notably, in the field of CADD, virtual screening (VS) and molecular docking techniques have become indispensable auxiliary methods to the labour- and resource-intensive high-throughput screening (HTS) experimental procedure [91]. By prioritizing the best compounds for further experimental validation, these computational methods provide a supplementary approach that streamlines the entire drug discovery process. Highly specific subsets have been successfully identified using computational screening of large compound libraries, depending on either complementarity to target structures (structure-based) or similarity to existing inhibitors (ligand-based). After that, the activity of these subgroups can be experimentally verified [92].

2.6.1. Homology modelling of GABRA2: Rationale and methods

The GABRA2 is unexplored largely due to unavailability of experimentally validated 3-D structure. It is aim to predict the 3D structure of the GABRA2 protein using homology modelling and further investigate the binding interactions between various compounds and GABRA2 receptor using Molecular docking to identify a lead compound that will potentially inhibit GABRA2 with a great efficiency. The GABRA2 is associated in regulating the major inhibition of neurotransmitter impulses. Although, according to some researchers Serotonin transporter gene (SLC6A4) is also associated with major inhibitory neuro-transmission in adult brain, leading to anxiety [93]. The genetic variations influence neuro transmission regulation, alterations in neural circuits, stress mediated pathways and many. The rationale of Homology modelling is to understand the ligand-binding pocket geometry and gate-channel mechanisms for the recent research in computational drug discovery that has demonstrated significant advancements in VS and MD methodologies, allowing for more precise identification of potential drug candidates. The methods typically involve the selection of the template, followed by the sequence alignment, then to generate models and evaluate energy minimization, also to validate the results using Ramachandran plots. Such pioneering efforts emphasize the potential of computational tools in drug repurposing and the discovery of safer therapeutic alternatives [94] [95].

2.6.2. Molecular Docking and ADME profiling

MD, a popular computational technique in SBDD has been extensively employed in drug discovery. By identifying possible binding modes and evaluating binding affinity, MD aims to evaluate and determine molecular recognition at both a geometric and thermodynamic level. Initially, the interactions between target molecules were the main focus of MD. However, in the past 10 years, nucleic acid-ligand docking, protein-protein docking, and nucleic-acid protein-ligand docking have all received more attention. Two interconnected steps make up docking: first, analysing various ligand

shapes inside the protein's active region; second, assessing these forms using a scoring system [96] .

A scoring formula and a search technique make up docking protocols. However, because the search space is so large, a comprehensive search is computationally infeasible. To minimize the problem's dimensionality and navigate such large search areas with efficiency, limitations, limits, and approximations, are utilized. A scoring function is required for a broad range of binding modes to result from protein-ligand interactions. It should be able to discriminate between all other modes that the search algorithm has examined and experimental binding modes. Currently, available docking methods use two types of scoring functions: A two-step grading function technique that ranks the resulting structures using a stricter scoring function after a reduced function directs the search approach, and Complete scoring functions score a protein-ligand conformation [97] [98].

Absorption, Distribution, Metabolism, Excretion, and Toxicity studies, or ADMET studies for short, assess a drug's pharmacokinetics. This is a critical step in drug development since it entails forecasting the behavior and effects of the medication in the body, including the amount absorbed orally and in the GI system. Neurotoxicity and Nephrotoxicity may result from poor absorption, which can also have a detrimental effect on distribution and metabolism. A novel medication must bind to its therapeutic target efficiently, but it also has to be able to get to the target site at high enough concentrations to safely provide the intended physiological impact [99].

Because ADMET features are taken now into account early in the drug development process, the no. of compounds that fail in clinical trials owing to inadequate ADMET profiles has dramatically decreased. Simply put, ADMET research enables us to comprehend the internal processing of drug molecules in living things. In light of this, ADMET is essential to CADD. Even those who are not familiar with CADD may submit data and analyze findings with ease using the user-friendly interface of the free SwissADME online application. When it comes to sophisticated techniques like iLOGP and the BOILED-Egg model, SwissADME offers a distinct advantage over other free web-based ADME and pharmacokinetics programs like pk-CSM and admetSAR [100] [101].

2.6.3. BBB Permeability and Drug-Repurposing

The BBB is a selectively permeable membrane, which is achieved by the combined actions of astrocytes, pericytes, and endothelial cells. Maintaining brain homeostasis, the barrier shields the brain from diseases and poisons by preventing them from entering the brain's circulation. The barrier is formed mostly by endothelial cells, and its activity is regulated by astrocytes and pericytes via several signaling pathways. A restricted set of solutes can pass through the BBB without the aid of facilitators. Only gases, like CO2 and O2, and tiny, lipid-soluble compounds, like Ethanol and antidepressants, with a MW of less than 400 Da or fewer than eight H-bonds, can

passively permeate over the blood-brain barrier. A crucial measure for assessing the BBB integrity is the barrier's permeability, which shows the degree of paracellular and transcellular movement. Administering medications to the CNS efficiently is still an enormous challenge in the therapy of NDs, even with great advancements in our knowledge of the cellular and molecular mechanisms that govern underlying illnesses and their medicines [102].

As a built-in defence mechanism, the BBB constitutes one of the CNS most important barriers. To protect the CNS from neurotoxic substances and to provide necessary nutrients and oxygen, the BBB must operate properly. The BBB has a variety of cell surface sensors and carriers that allow drugs to flow across and fulfil the high energy needs of the brain. Furthermore, lipophilic compounds have an easy time diffusing into the parenchyma of the brain. Therapeutics that can cross the blood-brain barrier can be created using these physiological traits [102] .

Drug repurposing, which usually includes phrases like 'drug repositioning', 're-profiling', 'therapeutic converting', 're-direction', 're-tasking', 'rescue', 'recycling', is the process of finding novel uses for currently approved pharmaceuticals. Finding novel pharmacological uses for medications that are FDA-approved, marketed, experimental, failing, or already in the process of discovery is part of this process. This strategy gives medications that have been authorized, halted, left on hold, and in the experimental stage a second chance at treating various illnesses. The term 'in silico drug repurposing' is frequently used to describe the computer method. Drug repositioning has been increasingly popular in recent years; now, one-third of newly approved pharmaceuticals are repurposed medications. These repurposed drugs currently make up around a quarter of the Pharma industry's annual income [103] [104].

The eminent approaches to drug-repurposing are: on-target and off-target strategies. A drug's known effects are used for a new purpose in on-target repurposing, resulting in diverse therapeutic outcomes while targeting the same biological target. Conversely, off-target repurposing involves using medications or drug candidates on novel targets for alternative therapeutic purposes, thereby introducing new indications and objectives. Activity-based repositioning is another name for the empirical strategy, which employs empirical testing to determine whether current medications have any novel applications. This approach involves conducting experiments on illnesses to analyze proteins, without requiring prior knowledge of the target proteins' structure. Among the empirical-based repositioning methods accessible are target testing, model organisms, cell assessments, and trials in patients. Nevertheless, in silico repositioning makes use of Computational biology and Bioinformatics techniques to virtually examine sizable public databases including medical and chemical information. The chemical interactions between therapeutic compounds and protein targets are examined in this method to identify putative bioactive substances. It's critical to expand our knowledge using a mix of computational and experimental techniques to

increase medication repositioning success rates. Repositioning medications to be more effective will be made possible by combining these strategies [105] [106].

2.7. Machine Learning in NDD Diagnosis and Biomarker Discovery

Advancements in high-throughput transcriptomic technologies have significantly revolutionised the advent of the research in NDDs. These advance approaches modulate explosive molecular data, in understanding the pathology of the disease. When ML is combined with the high-throughput analyses, it enhances the discovery of potential biomarkers and also helps to improve the classification of disease subtypes. Leveraging deeper with ML algorithms, meaningful patterns are identified from complex datasets, improving into early diagnosis and personalized treatment strategies.

2.7.1. High-throughput Transcriptomics in Dementia

High-throughput transcriptomic, especially RNA sequencing (RNA-seq), enables gene expression profiling in regions of the brain affected by the related disorder. Basically, it is a powerful technique which enables precise quantification of gene expression level in NDDs. Unlike traditional microarray techniques, RNA-seq provides a broad and dynamic range which can detect novel isoforms, transcripts, and low-abundance RNAs. It plays a crucial role in identifying differentially expressed genes (DEGs) in dementia subtypes. With increasing availability of multi-omics data, researchers are expanding RNA-seq applications by integrating it with proteomics, metabolomics, and epigenomics to uncover systemic disease mechanisms. Moreover, applying ML techniques to transcriptomic data enhances the accuracy of classification models, distinguishing between dementia subtypes based on gene expression patterns [107] [108].

2.7.2. ML Models for Subtype Classification

The complexity of NDDs, particularly dementia, necessitates computational models which are capable of distinguishing subtle and complex transcriptomic patterns with high accuracy. In recent studies, including the work by Spooner et al., a comparative analysis of multiple supervised learning algorithms has been explored to classify dementia subtypes using gene expression data. These ML models, Support Vector Classifier (SVC), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Naive Bayes (NB), and Random Forest (RF), have offers varied yet distinct advantages in handling high-dimensional biomedical datasets [109].

SVC is a kernel-based learning algorithm, particularly recognized for high-dimensional biological data due to its ability to manage an optimal space, that separates

classes within gene expression datasets, particularly separates dementia subtypes with limited sample size but larger feature space. SVC enhances its ability in modelling non-linear relationships with smaller datasets, also excels in differentiating the molecular variations by leveraging kernel functions such as, Radial Basis Function (RBF) or Polynomial Kernels (PK) [110].

Traditionally, LR is used to solve the binary classification problems, it remains relevant in the multi-class NDD classification. Despite its simple and traditional methods, LR remains a fundamental tool in disease classification tasks. It offers interpretability in predictive modeling by providing coefficients that directly indicate gene expression impact on dementia subtypes. However, LR assumes linear relationships within the dataset, which may limit its effectiveness in capturing complex gene expression interactions [111].

MLP, a type of feedforward artificial neural network, incorporates multiple layers of interconnected nodes to model complex feature interactions. In the context of dementia research, MLP has shown enhancements in capturing intricate transcriptomic signatures that may be overlooked by linear classifiers. The non-linear activation functions within MLP layers enable the model to detect higher-order relationships between gene sets and disease states, aligning well with the multifactorial and multifaceted nature of neurodegenerative processes. While MLP exhibits strong predictive power, it often requires large-scale training datasets and computational resources to prevent overfitting [112].

NB algorithm, grounded in Bayes' theorem is a probabilistic classifier operating in the assumption of feature independence. Despite this simplifying assumption, NB offers high-dimensional gene expression data efficiently, making it a lightweight yet effective choice for subtype differentiation. In the analysis of dementia-related gene expression data, NB has demonstrated robust baseline performance, particularly in identifying features that independently contribute to classification outcomes. Its probabilistic framework also facilitates transparency in prediction reasoning, which is advantageous in clinical translation [113].

RF, an ensemble learning technique has proven as to be one of the most reliable models in dementia research due to its ensemble approach, constructing multiple decision trees and aggregating their outputs for robust classification. Beyond prediction accuracy, RF ranks gene importance, enabling biomarker identification, reduces variance and enhances generalization alongside disease classification. Its ability to handle complex, nonlinear relationships and mitigate overfitting makes it particularly valuable for high-dimensional datasets. RF has consistently performed well in dementia classification tasks, demonstrating resilience against noise and overfitting, which are common challenges in omics datasets [114].

2.7.3. Feature Selection and Model Interpretability

In NDDs, high-dimensional transcriptomic datasets necessitate identifying a precise set of discriminative features to optimize model performance, minimize computational overhead, and clarify interpretability. This process is pivotal in classification tasks involving thousands of gene expression variables, as noise from non-informative features can obscure biologically meaningful signals. Researchers employ diverse strategies to pinpoint genes linked to dementia subtypes, including filter methods, as in statistical metrics like ANOVA, wrapper methods, like recursive feature elimination, and embedded techniques like feature importance scores from tree-based models. However, standalone applications of these methods often struggle with feature stability-the inconsistency of selected biomarkers across data partitions or experimental setups. To enhance reliability, ensemble-driven feature selection has gained traction. These approaches synthesize results from multiple algorithms or iterative workflows to prioritize robust biomarkers. For example, clustering-guided ensemble frameworks group correlated genes before selection, minimizing redundancy from collinear variables and aligning with the modular nature of biological pathways [115] [116].

In a 2022 study, a consensus-driven framework combining RF, SVM, and LASSO classifiers identified APOE4-independent gene signatures for Alzheimer's disease subtyping, with cross-dataset validation confirming stability. Such methods mitigate algorithmic bias and yield reproducible biomarkers amenable to functional analysis. Model interpretability remains equally critical for clinical adoption. While linear models like LR offer intrinsic transparency, complex architectures like deep neural networks rely on post hoc explanation tools. Techniques such as SHAP values quantify feature contributions globally and locally, while LIME generates instance-specific explanations. For instance, SHAP analysis in a 2023 AD study revealed that elevated GFAP expression and suppressed SYT1 levels drove classifier predictions, implicating astrogliosis and synaptic dysfunction in early pathology [117] [118].

2.7.4. Pathway Enrichment and Functional Relevance of Biomarkers

Pathway enrichment analysis of ML-based gene signatures establishes biological relevance by linking computational predictions to dysregulated processes in NDDs. This approach not only frames biomarkers mechanistically but also uncover potential therapeutic targets through systematic mapping of genes to perturbed cellular functions. In dementia-related disorders, enriched pathways frequently involve neuroinflammatory cascades, such as NF-κB activation in AD, interleukin signaling, and cytokine-cytokine receptor interactions, which underlie chronic inflammation and neuronal damage in different subtypes. For instance, microglial NF-κB pathway genes show elevated expression in early AD, correlating with synaptic pruning and cognitive deficits-a finding validated across multiple transcriptomic cohorts [119] [120] [121].

The ubiquitin-proteasome system (UPS) emerges as another critical pathway, with ML models consistently identifying UPS genes like UBE3A and PSMC4 in AD/PD biomarker panels. Impairments in this system promote pathological aggregation of tau and α-synuclein, as demonstrated by proteomic-pathway cross-validation in a study. The enriched pathways such as axon guidance and neuroactive ligand-receptor interaction suggest impaired neuronal connectivity, particularly relevant to bvFTD and PPA, while protein processing and ubiquitin mediated proteolysis indicate disrupted protein clearance, a hallmark of CBS and PSP. Additionally, pathways linked to oxidative stress (HIF-1 signalling, apoptosis) and metabolic dysfunction (TGF-β signalling, central carbon metabolism) suggest systemic contributions to neurodegeneration. These findings reinforce the hypothesis that dementia subtypes share overlapping yet distinct molecular signatures, providing potential targets for biomarker validation and therapeutic interventions. By bridging ML outputs with pathway biology, researchers can accelerate translational applications, from stratified diagnostics to mechanism-driven drug discovery [122] [123].

2.8. Integration of Computational Drug Design and Biomarker Discovery

The emergence of NDD research is crucial to unify molecular biomarker discovery with receptor-targeted therapeutic development. While transcriptomic studies and structural pharmacology have traditionally operated in isolation, advances in ML and multi-omics integration now enable synergistic frameworks that bridge these domains, accelerating translational outcomes. Gene-level approaches, including transcriptomic and epigenomic profiling, prioritize biomarker identification and disease subtyping through ML-driven analysis of gene expression patterns. These studies often lack direct therapeutic relevance, as identified biomarkers may not correspond to druggable protein targets. Conversely, receptor-level strategies focus on ligand-receptor docking, QSAR modeling, and binding affinity simulations using AutoDock Vina or GROMACS to discover compounds, frequently analyzing upstream regulatory mechanisms influencing drug efficacy.

Modern computational pipelines merge multi-omics and in silico pharmacology to prioritize targets and validate compounds within functional pathways. For instance, He et al. combined ML-based clustering with molecular docking to validate NPAS4 as a dual biomarker and therapeutic target for cognitive impairment. Similarly, Sharma & Bhatia et al. developed a Deep Learning (DL) model connecting dementia-associated gene networks to ligand-receptor binding profiles, enabling concurrent biomarker and drug candidate validation. Integrated approaches are particularly impactful for conditions like AD with comorbid anxiety, where genes such as GABRA2 and CRHR1 influence both synaptic plasticity and mood regulation. Molecular dynamics simulations revealed docosahexaenoic acid (DHA) binds RORα with high affinity, modulating circadian rhythms and anxiety pathways. While integrative models show promise, key challenges persist, like Standardizing multi-omics datasets across

cohorts, Balancing model complexity with clinical transparency, Translating in-silico predictions to in-vivo efficacy [124].

CHAPTER – 3

METHODOLOY

3.1. Computational Resources and Outline

This thesis aims to explore a dual-methodological approach bridging two complementary yet distinctive domains within NDD research: Receptor-based drug discovery and gene expression-based biomarker identification, as Objectives 1 and 2.

Databases Utilized:

Google Scholar (https://scholar.google.com/): It is employed as a comprehensive academic search engine to identify peer-reviewed literature, including journal articles, theses, conference papers, and books gathering full-text information.

PubMed (https://pubmed.ncbi.nlm.nih.gov/): Few medical sources like Medline, scientific publications, and digital books are cited extensively in PubMed. The references contain full-text information and abstracts accessed via PubMed Central as well as links to the author's site.

Protein Data Bank (https://www.rcsb.org/): It is a collection of 3-D structural information for important biological molecules including proteins, genetic material, and RNA, and has the RCSB PDB (RCSB.org) as its US data center. Undertaking research and offering instruction in the domains of biological sciences, wellness, power, and biotechnology are the main goals of the RCSB PDB [125].

UniProtKB (https://www.uniprot.org/help/uniprotkb): The UniProt Knowledgebase serves as a central platform for the collection of detailed, reliable and consistent annotations of protein functions.

ChEMBL (https://www.ebi.ac.uk/chembl/): ChEMBL is a repository on bioactive compounds exhibiting drug-like characteristics. It integrates chemical, bioactivity and genomic data altogether to support the translation of genomic information into effective new drugs [126].

DrugBank (https://go.drugbank.com/): It is an indispensable tool for any biopharmaceutical research because of its comprehensive and trustworthy drug data, which is arranged for easy access or software integration [127].

PubChem (https://pubchem.ncbi.nlm.nih.gov/): It is an open-access chemistry database that is run by the National Institute of Health (NIH) that allows people to submit and share scientific data. Hundreds of informational entries have been regularly

given by PubChem periodically since its founding, solidifying its position as an indispensable tool for scholars and the general public [128].

SwissSimilarity (https://www.swisssimilarity.ch/): This website allows to perform ligand-based virtual screening of several libraries of small molecules using different approaches. More promising lead compounds are being discovered using this technique than with the traditional ones falling into an adequate ADMET characteristics [129].

Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/): Source of publicly available functional genomics data repository supporting MIAME-compliant data submissions [130].

KEGG pathway (https://www.genome.jp/kegg/pathway.html): This database contains curated biological pathways, helping researchers understand molecular interactions within diseases [131].

Tools Utilized:

SWISS-MODEL (https://swissmodel.expasy.org/): SWISS-MODEL is a web-based integrated service dedicated to protein structure homology modelling. It guides the user in building protein homology models at different levels of complexity [132].

Swiss ADME (https://www.swissadme.ch/): During the drug development process, this online web interface helps users to predict ADME variables pharmacokinetic profiles, drug-likeliness, and medicinal chemistry compatibility for one or more small molecules [133].

Geoparse tool (https://pypi.org/project/GEOparse/): A tool used for extracting gene expression datasets from GEO, ensuring efficient molecular data retrieval [134].

