AUCKLAND UNIVERSITY OF TECHNOLOGY

DOCTORAL THESIS

# Multimodal, Personalized and Explainable AI in Mental Health: Early Diagnosis and Prognosis using Longitudinal Data

*Author:*
Sugam BUDHRAJA

*Supervisors:*
Dr. Maryam DOBORJEH
Asst Prof. Wilson GOH
Prof. Nikola KASABOV

*A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*

*in the*

Knowledge Engineering and Discovery Research Innovation (KEDRI)
School of Engineering, Computer and Mathematical Sciences

May 19, 2025

# Declaration of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor used artificial intelligence tools or generative artificial intelligence tools (unless it is clearly stated, and referenced, along with the purpose of use), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signed:

Date: August 4, 2024

*"What lies before us and what lies behind us are small matters compared to what lies within us. And when we bring what is within out into the world, miracles happen."*

Henry D. Thoreau

AUCKLAND UNIVERSITY OF TECHNOLOGY

# *Abstract*

Design and Creative Technologies
School of Engineering, Computer and Mathematical Sciences

Doctor of Philosophy

**Multimodal, Personalized and Explainable AI in Mental Health: Early Diagnosis and Prognosis using Longitudinal Data**

by Sugam BUDHRAJA

The early diagnosis and prognosis of mental health disorders remain formidable challenges, often complicated by their multifaceted origins and evolving presentation. This thesis tackles these challenges by developing explainable artificial intelligence (AI) models that combine multimodal and longitudinal data—spanning clinical, cognitive, genetic, and social domains—to improve the accuracy, transparency, and real-world usability of mental health predictions.

The work introduces several novel methods, including the Filter and Wrapper Stacking Ensemble (FWSE) for robust biomarker discovery in high-dimensional gene expression data, and the Dynamic Attention Gateway (DAG) for Liquid State Machines, which brings both interpretability and temporal sensitivity to time-series analysis. In integrating diverse data sources, the Mosaic Liquid State Machine (Mosaic LSM) architecture demonstrates that fusing clinical, neurocognitive, genetic, and social data can yield predictive models that outperform single-modality approaches in identifying individuals at ultra-high risk for mental illness.

Empirical results across real-world datasets—including the longitudinal LYRIKS study—show that these models not only match or exceed the accuracy of standard machine learning approaches but also offer clinicians transparent, case-by-case explanations for their predictions. For example, using pathway-level gene aggregation and attention-based mechanisms, prediction accuracies above 95% were achieved for certain high-risk mental health classifications, while simultaneously surfacing biologically meaningful markers that align with established literature.

To promote adoption beyond the data science community, this research also delivers NeuroGeMS, an open-source GUI software that makes advanced multimodal AI accessible for researchers in the biomedical domain. Taken together, these contributions lay groundwork for personalized and explainable AI in mental health.

# *Acknowledgements*

I would like to express my deepest gratitude to all those who have supported and guided me throughout my PhD journey. First and foremost, I would like to thank my supervisor, Dr Maryam Doborjeh, for her invaluable guidance, constant encouragement, and insightful feedback.

I am also immensely grateful to my co-supervisor, Asst Prof. Wilson Goh, for his continuous support, constructive criticism, advice on biological aspects of the research, and for always being available for discussions and advice. His contributions have greatly enriched this thesis.

I would also like to thank my mentor and supervisor Prof. Nikola Kasabov. His expertise, dedication, and unwavering support have been instrumental in shaping this research.

My heartfelt thanks go to my friends at the Auckland University of Technology, Nanayang Technological University, and Auckland Bioengineering Institute, especially Samuel, Gurleen, Balkaran, and Sherry. Their camaraderie, collaboration, and intellectual discussions have made this journey enjoyable and stimulating.

I also extend my thanks to my colleagues in this joint project, Dr. Zohreh Doborjeh, Prof. Edmund Lai, and Dr. Margaret Williams, for their invaluable support and collaboration.

I would like to extend my gratitude to the funding agencies, Ministry of Business, Innovation and Employment (MBIE) and Singapore Data Science Consortium (SDSC), for providing the financial support necessary for this research.

To my family, especially my parents, Sumeet and Rachna, my sister Anshul, and my relatives in New Zealand, Dheeraj and Vandana, for their unwavering support, patience, and encouragement throughout this journey. Their belief in me has been a source of strength and motivation.

Additionally, I acknowledge the use of generative AI tools in enhancing the wording of the introductions and conclusions in this thesis, for making complex terminology more accessible to a broader audience.

This thesis would not have been possible without the collective support and encouragement of all mentioned individuals. To each and every one of you, thank you.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **AI** | **A**rtificial **I**ntelligence |
| **ANOVA** | **AN**alysis **O**f **VA**riance |
| **AR** | **A**uto**r**egressive |
| **ARIMA** | **A**uto**r**egressive **I**ntegrated **M**oving **A**verage |
| **BACS** | **B**rief **A**ssessment of **C**ognition in **S**chizophrenia |
| **BAI** | **B**eck **A**nxiety **I**nventory |
| **CAARMS** | **C**omprehensive **A**ssessment (of) **A**t **R**isk **M**ental **S**tates |
| **CDSS** | **C**algary **D**epression **S**cale for **S**chizophrenia |
| **CPT** | **C**ontinuous **P**erformance **T**est |
| **CPTAC** | **C**linical **P**roteomic **T**umor **A**nalysis **C**onsortium |
| **DAG** | **D**ynamic **A**ttention **G**ateway |
| **DL** | **D**eep **L**earning |
| **DSM** | **D**iagnostic (and) **S**tatistical **M**anual (of Mental Disorders) |
| **EEG** | **E**lectro**e**ncephalo**g**raphy |
| **ER** | **E**ndoplasmic **R**eticulum |
| **fMRI** | **f**unctional **M**agnetic **R**esonance **I**maging |
| **FWSE** | **F**ilter (and) **W**rapper **S**tacking **E**nsemble |
| **GO** | **G**ene **O**ntology |
| **KNN** | **K**-**N**earest **N**eighbors |
| **LogReg** | **Log**istic **Reg**ression |
| **LSM** | **L**iquid **S**tate **M**achine |
| **LSTM** | **L**ong **S**hort-**T**erm **M**emory |
| **LUAD** | **L**ung **A**denocarcinoma |
| **LYRIKS** | **L**ongitudinal **Y**outh at **R**isk **S**tudy |
| **MA** | **M**oving **A**verage |
| **MCC** | **M**atthew's **C**orrelation **C**oefficient |
| **MCF-RFE** | **M**ulti-**C**riterion **F**usion-based **R**ecursive **F**eature **E**limination |
| **ML** | **M**achine **L**earning |
| **MLP** | **M**ulti**l**ayer **P**erceptron |
| **MMSE** | **M**ini-**M**ental **S**tate **E**xamination |
| **MoCA** | **M**ontreal **C**ognitive **A**ssessment |
| **MRI** | **M**agnetic **R**esonance **I**maging |
| **NGS** | **N**ext-**G**eneration **S**equencing |
| **NSCLC** | **N**on-**S**mall **C**ell **L**ung **C**ancer |
| **PANSS** | **P**ositive (and) **N**egative **S**yndrome **S**cale |
| **PDAC** | **P**ancreatic **D**uctal **A**denocarcinoma |
| **PerClose** | **P**erceptual **C**losure **T**ask |
| **PET** | **P**ositron **E**mission **T**omography |
| **RF** | **R**andom **F**orest |
| **RFE** | **R**ecursive **F**eature **E**limination |
| **RNN** | **R**ecurrent **N**eural **N**etwork |
| **SCID** | **S**tructured **C**linical **I**nterview for **DSM** |

| | |
|---|---|
| **SFS** | **S**ocial **F**unctioning **S**cale |
| **SnK** | **S**nakes i**n** (the) **G**rass |
| **SNR** | **S**ignal-to-**N**oise **R**atio |
| **SVM** | **S**upport **V**ector **M**achine |
| **TCGA** | **T**he **C**ancer **G**enome **A**tlas |
| **TWNFI** | **T**ransductive **W**eighted **N**euro-**F**uzzy **I**nference |
| **UHR** | **U**ltra-**H**igh **R**isk |
| **VSOLassoBag** | **V**ariable-**S**election **O**riented **L**asso **B**agging |
| **WWKNN** | **W**eighted **W**eighted **K**-**N**earest **N**eighbors |

*Dedicated to humankind, with the earnest hope that our collective efforts and understanding will continue to advance the field of mental health*

# Chapter 1

# Introduction

## 1.1 Background

Mental health disorders represent a significant public health challenge, affecting millions of individuals globally and imposing substantial socio-economic burdens. According to the World Health Organization (WHO), depression is the leading cause of disability worldwide, while anxiety disorders rank sixth (Organization et al., 2017). The complexity of mental health disorders, characterized by multifaceted symptoms and varying progression patterns, underscores the need for early diagnosis and effective prognosis to improve treatment outcomes and quality of life for affected individuals.