NetworkAnalyst 3.0 (https://www.networkanalyst.ca/): An online tool designed for pathway enrichment analysis and functional genomics research [135].

Synthetic Minority Oversampling Technique (SMOTE): A machine learning tool that generates synthetic data points to address class imbalances [136].

Software Utilized:

PyMol: In many scientific domains, such as computational chemistry and structural biology, molecular visualization software has grown to be a highly useful tool. A degree of detail and customization that would be unattainable in a laboratory setting is made possible by them when it comes to the visualization and analysis of the structures of molecules like proteins, nucleic acids, and tiny chemical compounds [137].

Open Babel: Used for format conversion and energy minimization of ligand structures before virtual screening [138].

AutoDock Vina: It is an open-source docking engines, a turnkey computational docking program that is based on a simple scoring function and rapid gradient-optimization conformational search. It was originally designed and implemented by Dr. Oleg Trott in the Molecular Graphics Lab, and it is now being maintained and develop by the Forli Lab at The Scripps Research Institute [139].

LigPlot+: The program automatically generates schematic 2-D representations of protein-ligand complexes from standard Protein Data Bank file input. It was designed to facilitate the innovative biotherapeutics and small molecule medications that are stable, optimized, and have attractive safety profiles [140].

Scikit-learn: A widely used ML library in Python, employed for implementing classification models on gene expression data [141].

StandardScaler: A feature scaling technique from Scikit-learn that ensures uniformity across gene expression values for accurate ML analysis [142].

3.2. Objective 1: Receptor-based Drug Discovery

3.2.1. Data Extraction

From the RCSB Protein Data Bank, the 3-D structural coordinates of the protein were identified. Since, the 3-D structure of the protein coded by the gene GABRA2 was not or readily available, we use the amino acid sequence associated with this gene which was available on UniProtkb [125]. In order to curate a dataset of the ligands having a comparable size and structure as the current anxiolytic drug Diazepam, we performed ligand based virtual screening using the smiles of Diazepam as reference on the SwissSimilarity web interface which calculates similarity scores between small molecules based upon 2D and 3D molecular fingerprints [129]. We selected Combined methods for virtual screening which combines linear path – based molecular fingerprints (FP2) and non – superpositional 3D fingerprints (Electroshape) methods to provide us with a similar molecule as the reference. PubChem and ChEMBL was used to download the screened ligand files in 3-D sdf format [126].

3.2.2. Homology Modeling

Homology modeling is a method to accurately predict 3-dimensional structure from amino acid sequence for protein whose 3-D models aren't yet determined experimentally. As the 3D structure of the protein encoded by the GABRA2 gene was not readily accessible, the corresponding amino acid sequence obtained from

UniProtKB was utilized for subsequent analysis. This sequence was then provided to SWISS-MODEL web tool which then perform sequence alignment to identify several templates for modelling [132]. The best template with high sequence similarity and identity index is selected and the 3D model is generated. The 3-D structure was then downloaded in pdb format for further assessment.

3.2.3. Target Receptor Preparation

The GABRA2 structures were created with the help of Autodock Vina. After loading the target receptor into AutoDock tools, the missing residues were fixed, water molecules were eliminated, polar hydrogen atoms were added, and Kolmann charges were added. Additionally, the PDBQT format was used to store the receptor structure. The receptor's visualisation and editing were also validated using PyMOL [137]. Subsequently, the optimization of the protein receptor's energy was also conducted using the Swiss-PDB Viewer software tool. This procedure entailed the repetitive application of molecular mechanics force fields alongside optimization methods. The primary aim was to attain a thermodynamically stable conformation, thereby augmenting the overall stability of the protein structure.

3.2.4. Ligand Molecules Preparation

To facilitate the repurposing of FDA-approved drugs against Diazepam, a collection of 520 drugs was sourced from PubChem and ChEMBL. The 3-dimensional structures of these ligands were initially generated in SDF (Structure Data File) format via PubChem. Subsequently, using OpenBabel tool, the ligand molecule was then converted from sdf to pdqt format while introducing torsion root and adding Gasteiger charges. Furthermore, configuration files for docking were prepared [138].

3.2.5. Molecular Docking Studies

An essential method for determining the ideal alignment and affinity for attachment of small molecules to proteins that serve as receptors is molecular docking. Molecular docking studies are conducted to predict binding potential of two compounds, be it protein-ligand, protein-protein or protein-nucleic acid. It is a computational study of understanding how two molecular structures bind in silico. After the preparation of ligand and receptor, screening of potential drug candidates, docking process was initialized with the help of AutoDock Tools [139]. Followed by defining the binding site, the grid box is prepared around the binding site of the receptor such that inappropriate docking is surpassed. Actual docking was performed using AutoDock vina 4, which is a command line-based application. Upon execution, Vina returns us with different binding poses each with a corresponding docking score reflecting the predicted binding affinity. When docking was finished, the findings were

methodologically retrieved in.csv format, revealing the binding affinities of every ligand to the receptor protein (GABRA2). The distinct output file for every docked ligand was carefully maintained, offering a thorough account of interactions and H-bond forms between the ligand and the protein. For a future 2D interaction study, every docked ligand's distinct output file was carefully saved. An in-depth description of the associations and formation of H-bonds between the binding component and the receptor was provided in this file. To examine and assess these interactions, LigPlot+ was utilized. Substances exhibiting binding energies below than control, Diazepam were chosen for more in-depth examination [140].

3.2.6. ADME Analysis

ADME is commonly known as Absorption, Distribution, Metabolism, and Excretion. SwissADME web tool was used to profile each drug on the basis of drug's physiochemical properties, water solubility, lipophilicity, potential and the pharmacokinetics studies. The ligand chemical structure is loaded in the SMILES format, representing the molecular structure in linear text form. Next, SwissADME evaluates whether the given compound has the potential to be a drug or not by checking if it follows Lipinski's Rule of Five. Moreover, it also gives insights and alert for any PAINS or structural warnings. The results are snapped in a reported excel sheet, to assess the potential efficacy, drug-likeliness, and safety of the ligand for further drug development. The objective of this screening was basically to find chemicals with both drug-like properties and advantageous ADMET profiles that were BBB permeable. Since it lowers expenses and lessens the chance that novel medications won't work out in clinical studies, assessing ADMET characteristics is essential [133].

3.2.7. Analysis and Visualisation

The top five complexes with the highest binding affinity to bind with the GABRA2 receptor were then visualised using PyMol 3.0.4. and in order to find out about the specific interactions between ligands and the receptor, LigPlot+ was used and map of ligands and their interacting residues was noted and saved.

3.2.8. Prediction of Biological Activity for Substances (PASS Analysis)

The PASS website was employed to forecast the pharmacological characteristics of certain compounds. In order to produce forecasts, this program evaluates compounds according to their structure-activity connections and compares them to known molecules. The Pa:Pi ratio indicates the probability that a molecule possesses specific biological characteristics. Drug development relies heavily on the ability to predict biological action.

3.3. Objective 2: Gene-expression based Biomarker Discovery

3.3.1. Data Acquisition

The gene expression dataset was accessed from the GEO database, which was thoroughly examined to discover studies that included gene expression data from patients suffering from different types of dementia and healthy patients. Using the Geoparse tool, GSE140830 dataset was extracted [134]. This was uniquely identified using GSM IDs, while the columns contained information corresponding to the expression levels of multiple genes across all samples. Moreover, class labels that clarify to the specific conditions or groups were integrated into the expression data from the metadata. This data was further rearranged into a structured format, where rows represented individual samples (GSM IDs), and columns, gene expression levels along with a final column referring another class label. This format allowed efficient preprocessing, statistical analysis, and ML implementation.

3.3.2. Data Preprocessing

The gene expression data was pre-processed and prepared for ML analysis. Such handling of missing values within the dataset, ensured its reliability for subsequent steps. Gene expression values were normalized using quantile normalization to correct variability across the samples. By doing so, we got values that maintained a normal distribution, limiting the effect of outliers. Following normalization, features were scaled using a StandardScaler that standardizes features by removing the mean and scaling to unit variance. This was a crucial step because it allowed scaling the input data to optimize the performance of the ML models since many algorithms can be sensitive to this. Additionally, class imbalance in the dataset was also handled using the Synthetic Minority Oversampling Technique (SMOTE) [136]. This oversampling approach produces modeled samples of the minority classes, ultimately presenting a balanced dataset.

3.3.3. Feature Engineering and Dimensionality Reduction Methods

Given the high-dimensional nature of gene expression datasets, feature engineering was performed to extract biologically meaningful signals. RF feature importance rankings identified the most predictive genes, contributing to refined biomarker selection. Further dimensionality reduction techniques, such as Principal Component Analysis (PCA), were explored to visualize gene clusters while retaining variance. Additionally, t-SNE was investigated as a nonlinear method to differentiate dementia subtypes based on gene expression patterns, enhancing interpretability in the classification framework.

3.3.4. Machine Learning and Model Training

Datasets, after its preprocessing, were separated into two subsets: 80% for training set and remaining 20% for testing set. A range of ML models was implemented, to classify the data, which include, Support Vector Classifier (SVC) with a linear kernel, Naïve Bayes, Logistic Regression, Multilayer Perceptron (MLP) Classifier, and Random Forest (RF) Classifier. Every model proved suitable for sustained prediction performance and high-dimensional gene expression data [143]. The RF classifier was emphasized because of its capacity to manage high-dimensional data efficiently and produce comprehensible feature importance rankings which is a crucial component in the identification of important biomarkers.

3.3.5. Hyperparameter Tuning and Model Optimisation

Hyperparameter tuning was applied to refine model architectures and maximize classification accuracy as ML models requires careful tuning to achieve optimal performance. Grid and random search methods, were explored to systematically optimize key parameters such as regularization strength in SVC, tree depth RF, and learning rate in MLP. Stratified cross-validation was used to ensure generalizability, minimizing overfitting while preserving predictive integrity. Ultimately, hyperparameter selection was guided by performance metrics, ensuring optimal balance between bias and variance.

3.3.6. Model Evaluation and Cross-validation Strategy

We performed cross-validation and tested on the reserved test set for model evaluation. A Stratified K-Fold cross-validation with 5 splits was performed to make sure that our model evaluation was robust. This approach evaluated the model's multiple times on different training data subsets with the average accuracy and standard deviation calculated as a measure to see how consistently each model performed. For each model, standard classification measures such as accuracy, precision, recall, and F1-score were calculated.

Accuracy = $true\ positives + true\ negatives / true\ positives + true\ negatives + false\ positives + false\ negatives$

Precision = $true\ positives / true\ positives + false\ positives$

Recall = $true\ positives / true\ positives + false\ negatives$

$F1\text{-}Score = 2\times Precision\times Recall / Precision + Recall$

These measures provided an in-depth evaluation of each model's ability to classify the gene expression data correctly. This was then followed by a comparative analysis of

the performance of the models, to identify which classifier performed best on the given dataset.

3.3.7. Biomarker Identification

Using the RF classifier, we performed biomarker identification, specifically highlighting the highest-ranking genes that contributed to prediction. The feature importance values were trained, and a bar plot was used to visualize the top 20 most significant features These key features outline potential biomarkers that could be crucial in distinguishing between different classes within the dataset. To delve deeper into class-specific insights, a separate analysis was performed to identify unique biomarkers of each target class. To this end, binary classification was performed for each class using the RF model. This allowed the identification of the top 20 features most relevant to each class. These features were saved as potential biomarkers, and a comprehensive file compiling all class-specific biomarkers was created to offer valuable biological insights.

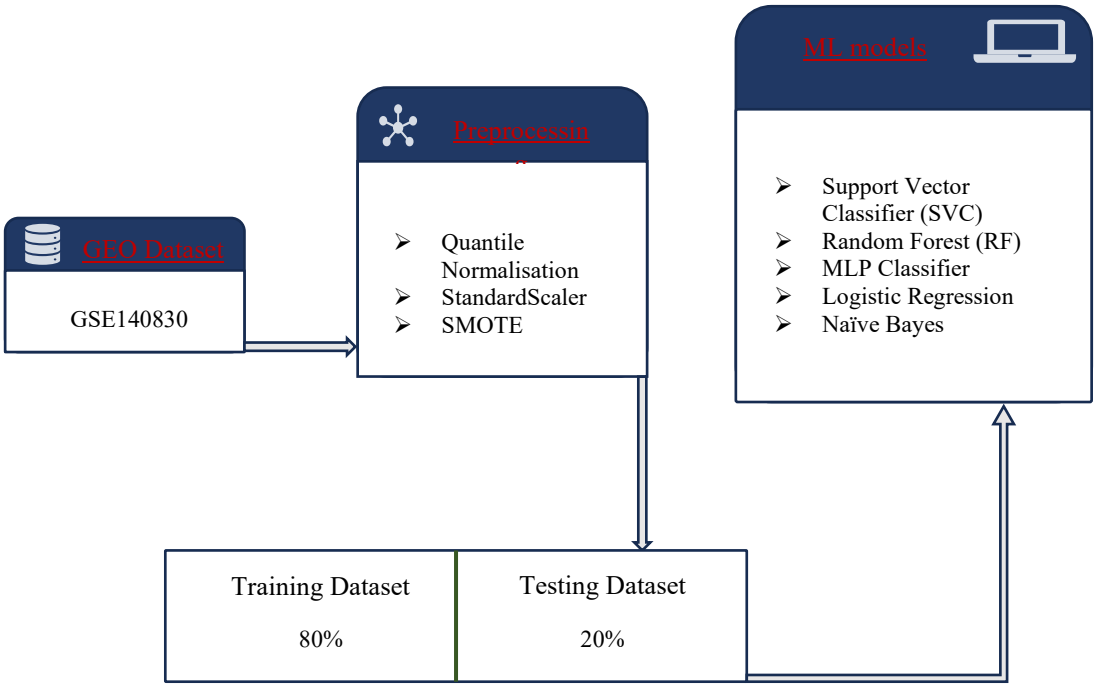


Fig 3.2. Pipeline of the methodology

3.3.8. Pathway Analysis Integration

To validate the biological relevance of the 20 top biomarkers identified through ML-based feature importance analysis, pathway enrichment analysis was performed on their target genes. It was performed using NetworkAnalyst 3.0, a web-based tool for comprehensive functional analysis. The target genes were analysed across databases, including KEGG pathway enrichment analysis to identify significantly enriched pathways. Pathways were ranked based on adjusted p-values (< 0.05), emphasizing biological relevance. Results were visualized using bar chart highlighting the top ranked pathways, ensuring clear communication of the findings.

CHAPTER – 4

RESULTS AND DISCUSSION

(Objective: 1)

4.1. Homology Modeling of GABRA2 gene

The modeling query submit for the gene sequence of GABRA2 returned 48 templates based upon the sequence alignment and we chose the Q5RCC5.1. A template because it showed highest identity (99.78%) to the reference sequence. The resulting protein had 451 amino-acids; the three-dimensional structure was further validated by additional quality parameters such as GMQE score which was 0.84 confirming the accuracy of the model. Also, we validated our model using RMSD and Ramachandran plot to ensure structural accuracy before docking, which confirmed that over 90% of the residues were in the most favoured regions, indicating good stereochemical quality. The structure and validations are shown in the below Fig 4.1.

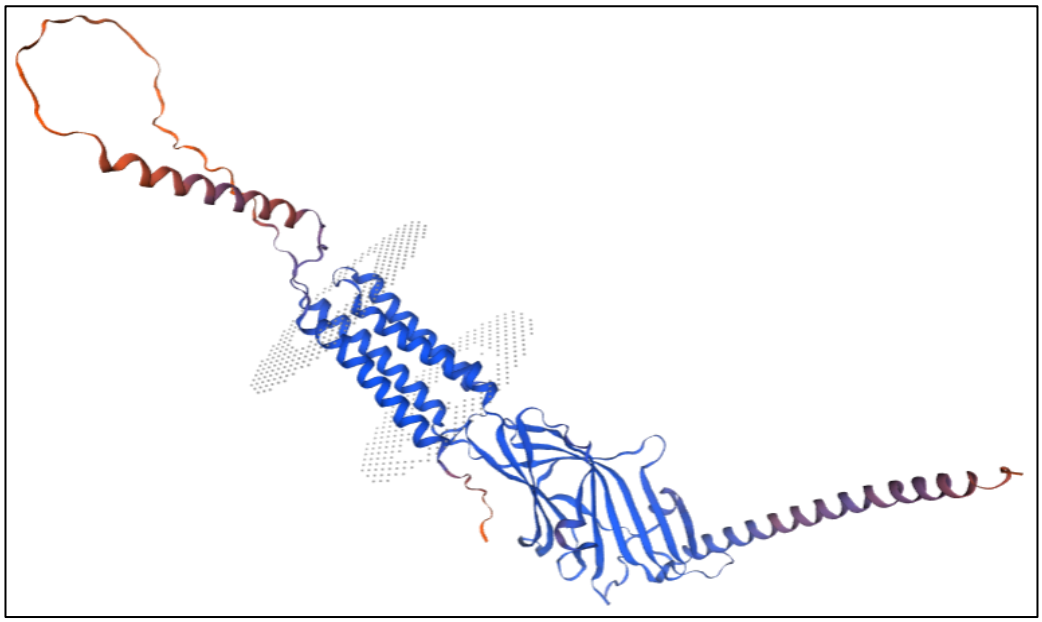


Fig 4.1.1. 3-D model retrieved for GABRA2 by Homology Modeling

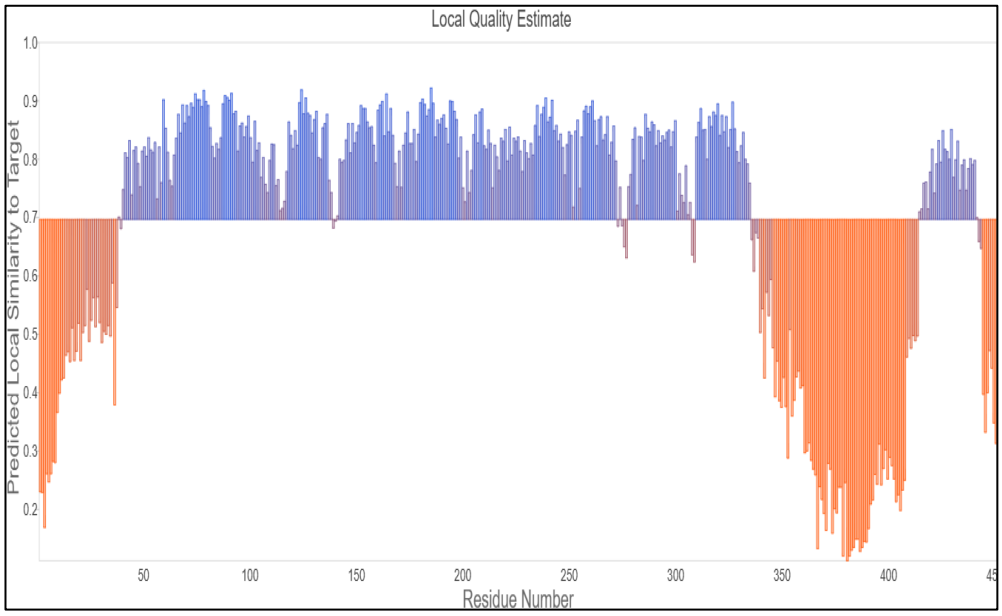


Fig. 4.1.2. Per-residue local QMEAN score plot.