The early diagnosis and prognosis of mental health disorders are critical for several reasons. First, timely identification of these disorders can lead to early interventions, which are often more effective and less costly than treatments administered at later stages (Insel, 2014). Second, accurate prognosis can guide personalized treatment plans, improving the likelihood of recovery and reducing the risk of relapse (Kessler et al., 2005). Despite these benefits, current diagnostic methods primarily rely on self-reported symptoms and clinical interviews, which can be subjective and prone to biases (Kupfer, First, and Regier, 2008).

Advancements in artificial intelligence (AI) and machine learning (ML) offer promising avenues to enhance the early diagnosis and prognosis of mental health disorders. AI techniques, particularly those involving multimodal data integration, can analyze vast and diverse datasets, uncovering patterns that may not be apparent through traditional methods (Bzdok and Meyer-Lindenberg, 2018). These advancements can facilitate a more holistic understanding of mental health, considering genetic, neurobiological, cognitive, and behavioral factors.

Multimodal approaches integrate data from various sources, such as clinical assessments, neuroimaging, genetic information, and digital phenotyping from smartphones and wearables (Cohen et al., 2020). This comprehensive perspective can lead to more accurate and personalized models for predicting mental health outcomes. For instance, studies have shown that combining neuroimaging data with genetic and clinical information can improve the prediction of schizophrenia onset (Koutsouleris et al., 2021).

The use of longitudinal data is particularly valuable in mental health research, as it allows for the tracking of changes in individuals over time (Achenbach and Rescorla, 2007). Longitudinal studies can identify early markers of mental health deterioration and provide insights into the temporal dynamics of disorders, enabling the development of models that predict future outcomes based on historical data (Green et al., 2010).

However, the integration of multimodal and longitudinal data presents significant challenges. These include the need for sophisticated data processing techniques, the management of high-dimensional datasets, and the requirement for models that can explain their predictions in a clinically meaningful way (Doshi-Velez and Kim, 2017). Addressing these challenges is crucial for translating AI research into practical tools for mental health diagnosis and prognosis.

## 1.2 Motivation

Mental health disorders are a significant global health concern, affecting nearly one billion people worldwide as of 2024, according to the World Health Organization (WHO) (Organization, 2022). This represents a substantial portion of the global population, with about one in eight individuals affected by mental health conditions at any given time. Furthermore, it is estimated that 50% of people will develop at least one mental health disorder in their lifetime (McGrath et al., 2023).

The COVID-19 pandemic has exacerbated mental health issues globally. The uncertainty, isolation, and economic instability caused by the pandemic have led to a significant increase in the prevalence of mental health disorders. A survey conducted by the Centers for Disease Control and Prevention (CDC) found that symptoms of anxiety disorder and depressive disorder increased considerably in the United States during the pandemic, with 31% of respondents reporting anxiety or depression, 13% reporting starting or increasing substance use, and 11% reporting serious consideration of suicide (Czeisler, 2020).

Children and adolescents have been particularly affected by the pandemic. School closures and social distancing measures have disrupted their routines and social interactions, leading to increased levels of stress, anxiety, and depression. A study published in *JAMA Pediatrics* reported that the prevalence of depressive and anxiety symptoms among children and adolescents has doubled during the COVID-19 pandemic (Racine et al., 2021).

The economic impact of mental health disorders is substantial, with global costs projected to reach $6 trillion by 2030 (Trautmann, Rehm, and Wittchen, 2016). These costs are driven by direct medical expenses, loss of productivity, and the economic burden placed on families and caregivers. In the workplace, mental health issues can lead to absenteeism, reduced performance, and increased turnover, further amplifying their economic impact (Jansson and Gunnarsson, 2018).

These alarming statistics underscore the urgent need for improved methods of diagnosing and prognosing mental health disorders. Traditional diagnostic methods, which rely heavily on self-reported symptoms and clinical interviews, can be subjective and inconsistent (Kupfer, First, and Regier, 2008). This subjectivity often results in delayed diagnoses and inadequate treatment, exacerbating the burden of mental health disorders.

## 1.3 Aim

The overarching aim of this thesis is to develop and validate multimodal, personalized, and explainable artificial intelligence (AI) models for the early diagnosis and prognosis of mental health disorders using longitudinal data. This objective is pursued through several specific aims, each addressing key aspects of the research problem.

### 1.3.1 Objectives

1. **To identify robust early markers of mental health**: Develop methods to discover and validate reproducible early biomarkers associated with mental health disorders, focusing on gene expression and other relevant data.