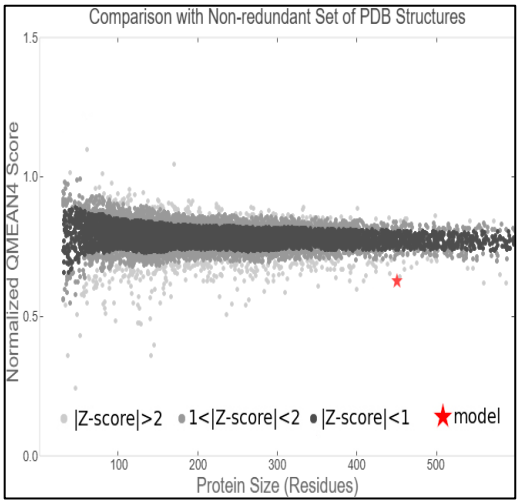


Fig 4.1.3. Global QMEAN Z-score comparison with PDB reference structures.

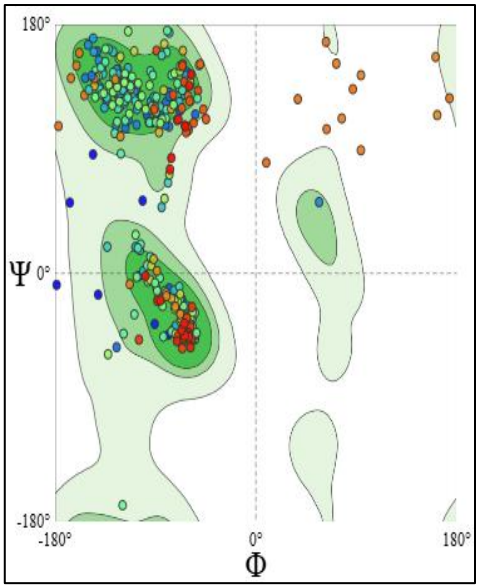


Fig 4.1.4. Ramachandran Plot showing 91.98% with favoured regions

4.2. Ligand-based Virtual Screening

To find possible drugs against predetermined biological targets, a computational strategy called virtual screening is used. It has become a crucial approach, in order to minimize empirical effort and to save time. Using SwissSimilarity, a ligand-based virtual screening approach to reduce the amount of in vitro work required, it returned us with 400 molecules that had structural resemblance with Diazepam (control). The reference medication, Diazepam has a binding energy of -7 kcal/mol, but of the screened compounds, 42 of the 400 FDA-approved medications showed a binding energy of -8 kcal/mol or lower, satisfying the requirement of having an enhanced binding affinity as shown in Table 4.1. The screened compounds were then searched on ligand structural databases such as, DrugBank and Pubchem for retrieving their 3D structure, only a 157 of the screened compounds had a 3D Structure available. These 157 files were then further processed. The Molecular Docked complex of GABRA2 with ligands with top binding 3 binding affinity is shown in Fig 4.2.

Table 4.1. Binding Affinity of Top 15 Ligands in accordance of the Reference Drug

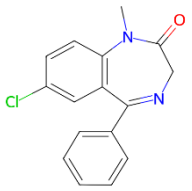
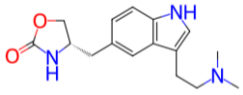
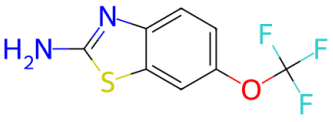
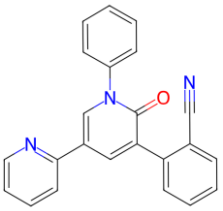
S.No.	Chembl ID	Drugs	Structure	Binding affinity (kcal/mol)
01.	**CHEMBL12**	**Diazepam (Control)**		**-7**
02.	CHEMBL1185	Zolmitriptan		-7.7
03.	CHEMBL744	Riluzole		-7.6
04.	CHEMBL1214124	Perampanel		-7.5

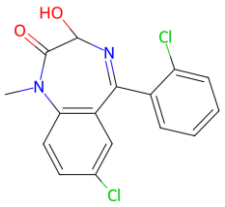
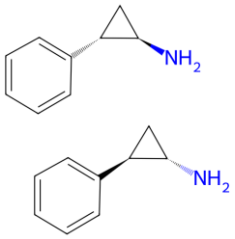
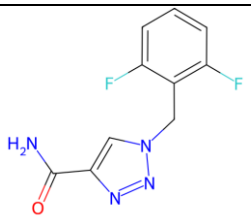
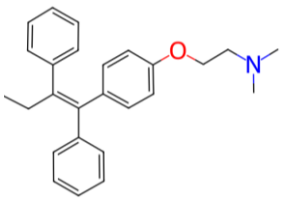
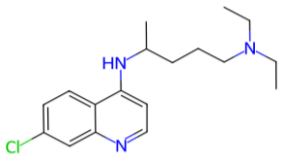
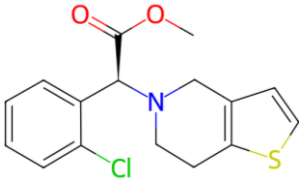
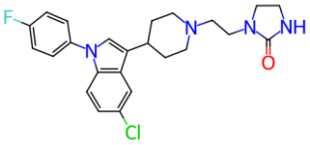
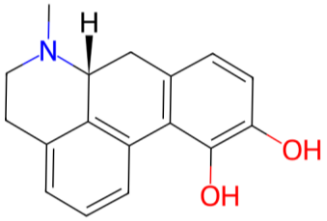
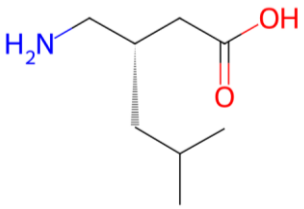
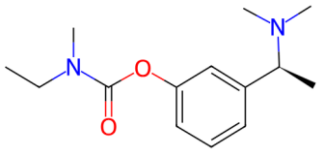
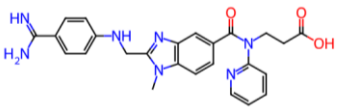
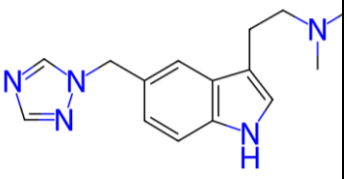
Table 4.1 (Continued)				
05.	CHEMBL22097	Lormetazepam		-7.5
06.	CHEMBL3989843	Tranylcypromine		-7.5
07.	CHEMBL1201754	Rufinamide		-7.5
08.	CHEMBL83	Tamoxifen		-7.4
09.	CHEMBL76	Chloroquine		-7.4
10.	CHEMBL1771	Clopidogrel		-7.3
11.	CHEMBL12713	Sertindole		-7.3
Continued on next page				

Table 4.1 (Continued)				
12.	CHEMBL53	Apomorphine		-7.3
13.	CHEMBL1059	Lyrica (Pregabalin)		-7.3
14.	CHEMBL636	Rivastigmine		-7.2
15.	CHEMBL48361	Dabigatran		-7.2
16.	CHEMBL905	Rizatriptan		-7.2

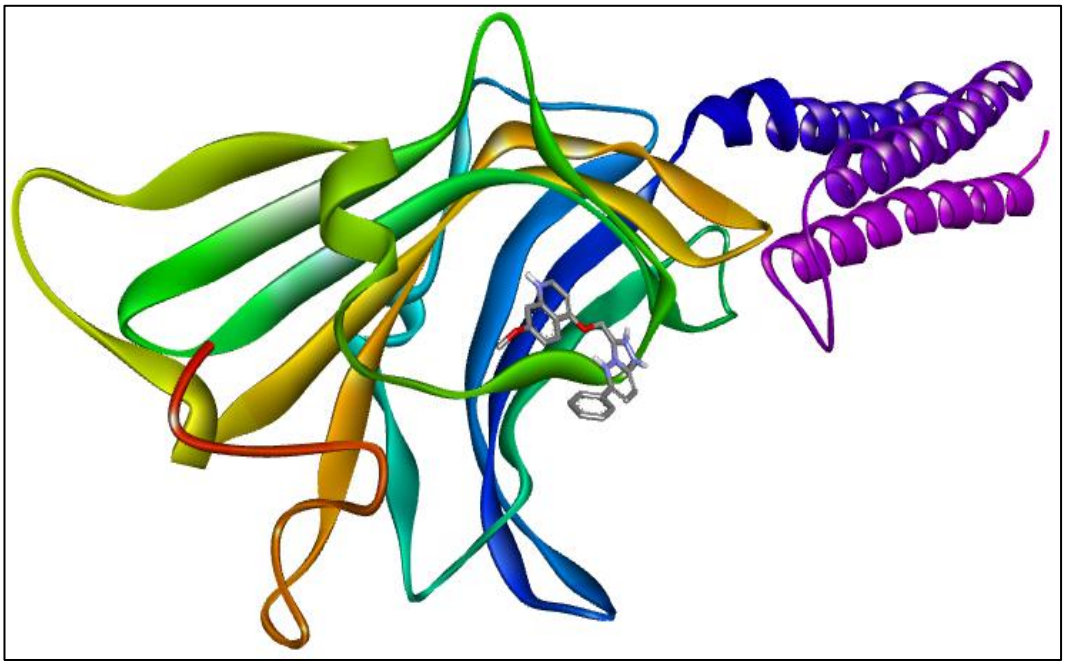


Fig 4.2.1. Molecular Docked Complex of GABRA2-Diazepam (Control)

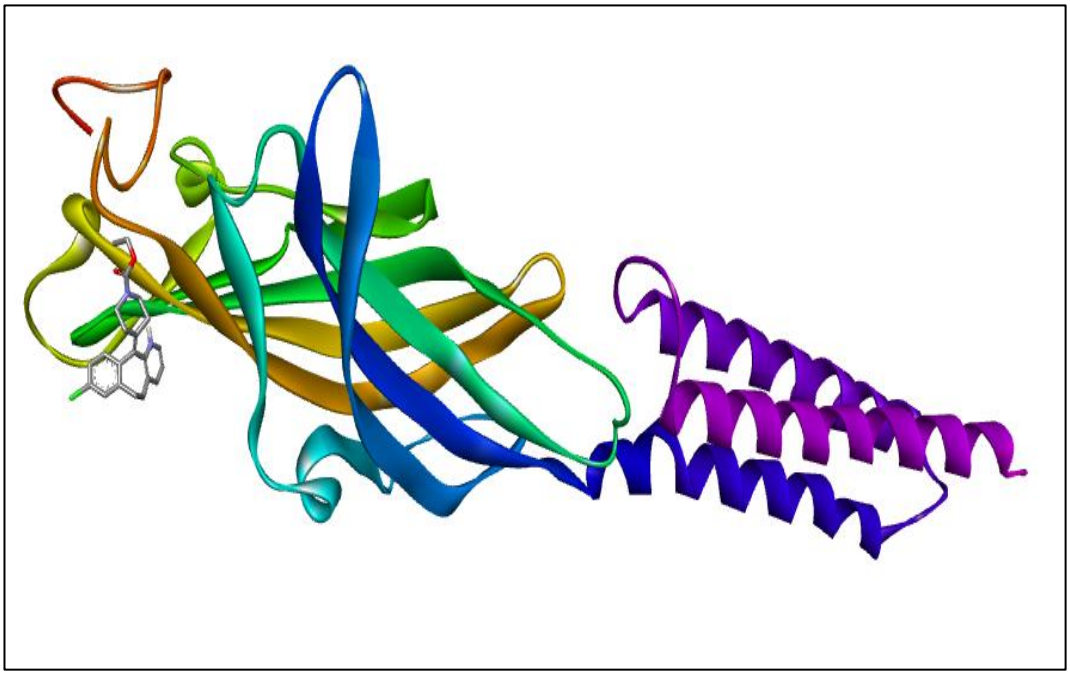


Fig 4.2.2. Molecular Docked Complex of GABRA2-Zolmitriptan

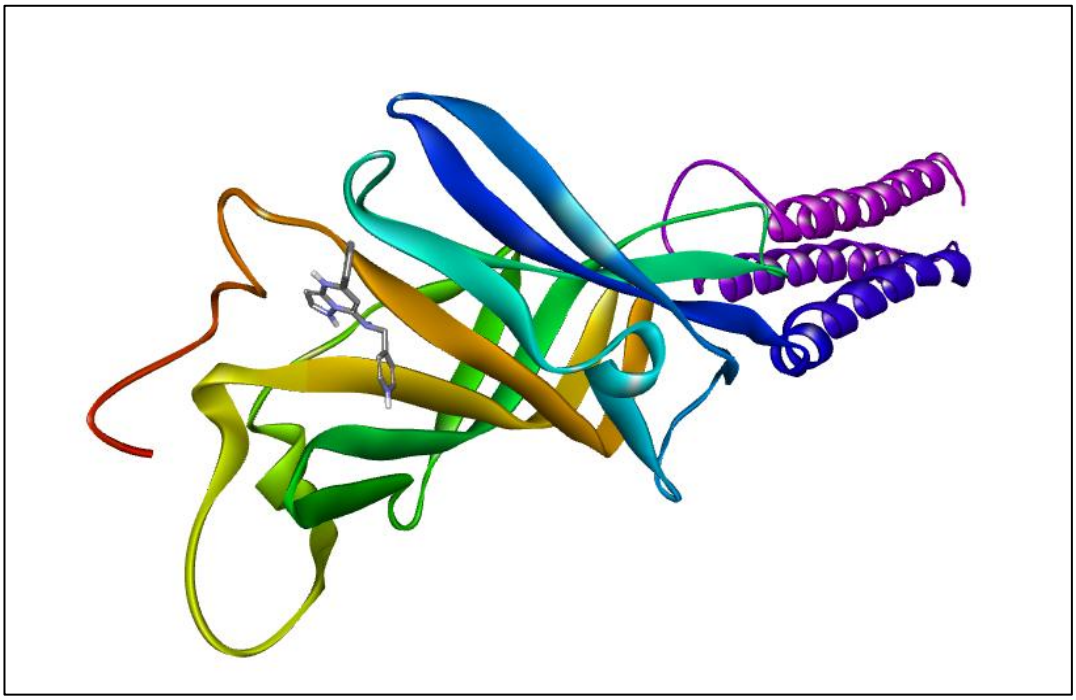


Fig 4.2.3. Molecular Docked Complex of GABRA2-Riluzole

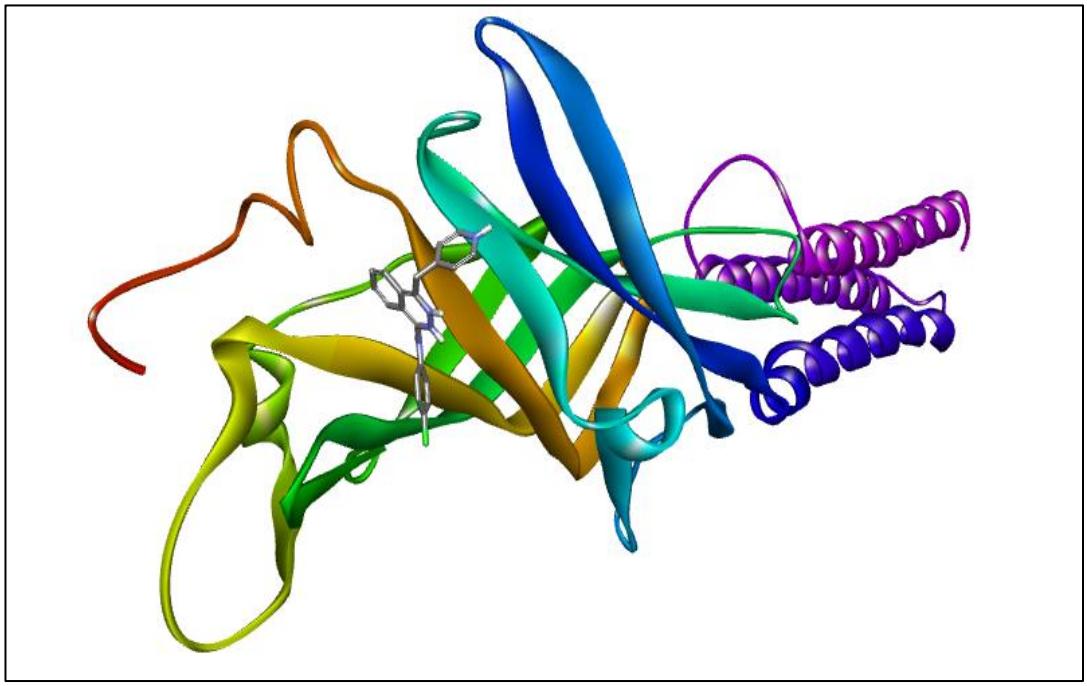


Fig 4.2.4. Molecular Docked Complex of GABRA2-Perampanel

4.3. Interaction Analysis

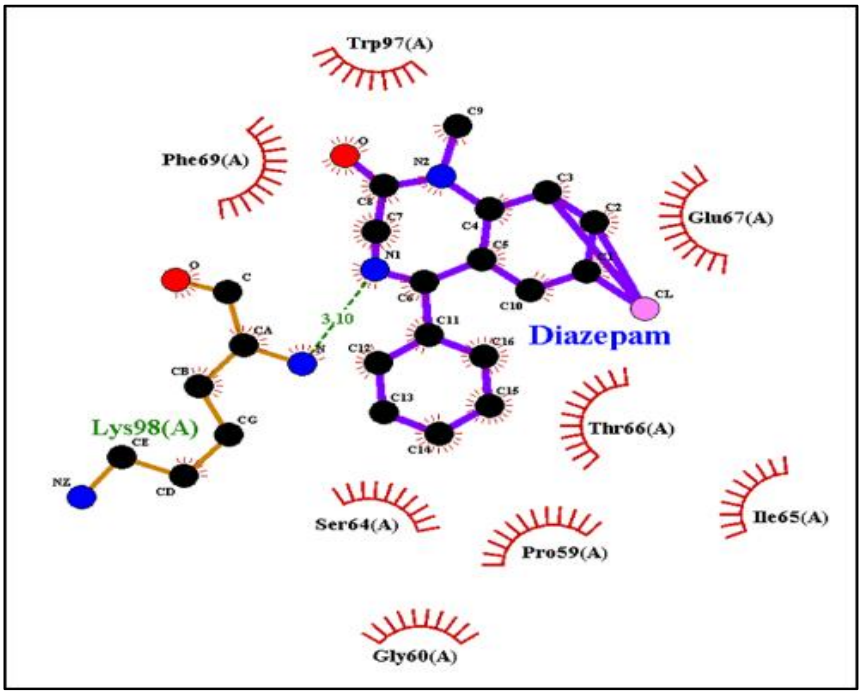
A thorough identification and documentation of the different intermolecular interactions between the ligands and the receptor were done after the interaction analysis of the top 20 ligands. To analyze the interaction between the GABRA2 receptor and the ligands, molecular docking studies were carried out using the 400 drugs that were retrieved after ligand based virtual screening. Out of the 400 drugs, 20 drugs exhibited a higher affinity to bind with the GABRA2 receptor, i.e., < -7 kcal/mol. These drugs also meet the evident cutoff of less than 1 Å RMSD (Root Mean Square Deviation) value. Of these 20 drugs, Zolmitriptan had the most negative binding affinity (-7.7 kcal/mol), suggesting it may form a stronger bond with the receptor compared to that of Diazepam. Table 4.2. provides an extensive analysis of these interactions, which include π-π stacking, hydrophobic interactions, electrostatic interactions, and H-bonds. The comprehensive explanation offers a significant understanding of the binding affinities and particular interaction processes that support the general stability and effectiveness of the ligand-receptor complexes.

Table 4.2. Interaction Networks of GABRA2 and Drugs

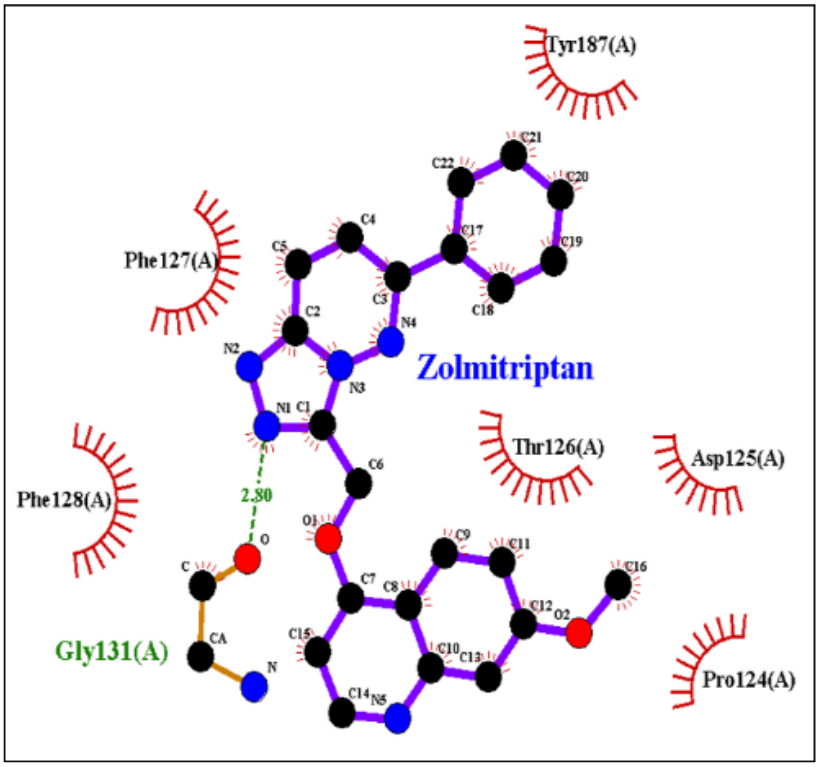
S.No.	Chembl ID	Drugs	Interacting Residues
01.	**CHEMBL12**	**Diazepam (Control)**	Pro59, Lys98, Thr153, Glu67, Thr66
02.	CHEMBL1185	Zolmitriptan	Tyr187, Leu160, Phe127, Phe128
03.	CHEMBL744	Riluzole	Thr153, Lys98, Phe69, Thr66, ILE65, GLU67
04.	CHEMBL1214124	Perampanel	GLU100, GLU67, PHE69, LYS98
05.	CHEMBL22097	Lormetazepam	LYS98, GLU67, PHE69
06.	CHEMBL3989843	Tranylcypromine	THR153, PHE69, GLU100, LYS98, LYS96, PHE69,
07.	CHEMBL1201754	Rufinamide	GLU100, LYS98, PHE69, GLU67, ILE65, SER64
08.	CHEMBL83	Tamoxifen	GLU100, THR153, LYS98, LYS96, THR66, GLU67, PHE69

Continued on next page

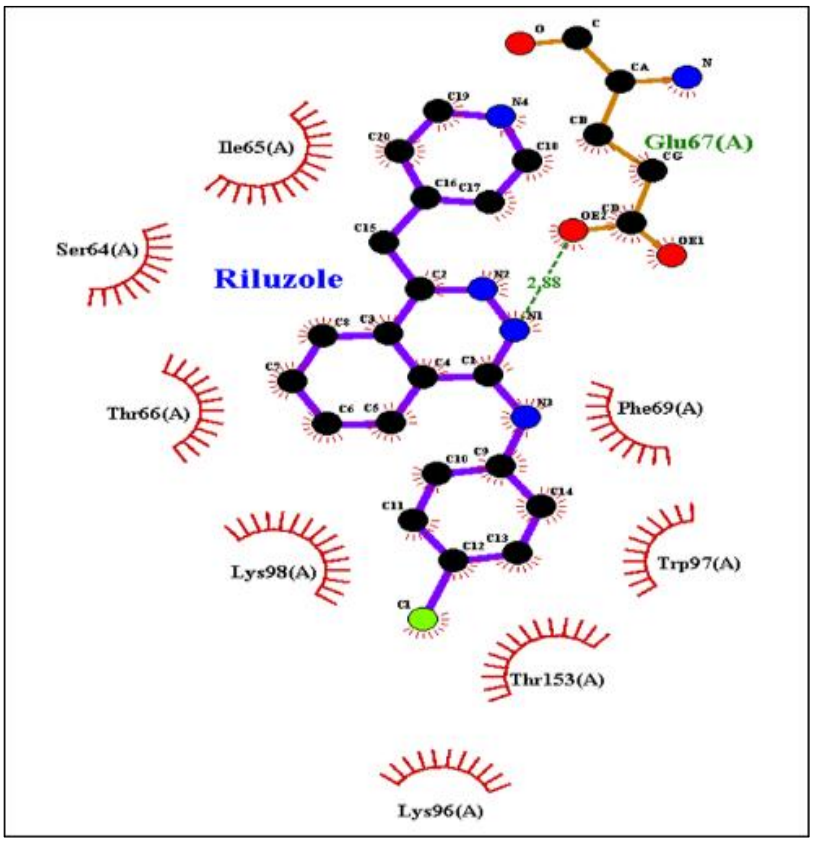
Table 4.2 (Continued)			
09.	CHEMBL76	Chloroquine	THR153, LYS98, PHE69, GLU67, SER64
10.	CHEMBL1771	Clopidogrel	LYS98, PRO59, GLU67, ILE65
11.	CHEMBL12713	Sertindole	LYS98, PRO59, GLU67, THR66, PHE69
12.	CHEMBL53	Apomorphine	THR153, LYS98, LYS96, PRO59, GLU67, PHE69
13.	CHEMBL1059	Lyrica (Pregabalin)	THR153, LYS98, GLU67, SER64, PHE69
14.	CHEMBL636	Rivastigmine	GLU100, THR153, LYS98, GLU67, PHE69
15.	CHEMBL48361	Dabigatran	LEU160, PHE93, PHE128, PHE127, LYS132, LYS133
16.	CHEMBL905	Rizatriptan	PRO59, LYS98, GLU67, THR153, LYS96, PHE69
17.	CHEMBL13280	Flunitrazepam	PRO59, LYS98, GLU67, THR153, LYS96, PHE69, SER64
18.	CHEMBL1070	Nabumetone	LYS96, LYS98, ARG58, PRO59, THR66, GLU67, ILE65
19.	CHEMBL452	Clonazepam	LEU160, PHE128, SER134, PHE128, TYR187
20.	CHEMBL939	Gefitinib	ASN114, MET141, LEU145, LYS144, ASN143, ARG159, PHE92
21.	CHEMBL4297446	Triptan	GLU100, THR153, LYS98, PHE69, ILE65, THR66


Trp97(A)
Phe69(A)
Glu67(A)
Diazepam
3.10
Lys98(A)
Thr66(A)
Ser64(A)
Pro59(A)
Ile65(A)
Gly60(A)

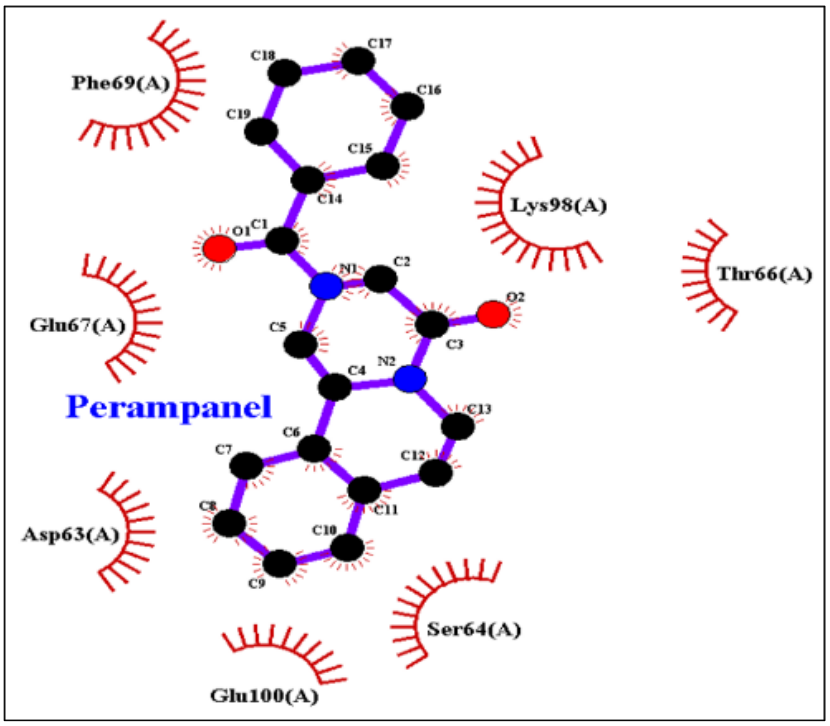
(a)


Tyr187(A)
Phe127(A)
Zolmitriptan
Thr126(A)
Asp125(A)
Phe128(A)
2.80
Gly131(A)
Pro124(A)

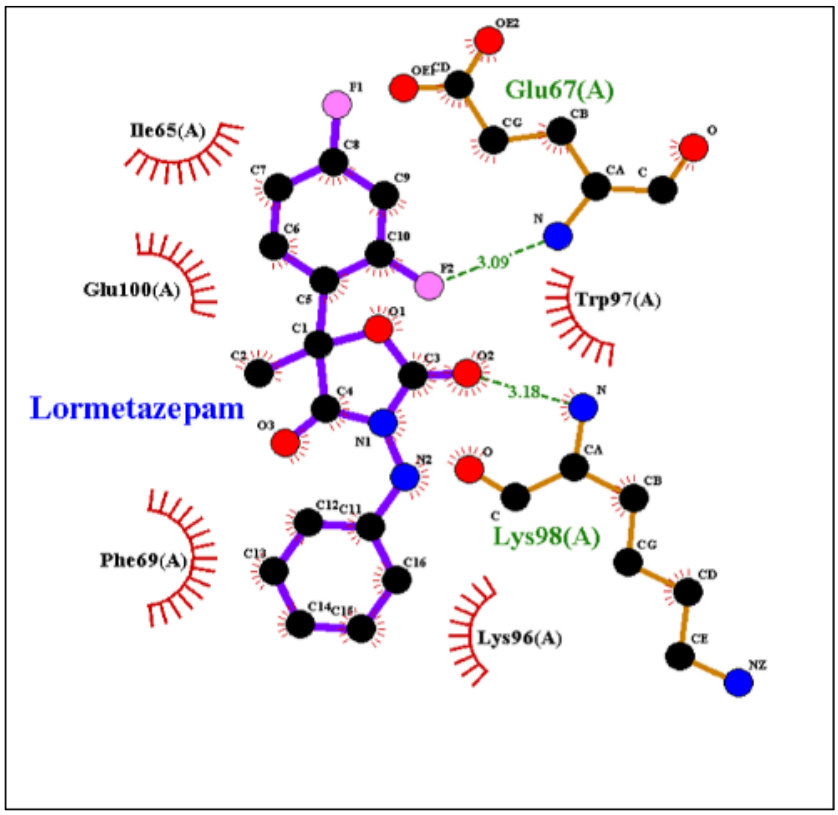
(b)


Ile65(A)
Glu67(A)
Ser64(A)
Riluzole
2.88
Thr66(A)
Phe69(A)
Lys98(A)
Trp97(A)
Thr153(A)
Lys96(A)

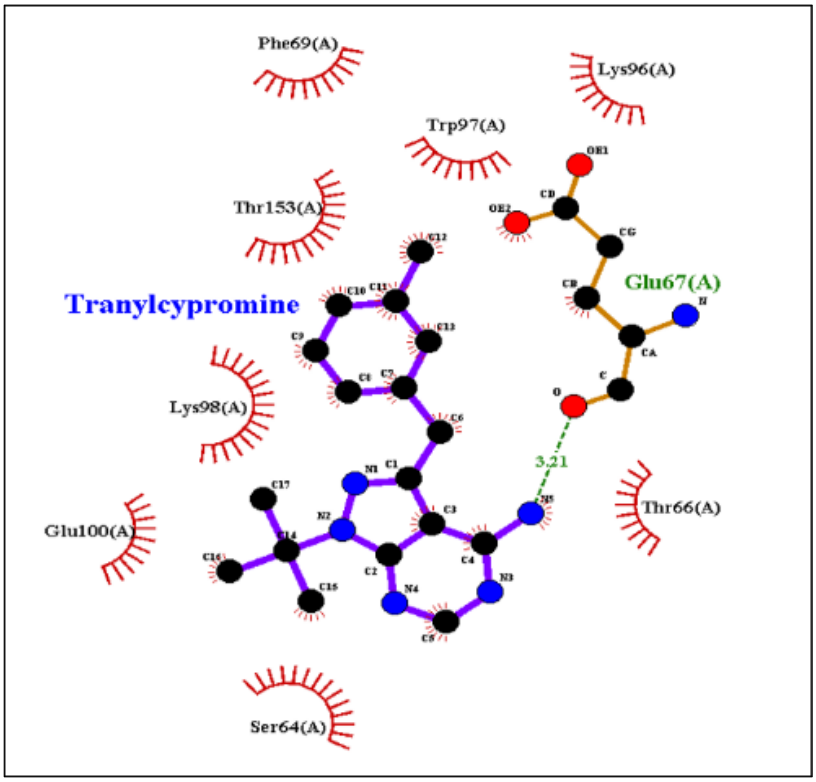
(c)


Phe69(A)
Lys98(A)
Thr66(A)
Glu67(A)
Perampanel
Asp63(A)
Ser64(A)
Glu100(A)

(d)


Glu67(A)
Ile65(A)
Glu100(A)
Trp97(A)
3.09
3.18
Lormetazepam
Lys98(A)
Phe69(A)
Lys96(A)

(e)


Phe69(A)
Lys96(A)
Trp97(A)
Thr153(A)
Glu67(A)
Tranylcypromine
Lys98(A)
3.21
Thr66(A)
Glu100(A)
Ser64(A)

(f)

Fig 4.3. 2-D structural representation of GABRA2 residues interacting to the (a) Diazepam, a reference drug (b) Zolmitriptan (c) Riluzole (d) Perampanel (e) Lormetazepam (f) Tranylcypromine

4.4. Pharmacokinetic (ADMET) Profiling

The strongest affinity ligands were subjected to pharmacokinetic study and evaluation of blood-brain barrier (BBB) permeability in order to find possible targets for Anxiety. The ligands' BBB permeability and pharmacokinetic characteristics were assessed using software program like SwissADME. As shown in the Table 2, all the ligands were abiding to the Lipinski's rule which had a high Gastrointestinal absorption (GI), it is studied to check the oral bioavailability of the drug. According to the rule, the ligand must meet other molecular properties, such as molecular weight (MW), molecular refractivity (MR), and lipophilicity (log Kp). In this study, MW < 500 Da, MR < 200 and log Kp < 5 were considered. These properties indicates that Riluzole and Tranylcypromine showed a few lead likeliness violations which makes them unfit to be taken as a drug. Besides Riluzole and Tranylcypromine, all other compounds showed an equivalent or better ADME assay than the control.

Following a thorough examination of a variety of pharmacokinetic and pharmacodynamic parameters, such as BBB permeability, solubility, GI absorption, adherence to drug-likeliness criteria (as outlined by Lipinski, Ghose, and Muegge), adherence to lead-likeliness criteria, hepatotoxicity, and AMES toxicity, for the top 5 ligands, it was determined that only Zolmitriptan displayed notably high binding affinity while adhering to drug-likeliness and lead-likeliness criteria. Additionally, Zolmitriptan exhibited no signs of toxicity. However, top five ligands have been shown in the given table.

Table 4.4. Physiochemical Properties of the top-five ligand molecules with highest binding affinity

Property	Zolmitriptan	Riluzole	Perampanel	Lormetazepam	Tranylcypromine	Diazepam (Control)
MW (g/mol)	287.36	234.2	349.38	335.18	133.19	284.74
MR	86.25	50.71	105.42	94.12	41.62	87.95
TPSA (Å²)	57.36	76.38	58.68	52.9	26.02	32.67
Ilogp	2.38	2.04	3.15	2.36	1.85	2.68

Table 4.4 (Continued)						
GI absorption	High	High	High	High	High	High
BBB permeability	Yes	Yes	Yes	Yes	Yes	Yes
Log Kp (cm/s)	-6.51	-5.17	-5.99	-6.61	-6.05	-5.91
Lipinski	Yes	Yes	Yes	Yes	Yes	Yes
Bioavailability score	0.55	0.55	0.55	0.55	0.55	0.55
Leadlikeliness	No	Yes	No	N	Yes	No

DISCUSSION

Anxiety, a neuropsychiatric disorder has been living undercover among the society and affecting almost all of us. Since, the symptoms of this disorder are not severe, this disorder is not considered of that much importance and is disregarded often times but it is a matter of concern as it affects the neurochemical systems of our brain. Anxiety is said to induce a state of heightened arousal and vigilance in brain circuits. Currently a widely used anxiolytic drug "Diazepam" is used to target a subunit of GABA/alpha-2 receptor (GABRA2) to inhibit the neuronal excitability in the brain but this drug is reported to have several side-effects ranging from common to severe to long term.

The adverse effects of diazepam commonly include drowsiness, confusion and problems in co-ordination while severe effects include hallucinations, delusions, sedative effects and unusual mood swings. On a longer diazepam can also cause withdrawal symptoms. Hence it is worth mentioning that the current cure to anxiety isn't as safe as it seems and there is a need to identify new alternatives to diazepam that can efficiently inhibit the GABRA2 receptor while having minimal side effects.

In this study we followed an intricate and efficient computational workflow to identify alternatives to diazepam. Ligand based virtual screening provided us with 400 compounds that have a similar structure to that of Diazepam. Following that we retrieved the 3D structures of GABRA2 receptor and the screened drugs. The final ligand and protein files were saved in pdbqt format after all heteroatoms were removed from the structure and polar hydrogens and Kollman/Gasteiger charges were added. These steps were done using Autodock tools to prepare the data for molecular docking. Molecular docking analysis revealed that 20 drugs exhibited a higher binding affinity to bind with the GABRA2 receptor than Diazepam. We further screened the ligand to top 5 on the basis of most negative binding affinities. Among those top 5 compounds, each passed Lipinski rule of five and exhibited high gastro-intestinal absorption, while Riluzole and Tranylcypromine showed violations of possible lead likeliness.

The interaction between the GABRA2 receptor and Diazepam has not been explored in previous studies, making this research a novel contribution to the field. As a preliminary investigation, the identification of novel GABRA2 inhibitors through virtual screening opens new avenues for therapeutic development in anxiety management. While the computational findings presented in this study are promising, experimental validation remains essential to confirm the efficacy of the identified compounds. Additionally, advancements in machine learning could further refine in silico screening for GABRA2 and related targets.

RESULTS AND DISCUSSION

(Objective: 2)

4.5. Data Collection

The microarray data that was selected for the study was obtained from GEO datasets via NCBI. The GSE140380 data set had six classes namely Control (281 samples), bvFTD (80 samples), PSP (54 samples), nfvPPA (47 samples), svPPA (44 samples) and CBS (36 samples). Notably, as shown in the Fig. 5.1., bvFTD accounts for the highest proportion (50–70%), while PSP and CBS each represent around 25%, highlighting the heterogeneity within NDDs. This figure illustrates the primary subtypes of dementia, including bvFTD, PPA, PSP, and CBS, each varying in prevalence and symptom expression.