2. **To enhance predictive accuracy using longitudinal data**: Utilize longitudinal data to track changes in individuals over time, improving the accuracy of diagnostic and prognostic models. Identifying temporal patterns of mental health deterioration.

3. **To make longitudinal AI models explainable**: Create AI models that provide meaningful explanations for their predictions. This objective is crucial for ensuring that the models can be trusted and used effectively.

4. **To integrate multimodal data sources**: Develop methods to effectively integrate various types of data, including clinical assessments, genetic information, cognitive performance, and social functioning. This integration aims to create a comprehensive dataset that captures the multifaceted nature of mental health disorders.

5. **To facilitate the adoption of multimodal AI methods and advance research in the biomedical domain**: Develop user-friendly tools and frameworks that make it easier for researchers in medical field to adopt and utilize multimodal AI methods.

### 1.3.2 Research Questions

The specific research questions addressed in this thesis include:

1. **What are the robust early biomarkers of mental health disorders, and how can they be identified and validated?**

2. **What are the key temporal patterns that can be identified using longitudinal data for predicting mental health outcomes?**

3. **How can AI models for time-series data be designed to provide explanations for their predictions?**

4. **How can multimodal data sources be effectively integrated to improve the diagnosis and prognosis of mental health disorders?**

5. **How can user-friendly software be developed to facilitate the use of multimodal AI models and promote advancements in biomedical research?**

## 1.4 Thesis Structure

This thesis contains six more chapters, each addressing specific aspects of the research objectives and answering the research questions outlined earlier. The structure of the thesis is as follows:

### Chapter 1: Introduction

The introduction provides an overview of the research problem, its significance, and the objectives and research questions that guide this study. It also outlines the motivation for the research, particularly in the context of the increasing prevalence of mental health disorders and the potential of AI to improve diagnosis and prognosis.

### Chapter 2: Review of Diagnosis and Prognosis in Mental Health

This chapter provides a comprehensive review of current methodologies in mental health diagnosis and prognosis, highlighting the limitations and challenges of existing approaches. It sets the stage for the development of new methods by identifying gaps in the current state of the art and justifying the need for advanced AI models.

### Chapter 3: Biomarker Discovery on Gene Data

This chapter introduces a novel method for biomarker discovery, focusing on the analysis of gene expression data. The chapter details the development and validation of the Filter and Wrapper Stacking Ensemble (FWSE) approach. This method is essential for identifying genetic markers associated with mental health disorders, which can enhance the understanding and prediction of these conditions. By identifying these biomarkers, the chapter addresses the need for better diagnostic tools and provides a foundation for more accurate AI models.

### Chapter 4: Explainable Modelling of Longitudinal Data

This chapter explores methods for developing explainable AI models that can analyze longitudinal data. Techniques such as the Dynamic Attention Gateway (DAG) for Liquid State Machines (LSMs) are presented. These methods ensure that the models provide interpretable and clinically meaningful explanations for their predictions, which is crucial for their adoption in clinical settings. By focusing on explainability, this chapter addresses the need for trust and transparency in AI models, making them more useful for interdisciplinary research teams.

### Chapter 5: Multimodal Longitudinal Data Integration

This chapter discusses the integration of diverse data sources using the Mosaic LSM method. It provides empirical analyses demonstrating the efficacy of multimodal data integration in enhancing predictive accuracy for mental health diagnostics. This integration is key to capturing the multifaceted nature of mental health disorders. By effectively combining various data types, the chapter advances the objective of creating comprehensive and robust AI models that reflect the complexity of mental health conditions.

### Chapter 6: Development of GUI Software for Multimodal Modelling

This chapter details the creation of NeuroGeMS, a graphical user interface software designed to facilitate easy modeling and analysis of multimodal data. The chapter showcases the software's application in real-world scenarios, emphasizing its utility for researchers of various backgrounds. By making advanced AI tools accessible, this chapter addresses the objective of facilitating the adoption of multimodal AI methods and advancing research in the biomedical domain.

### Chapter 7: Conclusion and Future Work

The conclusion summarizes the main findings of the thesis, discusses their implications, and outlines potential directions for future research. It reflects on how the research objectives were met and the research questions answered, highlighting the contributions to the field of mental health diagnostics and AI.

### Appendices

The appendices provide supplementary material that supports the main text of the thesis. This includes detailed descriptions of algorithms, additional data analyses, extended results, and other relevant information that adds depth to the research. The appendices ensure that all technical details and supporting information are available for reference.