4.6. Data Preprocessing

The gene expression data underwent quantile normalization to standardize expression levels across all samples, ensuring uniformity and facilitating comparability between them. This step was crucial for mitigating batch effects and reducing technical variability inherent in high-throughput experiments. Following normalization, a density map was generated to visually depict the distribution of expression levels across samples, providing a comprehensive view of the variation in the dataset. The analysis also revealed a class imbalance due to unequal sample sizes in different categories, which could have introduced bias in downstream analysis. To address this SMOTE was applied, which effectively balanced the dataset by generating synthetic samples for the underrepresented classes, thus ensuring that the classification models were not biased toward the majority class and maintained robust performance across all categories.

4.7. ML Implementation and Model Evaluation

The ML implementation and model evaluation elucidated varying levels of performance across the models tested in the gene expression dataset. Stratified K-Fold cross-validation, alongside test set evaluation was employed to ensure robust and unbiased results. For each model, key performance measures, including accuracy, precision, recall, and F1-score, were computed, while their performance was also analyzed using confusion matrices and classification reports.

Logistic Regression performed moderately, achieving a cross-validation accuracy of 86% (±0.00) and a test accuracy of 85%, indicating a consistent performance. While effective for linearly separable problems, its limitations to capture non-linear relationships in high-dimensional data may have constrained its performance. The

SVC with a linear kernel demonstrated exceptional performance, with a cross-validation accuracy of 98% (±0.00) and a test accuracy of 99%. This model effectively leveraged the linear separability of the dataset, achieving the highest accuracy of among the models tested. Naive Bayes achieved a cross-validation accuracy of 63% (±0.04) and test accuracy of 67%, despite its independency assumption. Its relatively weak performance indicates its limitations in handling the complex interdependencies inherent in high-dimensional gene expression data.

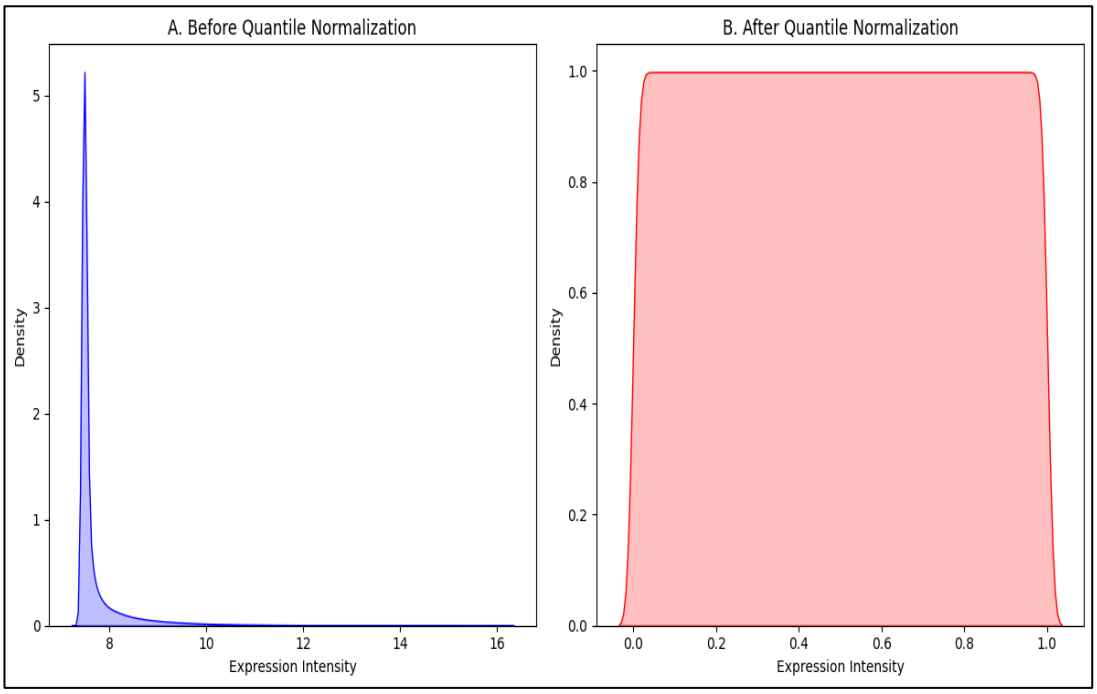


Fig 5.2. Comparison of data distributions before and after quantile normalization.

MLP Classifier achieved good performance with a cross-validation accuracy of 91% (±0.01) and a test accuracy of 93%. The ability to model complex, it effectively captured non-linear relationships thus the model demonstrated consistent and reliable performance across both training and testing. The RF Classifier also yielded appreciable results, achieving a cross-validation accuracy of 93% (±0.01) and a test accuracy of 94%. Notably, RF excelled in feature selection by providing interpretable feature importance rankings, crucial for identifying key biomarkers within the dataset. Meanwhile, the RF classifier offered valuable insights into feature importance, making it a promising choice for identifying potential biomarkers.

These findings in Table I and Fig 3. highlights the effectiveness of ML approaches in analyzing gene expression data and demonstrates the importance of model selection based on dataset's characteristics and objectives of the analysis.

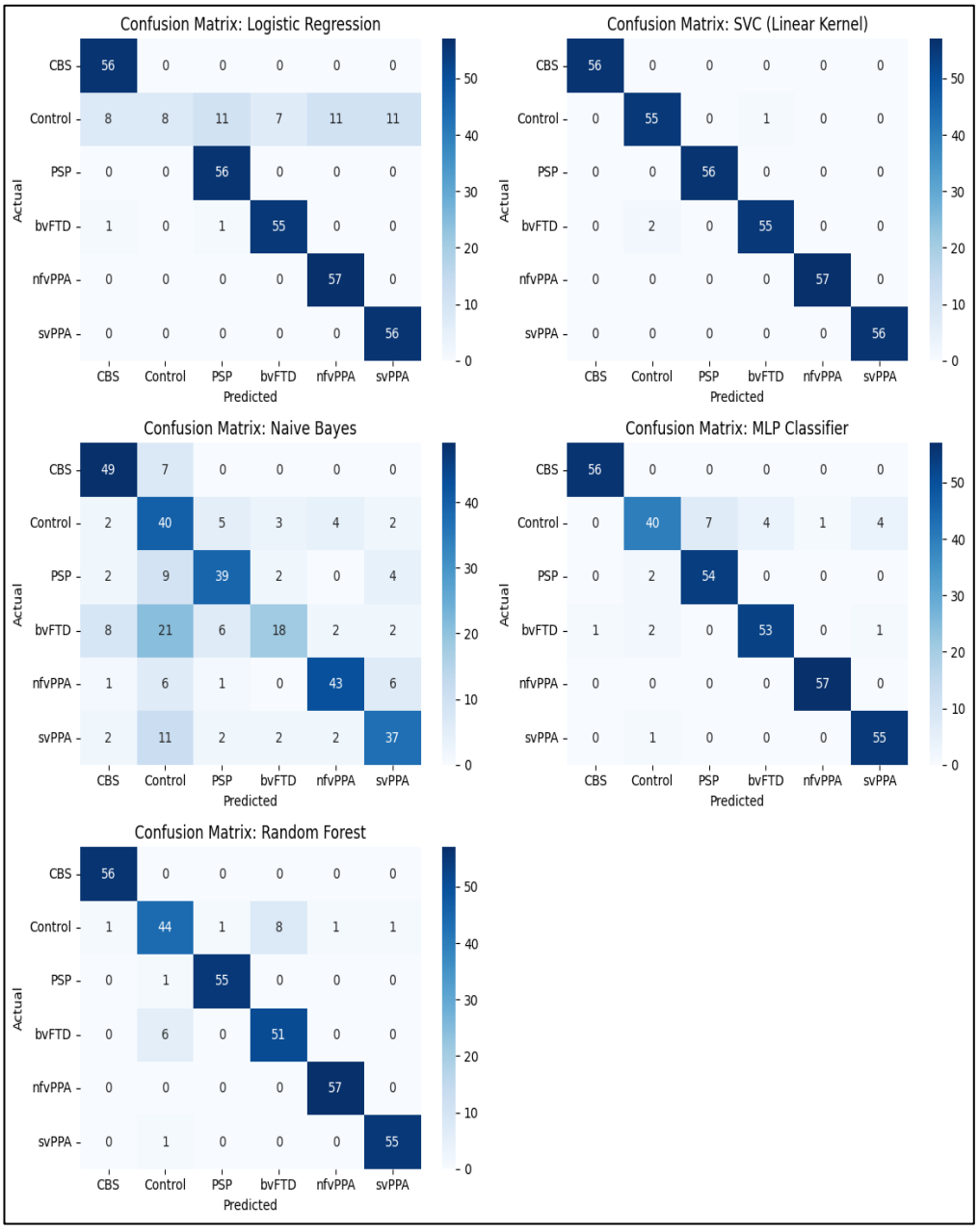


Fig 5.3. Comparative analysis of confusion matrices across 5 different machine learning classifiers: Logistic Regression, SVC (Linear Kernel), Naive Bayes, MLP Classifier, and Random Forest. Each matrix displays the classification performance across five classes (CES, Control, PSP, bvFTD, nfvPPA).

Table 5.1. Classification Performance of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-score
SVC (Linear Kernel)	**0.99**	**0.99**	**0.99**	**0.99**
Logistic Regression	0.85	0.87	0.85	0.80
MLP Classifier	0.93	0.93	0.93	0.93
Naïve Bayes	0.67	0.70	0.67	0.67
Random Forest	0.94	0.94	0.94	0.94
SGD Classifier	0.92	0.92	0.92	0.92

4.8. Biomarker Identification Analysis

The analysis successfully identified the top 20 influential genes exerting impact on classification outcomes using the RF classifier. These obtained features were ranked based on their importance scores and visualized in the bar plot. MOBP emerged as the most critical biomarker, followed by PDIA5 and RAB43.2. The importance scores for these genes suggest their significant role in distinguishing between the target classes.

To gain more granular insights, a binary classification approach was adopted for each target class. This allowed the identification of class-specific biomarkers, highlighting unique features most relevant to each class. The identified biomarkers were systematically compiled into a comprehensive file to provide a holistic view of potential biological indicators. In Fig 5.4. it is shown that how these results highlight key genetic features that could play a vital role in the studied classification task and offer valuable biological insights for future research and validation studies.

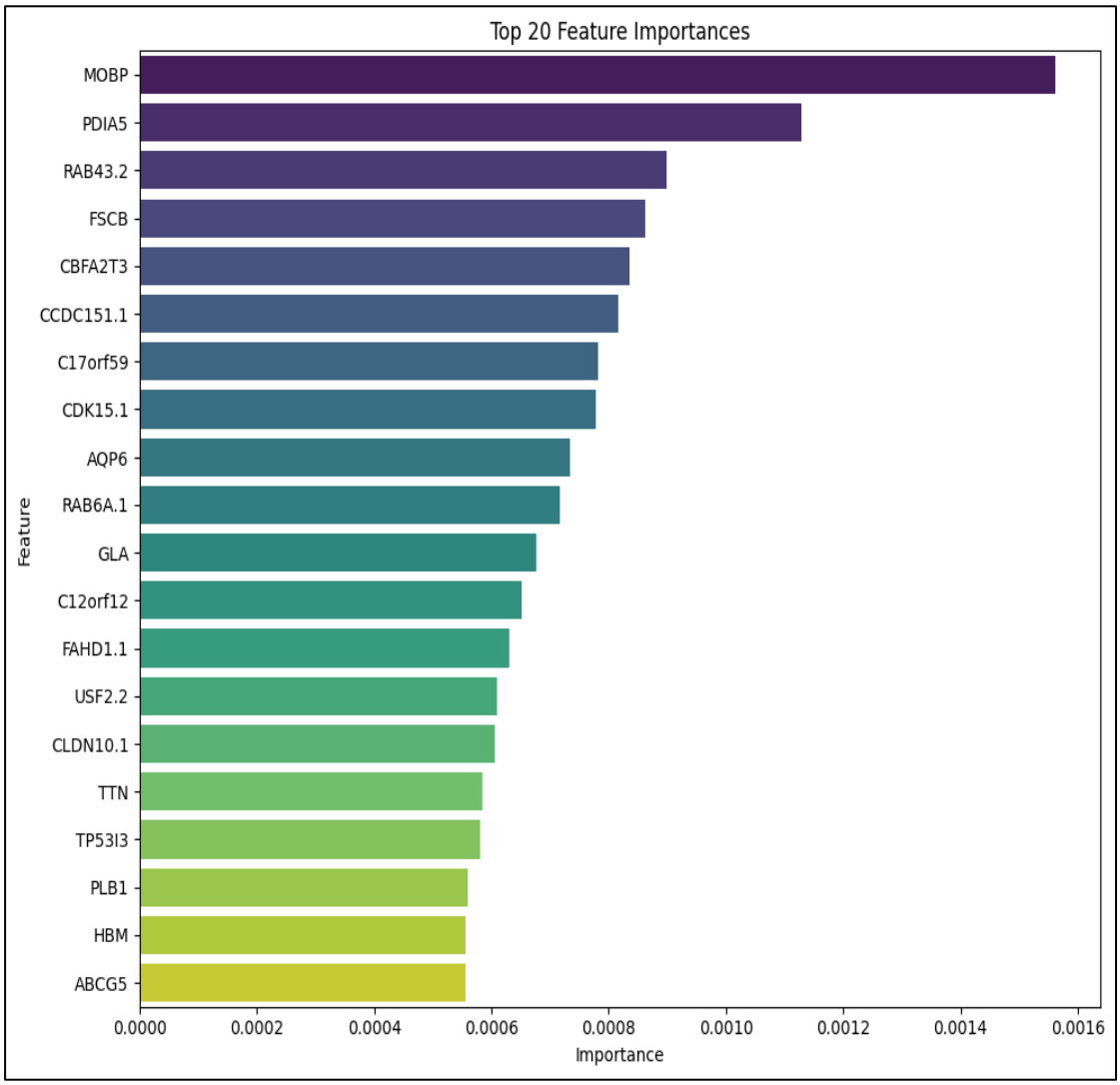


Fig 5.4. Top 20 feature importance scores derived from ML analysis

4.9. Pathway Enrichment Analysis of Identified Biomarkers

From Fig. 5.5., we can see that the enriched pathways such as axon guidance and neuroactive ligand-receptor interaction suggest impaired neuronal connectivity, particularly relevant to bvFTD and PPA, while protein processing and ubiquitin mediated proteolysis indicate disrupted protein clearance, a hallmark of CBS and PSP. Additionally, pathways linked to oxidative stress (HIF-1 signalling, apoptosis) and metabolic dysfunction (TGF-β signalling, central carbon metabolism) suggest systemic contributions to neurodegeneration. These findings reinforce the hypothesis

that dementia subtypes share overlapping yet distinct molecular signatures, providing potential targets for biomarker validation and therapeutic interventions.

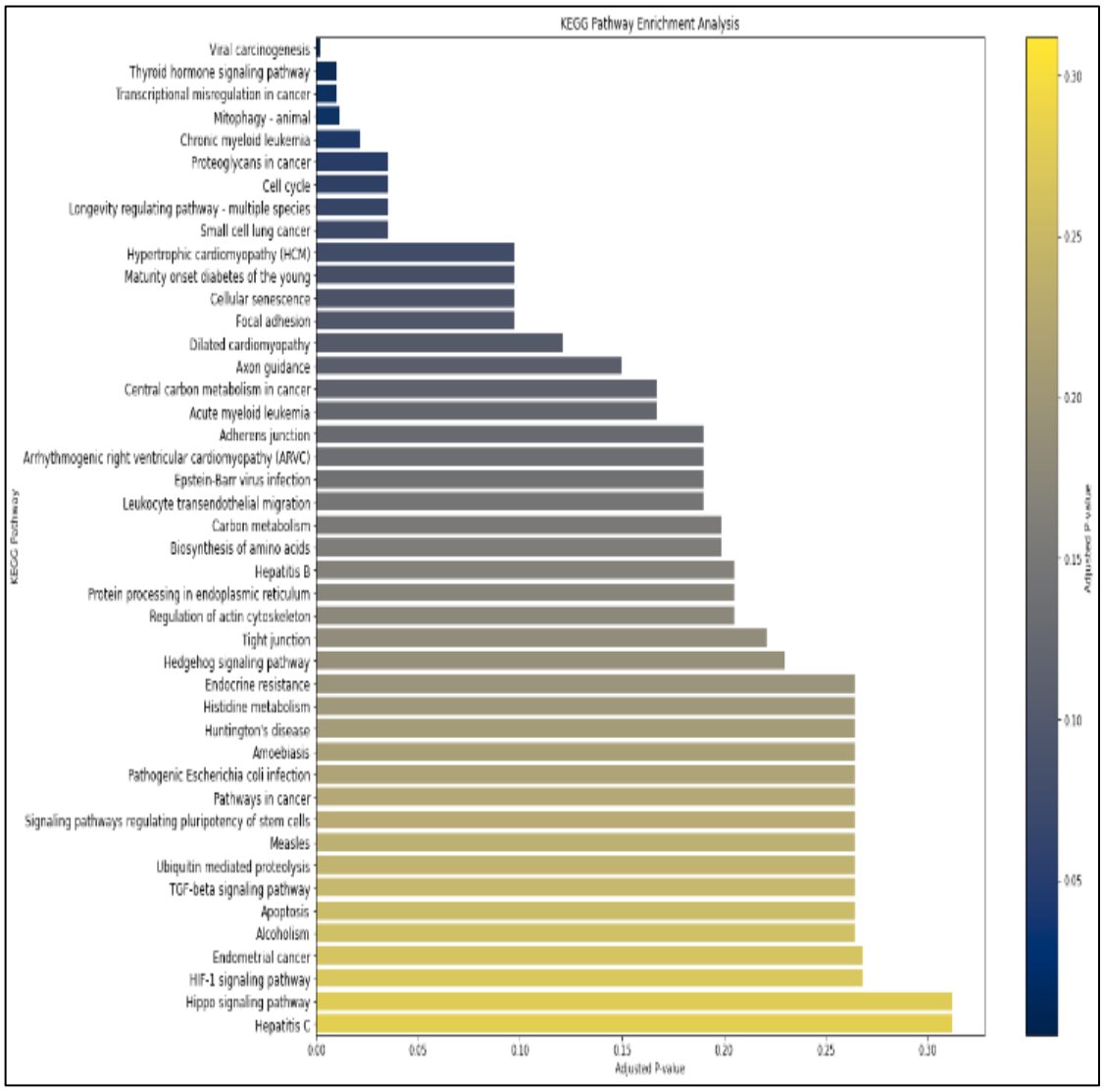


Fig 5.5. KEGG Pathway Enrichment Analysis showing significantly enriched biological pathways ranked by adjusted p-values

DISCUSSION

The present study highlights the pivotal role of gene expression analysis with ML methodologies in advancing the understanding and differentiation of dementia subtypes. Using GEO database data, biomarkers with diagnostic and therapeutic potential were identified across six categories: bvFTD, PSP, CBS, nfvPPA, svPPA, and controls. Through quantile normalization, the preprocessing steps ensured data integrity and unbiased analysis, while SMOTE handled the class imbalance along with technical variability.

Among the tested ML models, the highest accuracy was achieved by the SVC with a linear kernel, reaching 99%, followed by the RF classifier at 94%. These findings are consistent with recent advancements in deploying ML algorithms against intricate genomic datasets, with SVC recognized for its prediction accuracy in classification tasks and RF distinguished for its ability to identify feature importance. SVC outperformed other classifiers, while RF excelled in biomarker identification, highlighting its value in gene expression studies for predictive accuracy.

The identified biomarkers, including MOBP, PDIA5, and RAB43.2, align with merging research on molecular markers in neurodegenerative diseases. MOBP is linked to myelin-associated processes central to these conditions. Pathway enrichment analysis highlights axon guidance and neuroactive ligand-receptor interaction in bvFTD and PPA, while disrupted protein processing pathways distinguish CBS and PSP. These findings emphasize both shared and distinct molecular mechanisms across dementia subtypes.

Recent advancements underscore the transformative role of ML in precision medicine, particularly in subtype-specific diagnostics and targeted therapies. Predictive analytics, for example, can now accurately differentiate bvFTD from AD, paving the way for personalized treatment strategies. Future research should emphasize on validation using independent cohorts and preferably multi-omics datasets to contribute to a comprehensive understanding of dementia pathophysiology. Furthermore, enhancing the interpretability of ML models is crucial for clinical implementation, in which explainability of ML model is vital for informed decision-making. The combination of ML approaches and gene expression analysis has proved to be a promising pathway for advancing dementia research, both in the context of deciphering the molecular underpinnings of the disease and for predicting its clinical course. This establishes a framework toward growth in precision medicine for neurodegenerative disorders through discovery of robust biomarkers and delineation of molecular mechanisms. The findings mark an important advance for improved diagnostics and targeted interventions in dementia care.

CHAPTER – 5

CONCLUSION, FUTURE PERSPECTIVES AND SOCIAL IMPACT

NDDs presents an increasingly critical and global epidemic and require intensive research to improve effective curative interventions. The present study aimed to address this challenge through a dual-computational strategy: virtual screening and gene expression analysis, and have been independently explored in NDDs. This thesis is one of the first to present a converged in silico strategy, enhancing our understanding of both therapeutic targeting and disease subtype classification.

The first objective led to the identification of several drug-like compounds with higher binding affinities toward GABRA2, along with favorable pharmacokinetic properties where this study aims to suggest alternatives to Diazepam for anxiety treatment by targeting the GABRA2 receptor which concludes that Zolmitriptan, along with others compounds are identified to exhibit stronger binding affinities than Diazepam to bind with the GABRA2 receptor, indicating their potential as alternatives treatments for anxiety. Using advanced computational methods such as homology modelling, virtual screening and molecular docking, we successfully identified safer drug candidates that could reduce side effects associated with Diazepam. While these findings are promising, there is need for further experimental validation to confirm the therapeutic potential. Overall, this study lays a thorough groundwork for developing more effective and safer therapies for anxiety by targeting the GABRA2 receptor.

The second objective successfully demonstrated the use of ML models (SVC, RF) in achieving high classification accuracy (up to 99%) for dementia subtypes, and uncovered biologically relevant genes (e.g., MOBP, PDIA5, RAB43) linked to key molecular pathways where this study demonstrates the combination of gene expression analysis with ML to classify different subtypes of dementia and discover biomarkers with diagnostic and therapeutic relevance. By leveraging the GEO database and applying a rigorous data preprocessing approach, we fortified the data to ensure it was free from biasness. The SVC out of all the assessed models achieved the best accuracy (99%) while the RF classifier was helpful for biomarker selection. Biomarker(s) identified i.e. MOBP, PDIA5 and RAB43.2, complement emerging studies in neurodegenerative disorders, bringing attention to both common and distinct pathophysiological pathways among dementia subtypes.

These results show the ML's potential to revolutionize precision medicine, increasing diagnostic accuracy and refining treatment strategies. Future studies should aim to

validate these biomarkers in independent cohorts and may adopt a multi-omics approach to construct a more extensive picture of dementia pathophysiology. In addition, to enable clinical translation, improving the interpretability of ML models will also be key, ensuring their applicability to real world decision-making. By advancing computational and molecular insights, this study builds the foundation for precision medicine in dementia research, paving the way for improved and novel potential diagnostics and personalized interventions.

Integrating these complementary findings, this work presents a novel and cohesive computational workflow for addressing critical gaps in the understanding of NDDs. Together, these findings offer both therapeutic and diagnostic insights into NDDs.

CHAPTER – 6

REFERENCES

[1] A. R. Monteiro, D. J. Barbosa, F. Remião, and R. Silva, "Alzheimer's disease: Insights and new prospects in disease pathophysiology, biomarkers and disease-modifying drugs," *Biochem Pharmacol*, vol. 211, p. 115522, May 2023, doi: 10.1016/j.bcp.2023.115522.

[2] L. Fornari Laurindo *et al.*, "Immunological dimensions of neuroinflammation and microglial activation: exploring innovative immunomodulatory approaches to mitigate neuroinflammatory progression," *Front Immunol*, vol. 14, Jan. 2024, doi: 10.3389/fimmu.2023.1305933.

[3] "World Health Organization (WHO)." Accessed: May 11, 2025. [Online]. Available: https://www.who.int/

[4] "Dementia." Accessed: May 11, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia

[5] H. Kam and H. Jeong, "Pharmacogenomic Biomarkers and Their Applications in Psychiatry," *Genes (Basel)*, vol. 11, no. 12, p. 1445, Nov. 2020, doi: 10.3390/genes11121445.

[6] Q. Su *et al.*, "Multi-omics analysis reveals GABAergic dysfunction after traumatic brainstem injury in rats," *Front Neurosci*, vol. 16, Nov. 2022, doi: 10.3389/fnins.2022.1003300.

[7] S. Yi, L. Wang, H. Wang, M. S. Ho, and S. Zhang, "Pathogenesis of α-Synuclein in Parkinson's Disease: From a Neuron-Glia Crosstalk Perspective," *Int J Mol Sci*, vol. 23, no. 23, p. 14753, Nov. 2022, doi: 10.3390/ijms232314753.

[8] A. Nott and I. R. Holtman, "Genetic insights into immune mechanisms of Alzheimer's and Parkinson's disease," *Front Immunol*, vol. 14, Jun. 2023, doi: 10.3389/fimmu.2023.1168539.

[9] R. W. Olsen and W. Sieghart, "GABAA receptors: Subtypes provide diversity of function and pharmacology," *Neuropharmacology*, vol. 56, no. 1, pp. 141–148, Jan. 2009, doi: 10.1016/j.neuropharm.2008.07.045.

[10] J. You *et al.*, "Investigation into the Association between Neurotransmitters, Immune Features, and Lung Adenocarcinoma: A Multi-Omics Approach to the Identification of GABA-Related Features Employing 101 Combinatorial Machine Learning Computational Frameworks," Jun. 25, 2024. doi: 10.21203/rs.3.rs-4483010/v1.

[11] P. Chaturvedi, R. Khan, P. Sahu, A. Ludhiadch, G. Singh, and A. Munshi, “Role of Omics in Migraine Research and Management: A Narrative Review,” *Mol Neurobiol*, vol. 59, no. 9, pp. 5809–5834, Sep. 2022, doi: 10.1007/s12035-022-02930-3.

[12] E. C. B. Johnson *et al.*, “Large-scale proteomic analysis of Alzheimer’s disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation,” *Nat Med*, vol. 26, no. 5, pp. 769–780, May 2020, doi: 10.1038/s41591-020-0815-6.

[13] A. F. Khan and Y. Iturria-Medina, “Beyond the usual suspects: multi-factorial computational models in the search for neurodegenerative disease mechanisms,” *Transl Psychiatry*, vol. 14, no. 1, p. 386, Sep. 2024, doi: 10.1038/s41398-024-03073-w.

[14] M. Mottaqi, P. Zhang, and L. Xie, “Integrating Interpretable Machine Learning and Multi-omics Systems Biology for Personalized Biomarker Discovery and Drug Repurposing in Alzheimer’s Disease,” Mar. 28, 2025. doi: 10.1101/2025.03.24.644676.

[15] S. Nisar and M. Haris, “Neuroimaging genetics approaches to identify new biomarkers for the early diagnosis of autism spectrum disorder,” *Mol Psychiatry*, vol. 28, no. 12, pp. 4995–5008, Dec. 2023, doi: 10.1038/s41380-023-02060-9.

[16] D. Jeremic, L. Jiménez-Díaz, and J. D. Navarro-López, “Targeting epigenetics: A novel promise for Alzheimer’s disease treatment,” *Ageing Res Rev*, vol. 90, p. 102003, Sep. 2023, doi: 10.1016/j.arr.2023.102003.

[17] Y. Han *et al.*, “The fusion of multi-omics profile and multimodal EEG data contributes to the personalized diagnostic strategy for neurocognitive disorders,” *Microbiome*, vol. 12, no. 1, p. 12, Jan. 2024, doi: 10.1186/s40168-023-01717-5.

[18] Y. Hu, Y. Liu, Q. Zhu, Y. Chen, and Y. Zeng, “Identification of Novel Biomarkers Related to Vesicle Trafficking in Alzheimer’s Disease Using Bioinformatics Approaches,” *Neurochem Res*, vol. 50, no. 3, p. 157, Jun. 2025, doi: 10.1007/s11064-025-04410-1.

[19] I. Gyertyán, “How can preclinical cognitive research further neuropsychiatric drug discovery? Chances and challenges,” *Expert Opin Drug Discov*, vol. 15, no. 6, pp. 659–670, Jun. 2020, doi: 10.1080/17460441.2020.1739645.

[20] M. Grigore *et al.*, “Biomarkers of cognitive and memory decline in psychotropic drug users,” *J Neural Transm*, vol. 132, no. 1, pp. 39–59, Jan. 2025, doi: 10.1007/s00702-024-02837-4.

[21] “Automated quantitative structural magnetic resonance imaging volumetrics towards clinical application in Alzheimer’s disease”.

[22] A. Antonioni, E. M. Raho, E. Granieri, and G. Koch, “Frontotemporal dementia. How to deal with its diagnostic complexity?,” *Expert Rev Neurother*, vol. 25, no. 3, pp. 323–357, Mar. 2025, doi: 10.1080/14737175.2025.2461758.

[23] L. Donato, D. Mordà, C. Scimone, S. Alibrandi, R. D’Angelo, and A. Sidoti, “How Many Alzheimer–Perusini’s Atypical Forms Do We Still Have to Discover?,” *Biomedicines*, vol. 11, no. 7, p. 2035, Jul. 2023, doi: 10.3390/biomedicines11072035.

[24] “Neurobiology of Disease - Google Books.” Accessed: May 11, 2025. [Online]. Available: https://books.google.co.in/books?id=beDeDAAAQBAJ&pg=PA107&redir_esc=y#v=onepage&q&f=false

[25] D. Martino, A. J. Espay, A. Fasano, and F. Morgante, “Disorders of Movement,” 2016, doi: 10.1007/978-3-662-48468-5.

[26] J. You *et al.*, “Investigation into the Association between Neurotransmitters, Immune Features, and Lung Adenocarcinoma: A Multi-Omics Approach to the Identification of GABA-Related Features Employing 101 Combinatorial Machine Learning Computational Frameworks,” Jun. 25, 2024. doi: 10.21203/rs.3.rs-4483010/v1.

[27] S. Koga, K. A. Josephs, I. Aiba, M. Yoshida, and D. W. Dickson, “Neuropathology and emerging biomarkers in corticobasal syndrome,” *J Neurol Neurosurg Psychiatry*, vol. 93, no. 9, pp. 919–929, Sep. 2022, doi: 10.1136/jnnp-2021-328586.

[28] “XI Convegno Nazionale: Firenze, 17–19 marzo 2016, Palazzo dei Congressi – Villa Vittoria,” *Journal of Alzheimer’s Disease*, vol. 52, no. s1, pp. S1–S97, May 2016, doi: 10.3233/JAD-169001.

[29] C. Wang, J. Charles, G. Morrison, T. Wotherspoon, and C. Preuss, “Pharmacology of Human Anxiety,” *Anxiety, Gut Microbiome, and Nutraceuticals: Recent Trends and Clinical Evidence*, pp. 85–107, Jan. 2023, doi: 10.1201/9781003333821-5/PHARMACOLOGY-HUMAN-ANXIETY-CARRIE-WANG-JONATHAN-CHARLES-GRANT-MORRISON-THOMAS-WOTHERSPOON-CHARLES-PREUSS.

[30] “An Update on Anxiety Disorders: Etiological, Cognitive & Neuroscientific Aspects - Marwa Azab - Google Books.” Accessed: May 11, 2025. [Online]. Available: https://books.google.co.in/books?id=1hCcEAAAQBAJ&redir_esc=y

[31] A. J. Rosellini and T. A. Brown, “Anxiety and Fear-Related Disorders: Generalized Anxiety Disorder,” *Tasman’s Psychiatry*, pp. 1–36, 2023, doi: 10.1007/978-3-030-42825-9_74-1.

[32] B. Bandelow *et al.*, “Biological markers for anxiety disorders, OCD and PTSD: A consensus statement. Part II: Neurochemistry, neurophysiology and neurocognition,” *The World Journal of Biological Psychiatry*, vol. 18, no. 3, pp. 162–214, Apr. 2017, doi: 10.1080/15622975.2016.1190867.

[33] L. Hantsoo and C. N. Epperson, “Anxiety Disorders Among Women: A Female Lifespan Approach,” *Focus (Madison)*, vol. 15, no. 2, pp. 162–172, Apr. 2017, doi: 10.1176/appi.focus.20160042.

[34] M. Katherine, “Anxiety Disorders in Women: Gender-Related Modulation of Neurobiology and Behavior”.

[35] M. Pillerová *et al.*, “Molecular actions of sex hormones in the brain and their potential treatment use in anxiety disorders,” *Front Psychiatry*, vol. 13, Sep. 2022, doi: 10.3389/fpsyt.2022.972158.

[36] A. Oake and Y. V. Pathak, “Anxiety Disorders: Background, Anatomy, and Pathophysiology,” *Anxiety, Gut Microbiome, and Nutraceuticals: Recent Trends and Clinical Evidence*, pp. 33–47, Jan. 2023, doi: 10.1201/9781003333821-2/ANXIETY-DISORDERS-BACKGROUND-ANATOMY-PATHOPHYSIOLOGY-ASHLEY-OAKE-YASHWANT-PATHAK.

[37] J. R. F. Lisboa, J. D. R. Souza, F. V. Gomes, F. S. Guimarães, and J. A. S. Crippa, “Biomarkers in Anxiety Disorders,” in *Biomarkers in Neuropsychiatry*, Cham: Springer International Publishing, 2023, pp. 233–265. doi: 10.1007/978-3-031-43356-6_15.

[38] M. Hoefer *et al.*, “Fear conditioning in frontotemporal lobar degeneration and Alzheimer’s disease,” *Brain*, vol. 131, no. 6, pp. 1646–1657, Jan. 2008, doi: 10.1093/brain/awn082.

[39] K. Rascovsky *et al.*, “Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia,” *Brain*, vol. 134, no. 9, pp. 2456–2477, Sep. 2011, doi: 10.1093/brain/awr179.

[40] A. Antonioni, E. M. Raho, E. Granieri, and G. Koch, “Frontotemporal dementia. How to deal with its diagnostic complexity?,” *Expert Rev Neurother*, vol. 25, no. 3, pp. 323–357, Mar. 2025, doi: 10.1080/14737175.2025.2461758.

[41] M. L. Gorno-Tempini *et al.*, “Classification of primary progressive aphasia and its variants,” *Neurology*, vol. 76, no. 11, pp. 1006–1014, Mar. 2011, doi: 10.1212/WNL.0b013e31821103e6.

[42] K. A. Jellinger, “Pathomechanisms of neuropsychiatric disturbances in atypical parkinsonian disorders: a current view,” *J Neural Transm*, vol. 132, no. 4, pp. 495–518, Apr. 2025, doi: 10.1007/s00702-025-02890-7.

[43] A. L. Seritan, “Advances in the Diagnosis and Management of Psychotic Symptoms in Neurodegenerative Diseases: A Narrative Review,” *J Geriatr Psychiatry Neurol*, vol. 36, no. 6, pp. 435–460, Nov. 2023, doi: 10.1177/08919887231164357.

[44] J. Jalil, D. Volle, T. Zhu, and M. Sassounian, “Depression, Anxiety, and Other Mood Disorders,” in *Geriatric Medicine*, Cham: Springer International Publishing, 2024, pp. 1111–1153. doi: 10.1007/978-3-030-74720-6_88.

[45] J. Trettel, Z. Chemali, and K. R. Daffner, “Managing Neuropsychiatric Symptoms of Neurodegenerative Diseases,” in *Neurodegenerative Disorders*, Cham: Springer International Publishing, 2016, pp. 233–253. doi: 10.1007/978-3-319-23309-3_13.

[46] E. Küpeli Akkol, I. Tatlı Çankaya, G. Şeker Karatoprak, E. Carpar, E. Sobarzo-Sánchez, and R. Capasso, “Natural Compounds as Medical Strategies in the Prevention and Treatment of Psychiatric Disorders Seen in Neurological Diseases,” *Front Pharmacol*, vol. 12, May 2021, doi: 10.3389/fphar.2021.669638.

[47] “Dementia: What It Is, Causes, Symptoms, Treatment & TypesWhat It Is, Causes, Symptoms, Treatment & Types.” Accessed: May 16, 2025. [Online]. Available: https://my.clevelandclinic.org/health/diseases/9170-dementia

[48] “Dementia.” Accessed: Oct. 15, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia

[49] “Primary Care Dementia Data, January 2025 - NHS England Digital.” Accessed: May 16, 2025. [Online]. Available: https://digital.nhs.uk/data-and-information/publications/statistical/primary-care-dementia-data/january-2025

[50] S. Dattola, A. Ielo, G. Varone, A. Cacciola, A. Quartarone, and L. Bonanno, “Frontotemporal dementia: a systematic review of artificial intelligence approaches in differential diagnosis,” *Front Aging Neurosci*, vol. 17, Apr. 2025, doi: 10.3389/fnagi.2025.1547727.

[51] K. Poonam, V. S. Kotra, R. Guha, and P. P. Chakrabarti, “Hierarchical Classification of Frontotemporal Dementia Subtypes Utilizing Tabular-to-Image Data Conversion with Deep Learning Methods,” 2025, pp. 386–401. doi: 10.1007/978-3-031-78195-7_26.

[52] S. N. Amin, A. B. Mohd, C. A. Samain, and R. A. Ghannam, “Frontotemporal dementia,” in *Essential Guide to Neurodegenerative Disorders*, Elsevier, 2025, pp. 131–146. doi: 10.1016/B978-0-443-15702-8.00008-7.

[53] S. Dattola, A. Ielo, G. Varone, A. Cacciola, A. Quartarone, and L. Bonanno, “Frontotemporal dementia: a systematic review of artificial intelligence approaches in differential diagnosis,” *Front Aging Neurosci*, vol. 17, Apr. 2025, doi: 10.3389/fnagi.2025.1547727.

[54] A. A. Villaseñor and L. G. Saidi, “EXPLORING THE DIAGNOSTIC AND THERAPEUTIC JOURNEY OF PRIMARY PROGRESSIVE APHASIA: A CASE STUDY FROM A RURAL SETTING,” *Innov Aging*, vol. 8, no. Supplement_1, pp. 954–954, Dec. 2024, doi: 10.1093/geroni/igae098.3076.

[55] J. L. Robinson *et al.*, “Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated,” *Brain*, vol. 141, no. 7, pp. 2181–2193, Jul. 2018, doi: 10.1093/brain/awy146.

[56] F. Langerscheidt, T. Wied, M. A. Al Kabbani, T. van Eimeren, G. Wunderlich, and H. Zempel, “Genetic forms of tauopathies: inherited causes and implications of Alzheimer’s disease-like TAU pathology in primary and secondary tauopathies,” *J Neurol*, vol. 271, no. 6, pp. 2992–3018, Jun. 2024, doi: 10.1007/s00415-024-12314-3.