## 1.5 Significance

### 1.5.1 Contribution

The research conducted in this thesis makes the following key contributions to the field of mental health diagnostics and artificial intelligence:

- **Development of Multimodal Integration Techniques**: The thesis introduces novel methods for integrating diverse data sources, such as clinical assessments, neuroimaging, genetic data, and digital phenotyping. These techniques enhance the ability to capture the multifaceted nature of mental health disorders and improve diagnostic accuracy.
- **Enhancement of Predictive Models Using Longitudinal Data**: By utilizing longitudinal data, this research develops methods that can identifies temporal patterns and early markers of mental health deterioration, leading to more accurate and timely predictions.
- **Explainable AI Model for Longitudinal Data**: The development of explainable AI model for time-series data ensures that the predictions are interpretable, facilitating their adoption and increasing trust among healthcare professionals.
- **User-Friendly Tools for Researchers**: The creation of NeuroGeMS, a graphical user interface software, makes advanced AI tools accessible to medical researchers, facilitating the adoption of multimodal AI methods and advancing research in the biomedical domain.

### 1.5.2 Publications

The following peer-reviewed publications have emerged from the research conducted in this thesis:

1. Filter and wrapper stacking ensemble (FWSE): a robust approach for reliable biomarker discovery in high-dimensional omics data (Budhraja et al., 2023a).
2. Mosaic LSM: A liquid state machine approach for multimodal longitudinal data analysis (Budhraja et al., 2023b).

3. NeuroGeMS: An open-source GUI software for multimodal modelling in biomedical research and applications (Budhraja et al., 2024).

Related works published as part of collaborative efforts under the NZ–SG Data Science Catalyst Project:

1. Constrained neuro fuzzy inference methodology for explainable personalised modelling with applications on gene expression data (Singh et al., 2023).
2. Investigation of social and cognitive predictors in non-transition ultra-high-risk'individuals for psychosis using spiking neural networks (Doborjeh et al., 2023).
3. RNA-sequencing of peripheral whole blood of individuals at ultra-high-risk for psychosis–A longitudinal perspective (Tan et al., 2023).
4. A generalisability theory approach to quantifying changes in psychopathology among ultra-high-risk individuals for psychosis (Doborjeh et al., 2024).

### 1.5.3 Presentations

The research findings have been presented at various conferences and seminars, contributing to the dissemination and discussion of the work within the scientific community:

1. "Personalized Modelling for Medical Applications," AI Symposium, *National Cancer Centre Singapore (NCCS)*, 2021.
2. "FWSE: A novel ensemble feature selection method for biomarker discovery," PG Research Symposium, *Auckland University of Technology*, 2022.
3. "Predicting future status of Ultra High-Risk patients on multimodal longitudinal data," AIRA Conference, *University of Canterbury*, 2022.
4. "Mosaic LSM: A Liquid State Machine Approach for Multimodal Longitudinal Data Analysis," IJCNN, *Gold Coast*, 2023.
5. "Explainable AI and Multimodal Modelling for Diagnosis and Prognosis in Mental Health," iAIM, *NTU Singapore*, 2023.
6. "NeuroGeMS: Open-source GUI Software for Multimodal Modelling," ICONIP, *Auckland University of Technology*, 2024

# Chapter 2

# Review of Diagnosis and Prognosis in Mental Health

## 2.1 Introduction

Mental health disorders represent a significant global health challenge, affecting millions of individuals worldwide and imposing substantial burdens on healthcare systems, economies, and societies. The accurate and timely diagnosis and prognosis of these disorders are crucial for effective intervention and treatment. This chapter provides a comprehensive review of current approaches, challenges, and emerging technologies in the field of mental health diagnostics and prognostics.

We begin by exploring the critical importance of early detection in mental health disorders, highlighting how timely identification can significantly improve patient outcomes and reduce the overall burden of disease. The chapter then dives into the evolving landscape of diagnostic methodologies, contrasting traditional unimodal approaches with more recent multimodal strategies that promise a more holistic understanding of mental health conditions.

A key focus of this review is the concept of Ultra High Risk (UHR) in mental health, particularly in the context of psychosis. We examine the characteristics and significance of UHR states, their role in early intervention strategies, and the challenges in accurately identifying individuals at risk. To illustrate the practical application of multimodal approaches in UHR research, we introduce the Longitudinal Youth-At-Risk Study (LYRIKS) dataset, a rich source of multimodal data on individuals at risk of developing psychosis.

The chapter also emphasizes the power of longitudinal data in mental health research and clinical practice. We explore various analytical techniques for longitudinal data, ranging from traditional statistical methods to advanced machine learning approaches, including deep learning and neuromorphic computing models like Liquid State Machines.