[57] I. Cordts *et al.*, “TDP-43 Proteinopathy Specific Biomarker Development,” *Cells*, vol. 12, no. 4, p. 597, Feb. 2023, doi: 10.3390/cells12040597.

[58] N. P. Visanji, A. E. Lang, and G. G. Kovacs, “Beyond the synucleinopathies: alpha synuclein as a driving force in neurodegenerative comorbidities,” *Transl Neurodegener*, vol. 8, no. 1, p. 28, Dec. 2019, doi: 10.1186/s40035-019-0172-x.

[59] F. J. Padilla-Godínez *et al.*, “α-synuclein and tau: interactions, cross-seeding, and the redefinition of synucleinopathies as complex proteinopathies,” *Front Neurosci*, vol. 19, Mar. 2025, doi: 10.3389/fnins.2025.1570553.

[60] P. J. Sampognaro *et al.*, “Mutations in α-synuclein, TDP-43 and tau prolong protein half-life through diminished degradation by lysosomal proteases,” *Mol Neurodegener*, vol. 18, no. 1, p. 29, May 2023, doi: 10.1186/s13024-023-00621-8.

[61] Y. Riku, M. Yoshida, Y. Iwasaki, G. Sobue, M. Katsuno, and S. Ishigaki, “TDP-43 Proteinopathy and Tauopathy: Do They Have Pathomechanistic Links?,” *Int J Mol Sci*, vol. 23, no. 24, p. 15755, Dec. 2022, doi: 10.3390/ijms232415755.

[62] R. Bobkier, K. Kovler, A. Tsapalov, and E. K. Czech, “‘Fusion of Horizons’: Part III. Rethinking radon Risk: Scientific advances and regulatory implications (since 1990s),” *J Environ Radioact*, vol. 286, p. 107707, Jun. 2025, doi: 10.1016/j.jenvrad.2025.107707.

[63] F. Singh, G. Ito, and F. Suomi, “Editorial: The contribution of autophagy to neuronal metabolism,” *Front Cell Dev Biol*, vol. 13, Apr. 2025, doi: 10.3389/fcell.2025.1604307.

[64] J. Burtscher, M. Kopp, M. Gassmann, and M. Burtscher, “Health benefits of life at moderate altitude: does hypoxia matter?,” *Front Physiol*, vol. 16, May 2025, doi: 10.3389/fphys.2025.1598275.

[65] M. Guarnaccia, V. La Cognata, G. Gentile, G. Morello, and S. Cavallaro, “Unraveling the missing heritability of amyotrophic lateral sclerosis: should we focus more on copy number variations?,” *Neural Regen Res*, Apr. 2025, doi: 10.4103/NRR.NRR-D-24-01604.

[66] A. Brooks *et al.*, “Dietary restrictions and pattern influence on Parkinson’s disease,” *Parkinsonism Relat Disord*, vol. 134, p. 107542, 2025, doi: 10.1016/j.parkreldis.2025.107542.

[67] P. Singh, M. Borkar, and G. Doshi, “Network pharmacology approach to unravel the neuroprotective potential of natural products: a narrative review,” *Mol Divers*, Apr. 2025, doi: 10.1007/s11030-025-11198-3.

[68] N. AL-Qawasmeh, C. Y. Suen, and E. Omar, “A Comprehensive Approach to Handwriting Analysis for Alzheimer’s Detection,” 2025, pp. 72–81. doi: 10.1007/978-3-031-87663-9_6.

[69] S. M. Sears and S. J. Hewett, “Influence of glutamate and GABA transport on brain excitatory/inhibitory balance,” *Exp Biol Med*, vol. 246, no. 9, pp. 1069–1083, May 2021, doi: 10.1177/1535370221989263.

[70] A. Kwakowsky, B. Calvo-Flores Guzmán, K. Govindpani, H. Waldvogel, and R. Faull, “Gamma-aminobutyric acid A receptors in Alzheimer’s disease: highly localized remodeling of a complex and diverse signaling pathway,” *Neural Regen Res*, vol. 13, no. 8, p. 1362, 2018, doi: 10.4103/1673-5374.235240.

[71] K. Vakili *et al.*, “Use of Drugs Affecting GABAA Receptors and the Risk of Developing Alzheimer’s Disease and Dementia: a Meta-Analysis and Literature Review,” *Mol Neurobiol*, Mar. 2025, doi: 10.1007/s12035-025-04821-9.

[72] A. Ghit, D. Assal, A. S. Al-Shami, and D. E. E. Hussein, “GABAA receptors: structure, function, pharmacology, and related disorders,” *Journal of Genetic Engineering and Biotechnology*, vol. 19, no. 1, p. 123, Dec. 2021, doi: 10.1186/s43141-021-00224-0.

[73] B. Calvo-Flores Guzmán, C. Vinnakota, K. Govindpani, H. J. Waldvogel, R. L. M. Faull, and A. Kwakowsky, “The GABAergic system as a therapeutic target for Alzheimer’s disease,” *J Neurochem*, vol. 146, no. 6, pp. 649–669, Sep. 2018, doi: 10.1111/jnc.14345.

[74] G. Carello-Collar *et al.*, “The GABAergic system in Alzheimer’s disease: a systematic review with meta-analysis,” *Mol Psychiatry*, vol. 28, no. 12, pp. 5025–5036, Dec. 2023, doi: 10.1038/s41380-023-02140-w.

[75] G. Carello-Collar *et al.*, “The GABAergic system in Alzheimer’s disease: a systematic review with meta-analysis,” *Mol Psychiatry*, vol. 28, no. 12, pp. 5025–5036, Dec. 2023, doi: 10.1038/s41380-023-02140-w.

[76] A. Ghit, D. Assal, A. S. Al-Shami, and D. E. E. Hussein, “GABAA receptors: structure, function, pharmacology, and related disorders,” *Journal of Genetic Engineering and Biotechnology*, vol. 19, no. 1, p. 123, Dec. 2021, doi: 10.1186/s43141-021-00224-0.

[77] W. Wang, W. Fu, H. Zhu, J. Ma, J. Zhang, and J. Qi, “Progress in GABAA receptor agonists for insomnia disorder,” *Front Pharmacol*, vol. 15, Nov. 2024, doi: 10.3389/fphar.2024.1432726.

[78] J. J. Kim and R. E. Hibbs, “Direct Structural Insights into GABAA Receptor Pharmacology,” *Trends Biochem Sci*, vol. 46, no. 6, pp. 502–517, Jun. 2021, doi: 10.1016/J.TIBS.2021.01.011/ASSET/24989CD9-7938-4B09-B81F-3B0849321C34/MAIN.ASSETS/GR4.JPG.

[79] J. R. Atack, “Development of Subtype-Selective GABAA Receptor Compounds for the Treatment of Anxiety, Sleep Disorders and Epilepsy,” in *GABA and Sleep*, Basel: Springer Basel, 2010, pp. 25–72. doi: 10.1007/978-3-0346-0226-6_2.

[80] S. Kaur *et al.*, “Pharmacology of GABA and Its Receptors,” in *Frontiers in Pharmacology of Neurotransmitters*, Singapore: Springer Singapore, 2020, pp. 241–292. doi: 10.1007/978-981-15-3556-7_8.

[81] I. Arora, P. Mal, P. Arora, A. Paul, and M. Kumar, “GABAergic implications in anxiety and related disorders,” *Biochem Biophys Res Commun*, vol. 724, p. 150218, Sep. 2024, doi: 10.1016/j.bbrc.2024.150218.

[82] A. Ghit, D. Assal, A. S. Al-Shami, and D. E. E. Hussein, “GABAA receptors: structure, function, pharmacology, and related disorders,” *Journal of Genetic Engineering and Biotechnology*, vol. 19, no. 1, p. 123, Dec. 2021, doi: 10.1186/s43141-021-00224-0.

[83] K. R. Tan *et al.*, “Neural bases for addictive properties of benzodiazepines,” *Nature*, vol. 463, no. 7282, pp. 769–774, Feb. 2010, doi: 10.1038/nature08758.

[84] M. Koulentaki and E. Kouroumalis, “GABAA receptor polymorphisms in alcohol use disorder in the GWAS era,” *Psychopharmacology (Berl)*, vol. 235, no. 6, pp. 1845–1865, Jun. 2018, doi: 10.1007/s00213-018-4918-4.

[85] J. I. Gall *et al.*, “Insights into serotonergic and antioxidant mechanisms involved in antidepressant-like action of 2-phenyl-3-(phenylselanyl)benzofuran in

mice," *Prog Neuropsychopharmacol Biol Psychiatry*, vol. 102, p. 109956, Aug. 2020, doi: 10.1016/j.pnpbp.2020.109956.

[86] J. W. Błaszczyk, "Parkinson's Disease and Neurodegeneration: GABA-Collapse Hypothesis," *Front Neurosci*, vol. 10, Jun. 2016, doi: 10.3389/fnins.2016.00269.

[87] R. S. Benham *et al.*, "α2-containing γ-aminobutyric acid type A receptors promote stress resiliency in male mice," *Neuropsychopharmacology*, vol. 46, no. 12, pp. 2197–2206, Nov. 2021, doi: 10.1038/s41386-021-01144-w.

[88] E. Neumann, L. Küpfer, and H. U. Zeilhofer, "The α2/α3GABAA receptor modulator TPA023B alleviates not only the sensory but also the tonic affective component of chronic pain in mice," *Pain*, vol. 162, no. 2, pp. 421–431, Feb. 2021, doi: 10.1097/j.pain.0000000000002030.

[89] S. J. Nieto, E. N. Grodin, and L. A. Ray, "On the path toward personalized medicine: implications of pharmacogenetic studies of alcohol use disorder medications," *Expert Rev Precis Med Drug Dev*, vol. 5, no. 1, pp. 43–54, Jan. 2020, doi: 10.1080/23808993.2020.1724510.

[90] R. P. Dhavale, P. B. Choudhari, and M. S. Bhatia, "Computer Assisted Models for Blood Brain Barrier Permeation of 1, 5-Benzodiazepines," *Curr Comput Aided Drug Des*, vol. 17, no. 2, pp. 187–200, Apr. 2021, doi: 10.2174/1573409916666200131114018.

[91] T. Katsila, G. A. Spyroulias, G. P. Patrinos, and M.-T. Matsoukas, "Computational approaches in target identification and drug discovery," *Comput Struct Biotechnol J*, vol. 14, pp. 177–184, 2016, doi: 10.1016/j.csbj.2016.04.004.

[92] A. A. Alsaiari *et al.*, "Chlordiazepoxide against signalling, receptor and regulatory proteins of breast cancer: a structure-based in-silico approach," *Medical Oncology*, vol. 41, no. 5, p. 117, Apr. 2024, doi: 10.1007/s12032-024-02366-w.

[93] R. T. Maia, "Protein structure prediction by computational homology modeling: a brief explanation," *International Journal of Molecular Biology Open Access*, vol. 7, no. 1, pp. 118–120, Sep. 2024, doi: 10.15406/ijmboa.2024.07.00180.

[94] T. Hameduh, Y. Haddad, V. Adam, and Z. Heger, "Homology modeling in the time of collective and artificial intelligence," *Comput Struct Biotechnol J*, vol. 18, pp. 3494–3506, 2020, doi: 10.1016/j.csbj.2020.11.007.

[95] A. V. Sadybekov and V. Katritch, "Computational approaches streamlining drug discovery," *Nature*, vol. 616, no. 7958, pp. 673–685, Apr. 2023, doi: 10.1038/s41586-023-05905-z.

[96] V. Yadav, J. Reang, Vinita, and R. K. Tonk, “Ligand-based drug design (LBDD),” in *Computer Aided Drug Design (CADD): From Ligand-Based Methods to Structure-Based Approaches*, Elsevier, 2022, pp. 57–99. doi: 10.1016/B978-0-323-90608-1.00009-5.

[97] J. Wieczorek, R. S. Malik-Sheriff, Y. Fermin, H. E. Grecco, E. Zamir, and K. Ickstadt, “Uncovering distinct protein-network topologies in heterogeneous cell populations,” *BMC Syst Biol*, vol. 9, no. 1, p. 24, Dec. 2015, doi: 10.1186/s12918-015-0170-2.

[98] M. Oh, S. Park, S. Kim, and H. Chae, “Machine learning-based analysis of multi-omics data on the cloud for investigating gene regulations,” *Brief Bioinform*, vol. 22, no. 1, pp. 66–76, Jan. 2021, doi: 10.1093/bib/bbaa032.

[99] D. E. V. Pires, T. L. Blundell, and D. B. Ascher, “pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures,” *J Med Chem*, vol. 58, no. 9, pp. 4066–4072, May 2015, doi: 10.1021/acs.jmedchem.5b00104.

[100] F. Cheng *et al.*, “admetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties,” *J Chem Inf Model*, vol. 52, no. 11, pp. 3099–3105, Nov. 2012, doi: 10.1021/ci300367a.

[101] P.-H. Chu *et al.*, “Stem Cell-Derived Endothelial Cell Model that Responds to Tobacco Smoke Like Primary Endothelial Cells,” *Chem Res Toxicol*, vol. 33, no. 3, pp. 751–763, Mar. 2020, doi: 10.1021/acs.chemrestox.9b00363.

[102] W. M. Pardridge, “Blood-Brain Barrier and Delivery of Protein and Gene Therapeutics to Brain,” *Front Aging Neurosci*, vol. 11, Jan. 2020, doi: 10.3389/fnagi.2019.00373.

[103] S. Pushpakom *et al.*, “Drug repurposing: progress, challenges and recommendations,” *Nat Rev Drug Discov*, vol. 18, no. 1, pp. 41–58, Jan. 2019, doi: 10.1038/nrd.2018.168.

[104] M. D. Sweeney, Z. Zhao, A. Montagne, A. R. Nelson, and B. V. Zlokovic, “Blood-Brain Barrier: From Physiology to Disease and Back,” *Physiol Rev*, vol. 99, no. 1, pp. 21–78, Jan. 2019, doi: 10.1152/physrev.00050.2017.

[105] M. D. Barnhart *et al.*, “Phosphorylation of the smooth muscle master splicing regulator RBPMS regulates its splicing activity,” *Nucleic Acids Res*, vol. 50, no. 20, pp. 11895–11915, Nov. 2022, doi: 10.1093/nar/gkac1048.

[106] A. C. Yang *et al.*, “Dysregulation of brain and choroid plexus cell types in severe COVID-19,” *Nature*, vol. 595, no. 7868, pp. 565–571, Jul. 2021, doi: 10.1038/s41586-021-03710-0.

[107] E. C. B. Johnson *et al.*, “Large-scale proteomic analysis of Alzheimer’s disease brain and cerebrospinal fluid reveals early changes in energy metabolism

associated with microglia and astrocyte activation," *Nat Med*, vol. 26, no. 5, pp. 769–780, May 2020, doi: 10.1038/s41591-020-0815-6.

[108] H. Mathys *et al.*, "Single-cell transcriptomic analysis of Alzheimer's disease," *Nature*, vol. 570, no. 7761, pp. 332–337, Jun. 2019, doi: 10.1038/s41586-019-1195-2.

[109] A. Spooner *et al.*, "A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction," *Sci Rep*, vol. 10, no. 1, p. 20410, Nov. 2020, doi: 10.1038/s41598-020-77220-w.

[110] M. E. AlMansoori, S. Jemimah, F. Abuhantash, and A. AlShehhi, "Predicting early Alzheimer's with blood biomarkers and clinical features," *Sci Rep*, vol. 14, no. 1, p. 6039, Mar. 2024, doi: 10.1038/s41598-024-56489-1.

[111] K. K. F. Tsoi *et al.*, "Applications of artificial intelligence in dementia research," *Cambridge Prisms: Precision Medicine*, vol. 1, p. e9, Dec. 2023, doi: 10.1017/pcm.2022.10.

[112] S. Almubark and C. Autho, "Prediction of the Conversion from Mild Cognitive Impairment to Dementia using Multilayer Perceptron and Neuropsychological Test Data," *International Journal of Innovative Research in Science, Engineering and Technology. 2023*, vol. 4, pp. 1–006, 2023, doi: 10.35248/2376-0389.23.4.04.001-006.

[113] I. Ahammad *et al.*, "AITeQ: a machine learning framework for Alzheimer's prediction using a distinctive five-gene signature," *Brief Bioinform*, vol. 25, no. 4, May 2024, doi: 10.1093/bib/bbae291.

[114] M.-K. Park *et al.*, "A Transcriptomics-Based Machine Learning Model Discriminating Mild Cognitive Impairment and the Prediction of Conversion to Alzheimer's Disease," *Cells*, vol. 13, no. 22, p. 1920, Nov. 2024, doi: 10.3390/cells13221920.

[115] R. Shukla and T. R. Singh, "AlzGenPred - CatBoost-based gene classifier for predicting Alzheimer's disease using high-throughput sequencing data," *Sci Rep*, vol. 14, no. 1, p. 30294, Dec. 2024, doi: 10.1038/s41598-024-82208-x.

[116] G. Carello-Collar *et al.*, "The GABAergic system in Alzheimer's disease: a systematic review with meta-analysis," *Mol Psychiatry*, vol. 28, no. 12, pp. 5025–5036, Dec. 2023, doi: 10.1038/s41380-023-02140-w.

[117] "Decoding Neurodegeneration: Leveraging Machine Learning Approaches to Classify Single Cells and Identify Transcriptomic Features in Alzheimer's Disease and ALS." Accessed: May 16, 2025. [Online]. Available: https://studenttheses.uu.nl/handle/20.500.12932/48441

[118] S. M. Lundberg, P. G. Allen, and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions”, Accessed: May 16, 2025. [Online]. Available: https://github.com/slundberg/shap

[119] C. Pottier *et al.*, “Deciphering distinct genetic risk factors for FTLD-TDP pathological subtypes via whole-genome sequencing,” *Nat Commun*, vol. 16, no. 1, p. 3914, Apr. 2025, doi: 10.1038/s41467-025-59216-0.

[120] H. Peng, Y. Cheng, Q. Chen, and L. Qin, “Integrated Transcriptomic and Machine Learning Analysis Identifies EAF2 as a Diagnostic Biomarker and Key Pathogenic Factor in Parkinson’s Disease,” *Int J Gen Med*, vol. Volume 17, pp. 5547–5562, Nov. 2024, doi: 10.2147/IJGM.S486214.

[121] Z. Xie *et al.*, “Identification of therapeutic targets for Alzheimer’s Disease Treatment using bioinformatics and machine learning,” *Sci Rep*, vol. 15, no. 1, p. 3888, Jan. 2025, doi: 10.1038/s41598-025-88134-w.

[122] F. Wang, Y. Liang, and Q.-W. Wang, “Interpretable machine learning-driven biomarker identification and validation for Alzheimer’s disease,” *Sci Rep*, vol. 14, no. 1, p. 30770, Dec. 2024, doi: 10.1038/s41598-024-80401-6.

[123] S. Vatansever *et al.*, “Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions,” *Med Res Rev*, vol. 41, no. 3, pp. 1427–1473, May 2021, doi: 10.1002/med.21764.

[124] A. Kamondi, M. Grigg-Damberger, W. Löscher, H. Tanila, and A. A. Horvath, “Epilepsy and epileptiform activity in late-onset Alzheimer disease: clinical and pathophysiological advances, gaps and conundrums,” *Nat Rev Neurol*, vol. 20, no. 3, pp. 162–182, Mar. 2024, doi: 10.1038/s41582-024-00932-4.