By critically analyzing the existing literature and identifying gaps in current methodologies, this chapter aims to lay the groundwork for the development of advanced AI models that can enhance the diagnosis and prognosis of mental health disorders. The insights gained from this review will inform the subsequent chapters, where novel approaches and methods will be introduced and validated.

## 2.2 Mental Health Diagnostics

### 2.2.1 The Importance of Early Detection

The landscape of mental health disorders is extensive, encompassing a range of conditions that can profoundly impact an individual's quality of life. Early diagnosis

and prognosis of such disorders are vital steps, influencing the direction of treatment and overall patient outcomes. This section sheds light on the significance of early detection and intervention in mental health, a topic of pivotal importance in modern healthcare research and practice.

**Early Diagnosis**

Diagnosis refers to the process of identifying a disease or condition based on its signs, symptoms, and patient history. In the context of mental health, this often involves the recognition of patterns in behavior, mood, and cognitive function that align with established psychiatric disorders. The importance of early diagnosis of mental health disorders is discussed below.

1. **Reducing the Burden of Disease**: Mental health conditions can severely affect an individual's personal, social, and professional life if not identified early. Timely diagnosis can alleviate these effects, lessening the burden on the individual, their network, and healthcare systems.

2. **Prompt Intervention**: The early detection of mental health issues is crucial for commencing appropriate treatment promptly, often leading to improved outcomes. It can arrest the disorder's progression, lessen symptom severity, and enhance the patient's recovery prospects.

**Prognosis**

Prognosis involves predicting the future course and outcome of a diagnosed condition. It's about estimating the likely progression of the disorder, including its potential impacts on the patient's functional abilities and quality of life. The importance of prognosis in mental health is discussed below.

1. **Individualized Treatment Planning**: An accurate prognosis is essential for tailoring treatment to the individual's specific needs. It helps in predicting the course of the disorder, guiding clinicians in selecting the most effective treatment strategies and adjusting them over time.

2. **Predicting Treatment Response**: Understanding the likely progression of a mental health disorder assists in setting realistic expectations for treatment response and outcomes. This knowledge is vital for both clinicians and patients in managing the disorder effectively.

3. **Prevention of Relapse**: In mental health, the risk of relapse is a significant concern. Prognostic evaluation helps in identifying individuals at higher risk, enabling the implementation of strategies focused on relapse prevention.

Advancements in technology, particularly in the field of AI, have opened new avenues in the early diagnosis and prognosis of mental health disorders. The use of machine learning models in analyzing patterns within large datasets, including genetic, neuroimaging, and behavioral data, has shown promising results in predicting the onset and trajectory of mental health conditions.

### 2.2.2 Unimodal vs Multimodal Approaches

Unimodal approaches which focus on single modality or perspective for diagnosis and treatment, have been prevalent in the field for many years. While these methods have their merits, they also come with significant limitations, especially in the complex and multifaceted world of mental health disorders. Multimodal approaches, on the other hand, integrate diverse data types, offering a more holistic understanding of mental health disorders as discussed in section 5.2.2. This section outlines key differences between unimodal and multimodal approaches.

1. **Comprehensiveness in Assessment**
    - **Unimodal**: Unimodal diagnostics often fail to capture the multifaceted nature of mental health disorders, as they typically focus on a single dimension (e.g., psychological symptoms alone).
    - **Multimodal**: Multimodal diagnostics incorporate a variety of data sources, such as psychological assessments, physiological data, and social-behavioral patterns. This multidimensional approach results in a more thorough and nuanced understanding of the disorder.
2. **Resilience to Biased Interpretations**
    - **Unimodal**: Diagnoses based on a single modality can be prone to biases inherent in that specific method, leading to skewed or partial interpretations.
    - **Multimodal**: By integrating data from multiple sources, multimodal approaches minimize the risk of biases tied to any one modality. This results in a more balanced and objective assessment.
3. **Adaptability to Patient Diversity**
    - **Unimodal**: Unimodal methods may not adequately account for the diversity in patient backgrounds and experiences, potentially leading to less effective or inappropriate treatment strategies for certain groups.
    - **Multimodal**: Multimodal diagnostics can be more adaptable and sensitive to variations in cultural, genetic, and individual factors. This enhances the capacity to provide culturally competent and individualized care.
4. **Detecting Comorbid Conditions**
    - **Unimodal**: Focusing on a single aspect of mental health might overlook comorbid conditions, which are common in psychiatric disorders.
    - **Multimodal**: A multimodal approach is better equipped to identify comorbidities by analyzing a broader spectrum of data. This leads to a more comprehensive treatment plan that addresses all aspects of a patient's mental health.

In conclusion, the transition from unimodal to multimodal approaches in mental health diagnostics represents a paradigm shift towards more comprehensive, accurate, and personalized care. This shift aligns with the evolving understanding of mental health disorders as complex conditions influenced by a multitude of factors, necessitating an integrative approach to diagnosis and treatment.

### 2.2.3 Diverse Modalities in Mental Health

In the rapidly evolving field of mental health diagnostics, a diverse array of modalities have been employed to facilitate the accurate diagnosis and prognosis of mental health disorders. From established clinical evaluations to modern digital tracking, this section examines the effectiveness and roles of different data modalities in mental health.

#### Clinical Tests

At the foundation of mental health diagnosis are clinical tests, encompassing structured interviews and self-assessment scales that gauge symptoms and functionality. Tools like the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Nemeroff et al., 2013) are pivotal in this process. The reliability of these tests largely hinges on the clinician's skill and the patient's openness and accurate self-reporting.

#### Neuro-cognitive Assessments

These assessments are vital for examining cognitive functions impacted by mental health disorders, measuring aspects like memory, focus, executive functions, and processing speed. Instruments such as the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) and the Mini-Mental State Examination (MMSE) (Tombaugh and McIntyre, 1992) are prevalent, aiding in detecting cognitive impairments linked with various mental health conditions.

#### Gene Expression

Exploring gene expression provides insights into the genetic foundations of mental disorders. Advances in genomics have led to the identification of genetic markers that could predispose individuals to specific mental health issues. This area of study assists in deciphering the molecular activities that contribute to these disorders, paving the way for tailored treatment strategies.

#### Social Functioning

Evaluating an individual's social functioning is essential to grasp the effects of mental health conditions on daily life. This evaluation includes the person's capacity to sustain relationships, perform work roles, and participate in social settings. Tools like the Social Functioning Scale (SFS) (Birchwood et al., 1990) are employed to assess these aspects, with impairments often signaling underlying mental health problems.

#### Neuroimaging

Neuroimaging techniques, such as Magnetic Resonance Imaging (MRI) (Vlaardingerbroek and Boer, 2013) and Positron Emission Tomography (PET) (Bailey et al., 2005), have become invaluable in mental health research, enabling the non-invasive study of brain structures and functions. They provide insights into the neurological basis of mental health conditions and have been instrumental in enhancing our understanding of depression, anxiety, and bipolar disorder.

#### Digital Monitoring

A newer method, digital monitoring, uses smartphones and wearable devices to passively gather data on an individual's behavior and physiological state. This includes monitoring physical activity, sleep, social interactions, and potentially even language use. Digital monitoring provides a continuous, real-time view of an individual's mental health, supporting early detection and timely intervention.

Each modality offers unique insights and contributes differently to a more holistic understanding of these conditions. As research and technology continue to advance, the integration of these modalities promises to enhance the accuracy and efficacy of mental health diagnostics, leading to better outcomes for individuals affected by these disorders.

### 2.2.4 Understanding Ultra High Risk (UHR)

In the realm of mental health, Ultra High Risk (UHR) has emerged as a pivotal criterion for understanding and intervening in the early stages of mental health disorders. UHR refers to a state characterized by the presence of prodromal symptoms that are suggestive, though not definitive, of developing a full-blown psychiatric disorder, particularly psychosis. The identification of individuals at UHR is grounded in the recognition of subtle yet significant changes in thoughts, emotions, and behaviors that precede the onset of a diagnosable mental health condition.

#### Characteristics of UHR

- **Symptomatic Precursors**: Individuals at UHR often exhibit attenuated symptoms. These include mild forms of hallucinations, delusions, or thought disorders, which are less severe than those observed in full-blown psychosis.
- **Genetic Vulnerability**: A family history of mental health disorders, especially psychosis, is a common feature in UHR individuals, indicating a genetic predisposition.
- **Functional Decline**: There is often a noticeable decline in social, academic, or occupational functioning, aligning with the onset of prodromal symptoms.
- **Subjective Distress**: UHR individuals frequently experience a significant level of distress or anxiety about their symptoms, aware that their experiences are unusual but not severe enough to warrant a full psychiatric diagnosis.

Understanding and identifying Ultra High Risk is a critical aspect of early diagnosis and prognosis of mental health disorders. It represents a shift towards a more preventative and patient-centric approach in mental healthcare. Predicting risk rather than waiting for the full manifestation of a disease allows for timely interventions, improved outcomes, and a more efficient use of healthcare resources.