[125] H. M. Berman, “The Protein Data Bank,” *Nucleic Acids Res*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.

[126] D. Mendez *et al.*, “ChEMBL: towards direct deposition of bioassay data,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D930–D940, Jan. 2019, doi: 10.1093/nar/gky1075.

[127] C. Knox *et al.*, “DrugBank 6.0: the DrugBank Knowledgebase for 2024,” *Nucleic Acids Res*, vol. 52, no. D1, pp. D1265–D1275, Jan. 2024, doi: 10.1093/nar/gkad976.

[128] S. Kim *et al.*, “PubChem 2025 update,” *Nucleic Acids Res*, vol. 53, no. D1, pp. D1516–D1525, Jan. 2025, doi: 10.1093/nar/gkae1059.

[129] V. Zoete, A. Daina, C. Bovigny, and O. Michielin, “SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening,” *J Chem Inf Model*, vol. 56, no. 8, pp. 1399–1404, Aug. 2016, doi:

10.1021/ACS.JCIM.6B00174/ASSET/IMAGES/LARGE/CI-2016-00174V_0003.JPEG.

[130] T. Barrett *et al.*, "NCBI GEO: archive for functional genomics data sets—update," *Nucleic Acids Res*, vol. 41, no. D1, pp. D991–D995, Nov. 2012, doi: 10.1093/nar/gks1193.

[131] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Res*, vol. 44, no. D1, pp. D457–D462, Jan. 2016, doi: 10.1093/nar/gkv1070.

[132] S. Bienert *et al.*, "The SWISS-MODEL Repository—new features and functionality," *Nucleic Acids Res*, vol. 45, no. D1, pp. D313–D319, Jan. 2017, doi: 10.1093/nar/gkw1132.

[133] A. Daina, O. Michielin, and V. Zoete, "SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Sci Rep*, vol. 7, no. 1, p. 42717, Mar. 2017, doi: 10.1038/srep42717.

[134] "GEOparse · PyPI." Accessed: May 11, 2025. [Online]. Available: https://pypi.org/project/GEOparse/

[135] G. Zhou, O. Soufan, J. Ewald, R. E. W. Hancock, N. Basu, and J. Xia, "NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis," *Nucleic Acids Res*, vol. 47, no. W1, pp. W234–W241, Jul. 2019, doi: 10.1093/nar/gkz240.

[136] C. V., B. W., H. O., and KegelmeyerW. Philip, "SMOTE," *Journal of Artificial Intelligence Research*, Jun. 2002, doi: 10.5555/1622407.1622416.

[137] "PyMOL | pymol.org." Accessed: May 11, 2025. [Online]. Available: https://www.pymol.org/

[138] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox," *J Cheminform*, vol. 3, no. 1, p. 33, Dec. 2011, doi: 10.1186/1758-2946-3-33.

[139] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J Comput Chem*, vol. 31, no. 2, pp. 455–461, Jan. 2010, doi: 10.1002/jcc.21334.

[140] A. C. Wallace, R. A. Laskowski, and J. M. Thornton, "LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions," *"Protein Engineering, Design and Selection,"* vol. 8, no. 2, pp. 127–134, 1995, doi: 10.1093/protein/8.2.127.

[141] F. Pedregosa FABIANPEDREGOSA *et al.*, “Scikit-learn: Machine Learning in Python,” *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Nov. 2011, doi: 10.5555/1953048.2078195.

[142] “StandardScaler — scikit-learn 1.6.1 documentation.” Accessed: May 11, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

[143] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, “A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis,” *Bioinformatics*, vol. 21, no. 5, pp. 631–643, Mar. 2005, doi: 10.1093/bioinformatics/bti033.

LIST OF PUBLICATIONS

1. Conference publication in 2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) entitled "In Silico Identification of Novel GABRA2 Inhibitors Leveraging the Power of Virtual Screening." **DOI:** 10.1109/UPCON62832.2024.10983394.

IEEE Advancing Technology for Humanity

IEEE UP SECTION (INDIA)

11th IEEE UP Section International Conference on
Electrical, Electronics and Computer Engineering (UPCON-2024)
29th November - 01st December 2024

CERTIFICATE
OF PRESENTATION

S. No.164

This is to certify that

Dr./Prof./Mr./Ms./Mrs. **Rishi Mrinal** from **Delhi Technological University** has presented a paper entitled **In silico Identification of Novel GABRA2 Inhibitors Leveraging the Power of Virtual Screening** during UPCON-2024 at Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, U. P., India

General Chair UPCON 2024

Convener UPCON 2024

2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)
Lucknow, India. Nov 29- Dec 01, 2024

2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) | 979-8-3503-7872-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/UPCON62832.2024.10983394

In silico Identification of Novel GABRA2 Inhibitors Leveraging the Power of Virtual Screening

Rishi Mrinal
Molecular Neuroscience and Functional Genomics Laboratory,
Department of Biotechnology,
Delhi Technological University,
Delhi, India
rishimrinal_23bio14@dtu.ac.in

Pravir Kumar
Molecular Neuroscience and Functional Genomics Laboratory,
Department of Biotechnology,
Delhi Technological University,
Delhi, India
pravirkumar@dtu.ac.in

***Abstract*— Anxiety, one of the most prevalent neuropsychiatric disorders, often remains underdiagnosed due to its subtle and varied symptoms. Among the key players in anxiety regulation, the GABRA2 receptor subunit of the GABA receptor system stands out for its critical role in maintaining neuronal inhibition and promoting calmness. Despite its significance, the GABRA2 receptor has been largely unexplored in drug discovery. Current therapeutics, including Benzodiazepines like Diazepam, target GABA receptors but are associated with undesirable side effects such as sedation and cognitive dysfunction, underscoring the need for safer alternatives. This study presents a pioneering effort to explore the GABRA2 receptor using advanced computational techniques, including virtual screening and molecular docking. By leveraging a novel and comprehensive in silico approach, we identified promising drug candidates as potential alternatives to current benzodiazepine-based therapies. Our findings not only emphasize the untapped therapeutic potential of targeting GABRA2 in anxiety management but also validate the power of computational methodologies in accelerating drug discovery.**

Keywords*— *Anxiety, GABRA2, Structure prediction, Diazepam, Molecular docking, Virtual Screening, Computational drug discovery

I. Introduction

Anxiety disorders are the most common mental disorders. The life-time prevalence rate of anxiety disorder for adolescents aged between 13-17 is 7.7%, while it is 6.6% in adults aged between 18-64 years [1]. Women are more prevalent to develop anxiety disorders and approximately twice as high as in male. Anxiety can be identified in patients by the presence of continuous fear, unrealistic worry about something to happen, and/or even unpleasant feelings of imminent death [2]. The prevalence of anxiety disorders is as follows: 10.3% for specific phobias, 6.0% for panic disorder, 2.7% for social anxiety disorder (phobia), and 2.2% for generalized anxiety disorder [3]. Anxiety disorders often accompanied by neuromuscular tension, restlessness, fatigue, and concentration deficit, resulting in significantly interfere with daily activities. Moreover, scientific evidence suggests that prolonged anxiety result in the development of more serious and detrimental health consequence, which lowered overall life expectancy of the individual.

Anxiolytics or antidepressants are commonly used medical intervention to treat or manage anxiety disorders. Among others, anxiolytic drugs Benzodiazepines are the classical class of drug compounds extensively used in the treatment of anxiety. Additionally, benzodiazepines are also useful in the treatment of many neurological and psychiatric disorders like insomnia, muscle relaxation, and epilepsy. Benzodiazepines function as positive allosteric modulators of gamma-aminobutyric acid (GABA) receptors, particularly the GABA-α2 receptor, also known as the GABRA2 receptor. By promoting increased binding of the inhibitory neurotransmitter GABA, benzodiazepines prevent aberrant neuronal activity in the central nervous system [4]. However, in addition to their positive therapeutic efficacy long-term benzodiazepines consumption may show side effects such as, anterograde amnesia, emotional blunting, and marked neonatal withdrawal symptoms when taken during pregnancy [5]. The front-line drugs of the class of benzodiazepines, used for the treatment are Diazepam, Oxazepam, Lorazepam, Alprazolam and Chlordiazepoxide, out of which the anxiolytic and pharmacological mechanisms of some of these drugs are still unclear. To understand these interactions, it is essential for the development of more targeted and effective therapeutic agents with potentially minimized adverse effects.

GABRA2 encodes alpha-2 subunit of the GABA-A receptor, which functions as a ligand-gated ion channel is a promising therapeutic target for the central nervous system that inhibits neurotransmitters. Diazepam, a widely prescribed benzodiazepine, works by modulating the GABAergic system especially GABRA2 where the therapeutic actions mediate through its binding to the binding site of the receptor. The GABRA2 is unexplored largely due to unavailability of experimentally validated 3-D structure. Through this paper, we aim to predict the 3D structure of the GABRA2 protein using homology modelling and further investigate the binding interactions between various compounds and GABRA2 receptor using Molecular docking to identify a lead compound that will potentially inhibit GABRA2 with a great efficiency.

II. Review of Literature

Anxiety triggers persistent feeling of unrealistic fear or danger in individuals. Although the specific cause of anxiety disorders is unknown, genetic and environmental factors are thought to have a significant impact. However, it is known that chronic illness, for instance cancer, lack of sleep, eating disorders can trigger anxiety [6]. The GABRA2 is associated in regulating the major inhibition of neurotransmitter impulses. Although, according to some researchers Serotonin transporter gene (SLC6A4) is also associated with major inhibitory neuro-transmission in adult brain, leading to anxiety [7]. The genetic variations influence neuro-transmission regulation, alterations in neural circuits, stress-mediated pathways and many.

2. Accepted Conference in 2025 IEEE Second International Conference On Advances in Modern Age Technologies for Health and Engineering Sciences entitled “Integrating Machine Learning and Gene Expression Analysis for Dementia Subtype Classification and Biomarker Discovery.”

Second International Conference on

Advances in Modern Age Technologies for Health and Engineering Science (AMATHE 2025)

24 - 25, April 2025 | Shivamogga, Karnataka, India

CERTIFICATE

This certificate is presented to

AE-CS 9001

Rishi Mrinal
Molecular Neuroscience and Functional Genomics Laboratory,
Department of Biotechnology,
Delhi Technological University, Delhi, India

for presenting the research paper entitled “Integrating Machine Learning and Gene Expression Analysis for Dementia Subtype Classification and Biomarker Discovery” in the Second International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE 2025) held at PES Institute of Technology and Management, Shivamogga, Karnataka, India during 24 - 25, April 2025. The conference is technically co-sponsored by IEEE Bangalore Section.

Dr. Likewin Thomas
Conference Chair

Dr. Prasanna Kumar H R
Organizing Chair

Organized by

PESITM | PES INSTITUTE OF TECHNOLOGY & MANAGEMENT
Shivamogga, Karnataka, India | www.pestrust.edu.in/pesitm

Technical Sponsors

IEEE Bangalore Section | IEEE

Integrating Machine Learning and Gene Expression Analysis for Dementia Subtype Classification and Biomarker Discovery

Rishi Mrinal
Molecular Neuroscience and Functional Genomics Laboratory,
Department of Biotechnology,
Delhi Technological University, Delhi, India
rishimrinal_23bio14@dtu.ac.in

Pravir Kumar
Molecular Neuroscience and Functional Genomics Laboratory,
Department of Biotechnology,
Delhi Technological University, Delhi, India
pravirkumar@dtu.ac.in

***Abstract*— Dementia, a progressive neurodegenerative disorder affecting millions worldwide, poses significant challenges in early diagnosis and targeted treatment due to overlapping clinical symptoms. While gene expression analysis has provided molecular insights into dementia subtypes, the integration of machine learning (ML) remains underexplored. This study utilizes gene expression data from the GEO database and ML techniques to enhance dementia subtype classification and identify potential biomarkers. Rigorous preprocessing, including quantile normalization and SMOTE, ensured data integrity and minimized bias. Among all the models evaluated, the Support Vector Classifier (SVC) achieved highest accuracy (99%), while the Random Forest (RF) classifier proved valuable for feature selection. Key biomarkers, including MOBP, PDIA5, and RAB43, were identified, with pathway enrichment analysis highlighting subtype-specific molecular mechanisms such as axon guidance and disrupted protein processing. These findings demonstrate the potential of ML-driven approaches in dementia research, offering improved diagnostic accuracy and biomarker discovery. Future studies should focus on validating these biomarkers in independent cohorts and integrating multi-omics data to advance precision medicine for neurodegenerative disorders.**

Keywords*— *Dementia subtypes, Machine learning, Gene expression analysis, Biomarker identification, Precision medicine, Neurodegenerative disorders

I. Introduction

Dementia is considered to be one of the broad category of human brain disorders, a life threatening disease that leads to a complex spectrum of progressive neurodegenerative conditions, each characterized by clinical features and neuropathological markers [1]. Globally over 50 million people have been affected by dementia and most of them are above the age gap of 60. Numerous studies have reported that the proper diagnosis and effective treatment of this disorder remains challenging [2].

The subtypes of dementia including, behavioural variant Frontotemporal Dementia (bvFTD), Progressive Supranuclear Palsy (PSP), Corticobasal Syndrome (CBS), and Primary Progressive Aphasia (PPA), poses significant challenges to precise diagnosis and the design of effective therapeutic strategies, largely due to overlapping symptoms and variability in disease progression. The identification of reliable biomarkers through gene expression analysis offers a promising avenue for advancing the understanding and differentiation of these subtypes [3]. However, discovering reliable biomarkers and identifying differentially expressed genes (DEGs) from a variety of studies may help to improve target selection, enabling a more personalized approach to the treatment.

Biomarkers are the measurable indicators of biological processes aiding in early diagnosis, risk assessment, and treatment monitoring. Traditionally, dementia was only diagnosed at the clinically symptomatic stage by neuropsychological examination alone, biomarkers offer a potential bridge (gap) between symptomatology and pathophysiology, providing insights into the neurobiological frameworks underlying the disorders. The biomarker identification can lead to the early detection of the disorder before the actual symptoms are clear. Gene expression analysis, is a term which involves the exploration of gene activity patterns at the transcriptional level submitting a detailed information of biological and cellular processes. The different methodologies such as microarrays, bulk RNA sequencing (RNA-Seq), and single-cell RNA sequencing (scRNA-Seq) allow researchers in examining different gene expression patterns across different cell conditions. By computational methods, such as network analysis, co-expression studies and differential expression analysis, we can identify genes and pathways associated with the disorder. The data for the analysis study is obtained from Gene Expression Omnibus (GEO) datasets, that is a repository which archives high-throughput gene expression datasets submitted by the scientific community globally.

Artificial Intelligence comprises of various Machine Learning (ML) techniques that enables analysis of complex datasets related to dementia and its subtypes for early disease detection, biomarker discovery as well as prognosis of the disease. The present study focuses on determining functionally related genes that integrates gene expression

3. Poster Presentation in SNCI (Society for Neurochemistry, India) 2025, Delhi Chapter at Jamia Hamdard University on “Computational Identification of GABRA2 Inhibitors for Targeted Modulation of Learning and Memory.”

Society for Neurochemistry, India (SNCI) Delhi Local Chapter
&
Department of Toxicology, School of Chemical and Life Sciences
Jamia Hamdard, New Delhi

Certificate of Appreciation

This is to certify that Prof./Dr./Ms./Mr. Ms. Rishi Mrinal
of Delhi Technological University has
Participate as **Delegate** in the Two Days National Symposium on **“Neurochemistry and Emerging Therapeutics: Challenges and Opportunities in Neuroscience”** held from 16th April 2025 to 17th April 2025 at Convention Centre, Jamia Hamdard, New Delhi. He/She has also presented in Young Investigator/Poster Session.

Prof. Mohammad Akram
Organizing Co-Chairperson

Prof. Suhel Parvez
Organizing Chairperson

Prof. Prakash Babu Phanithi
Secretary General (HQ), SNCI

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

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

PLAGIARISM VERIFICATION

Title of the Thesis: **Integrative Computational Approaches for Receptor-based Drug Discovery and Biomarker Identification in NDDs**

Total Pages: **58**

Name of the Student: **Rishi Mrinal**

Supervisor: **Prof. Pravir Kumar**

Department of Biotechnology, Delhi Technological University, Delhi - 110042

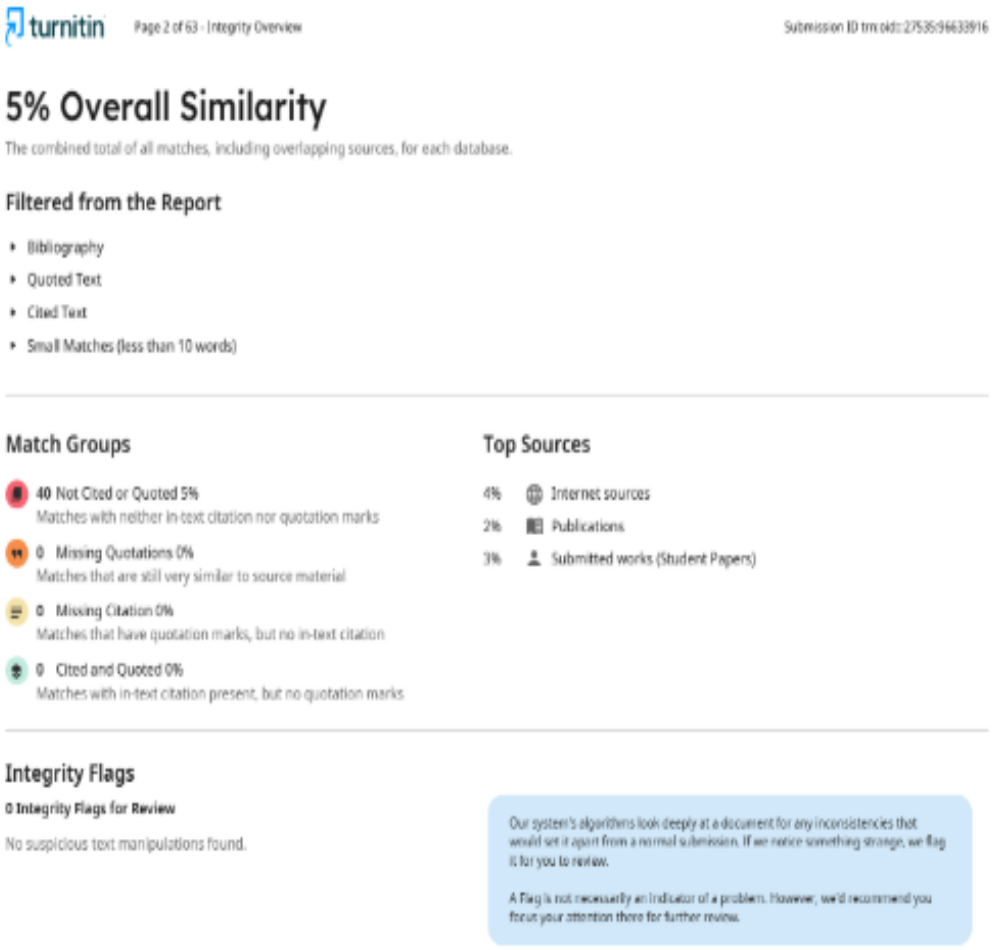
This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: **Turnitin,** Similarity Index: **5%,** Total Word Count: **16,815**

Date: 19.05.2025

Candidate's Signature

Signature of Supervisor



turnitin Page 2 of 63 - Integrity Overview Submission ID trn:oid:::2753:96633916

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Match Groups

- 40 Not Cited or Quoted 5%
  Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

4% Internet sources
2% Publications
3% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.