Studies tracking UHR individuals reveal significant insights: the original 2003 study (Yung et al., 2003) noted that about 40.8% of people identified as UHR developed psychosis within a year. However, recent studies have highlighted a notable decline in transition rates among individuals classified as UHR. A 2012 study (Fusar-Poli et al., 2012) reported lower transition rates, with approximately 22% developing psychosis within one year, 29% within two years, and 36% within three years. More recent work (Hartmann et al., 2016; Fusar-Poli et al., 2020; De Pablo et al., 2021)

suggests a further decline to approximately 10–20%, particularly in cohorts enrolled through early intervention services. This trend is likely driven by earlier detection, expanded access to care, and broader diagnostic criteria capturing more heterogeneous risk groups.

This evolution in findings underscores that while UHR is a critical criterion for early detection, it also encounters challenges, notably high rates of false positives (Yung and Nelson, 2013). This highlights the need for additional modalities beyond traditional clinical assessments to enhance predictive accuracy, as well as the need for temporally-aware and personalized approaches — such as those presented in this thesis — to refine risk stratification and improve long-term outcome prediction.

### 2.2.5 The LYRIKS Dataset: Multimodal UHR Data

The Longitudinal Youth-At-Risk Study (LYRIKS) is an observational study conducted in Singapore amongst youths aged 14-29 to identify biological risk factors that can help in identifying individuals at UHR for developing psychosis (Lee et al., 2013). The data collected in the study consists of a collection of clinical tests, cognitive assessments, gene expression, and social functioning of UHR individuals.

The LYRIKS study initially categorized participants into two groups: Control and UHR, based on the Comprehensive Assessment of At-Risk Mental States (CAARMS) (Yung et al., 2002). Over time, some participants originally classified as UHR showed negative CAARMS results and were labeled as 'Remitted'. Those who continued to meet the UHR criteria were termed 'Maintained', and those who developed psychosis were labeled as 'Converted'. The dataset thus provides a dynamic view of the progression of mental health status over time.

Data was collected periodically over two years, with clinical, cognitive, and social data gathered every six months and genetic data every twelve months. The richness of the dataset lies in its multimodal nature, encompassing:

1. **Clinical Tests**: Including CAARMS, PANSS (Positive and Negative Syndrome Scale, Kay, Fiszbein, and Opler, 1987), CDSS (Calgary Depression Scale for Schizophrenia, Addington, Addington, and Maticka-Tyndale, 1993), and BAI (Beck Anxiety Inventory, Beck et al., 1993).

2. **Cognitive Assessments**: Such as Snakes in the Grass (Snk) (Öhman, Flykt, and Esteves, 2001), Brief Assessment of Cognition in Schizophrenia (BACS) (Keefe et al., 1999), Continuous Performance Test (CPT) (Cornblatt et al., 1988), and Perceptual Closure Task (PerClose) (Doniger et al., 2000).

3. **Social Functioning**: Measured through the High-Risk Social Challenge Task (HiSoC) (Gibson et al., 2010).

4. **Gene Expression**: RNA-sequencing of peripheral whole blood samples.

Overall the LYRIKS study, taking into account all timepoints, involved 667 participants. However, not all participants provided data across every modality at each collection point. Data from a subset of 107 participants is used in all subsequent analyses because they had data across all modalities and most time steps. To address the few remaining missing values, temporal interpolation was used. This technique involves using data from previous and subsequent time steps to estimate and fill in missing values.

The primary task we will explore with this dataset is to predict the risk status of individuals at various future time points (0, 6, 12, 18, and 24 months ahead) using

different modalities, both individually and combined. The longitudinal nature of the data allows for an exploration of cross-sectional versus longitudinal data use in prediction. For longitudinal studies, three timesteps of data, equating to one year, are used as input for the models. Both 'Converted' and 'Maintained' participants are categorized under the 'At-risk' label, while 'Remitted' and 'Control' are categorized as 'Healthy'.

To maximize the dataset's utility and get more generalized estimates of performance, each participant's data at different time points was treated as a new sample in the predictive models. For instance, when predicting the risk status 6 months ahead, each participant could contribute up to 4 distinct samples (their data at 0, 6, 12, and 18 months to predict risk status at 6, 12, 18 and 24 months respectively). However, to maintain the integrity of the analysis and avoid data leakage, all data from an individual participant was either included entirely in the training set or entirely in the test set. This method ensured that the predictive performance was not artificially inflated by having training and test data from the same individual, while improving the robustness of the performance measured.

Overall, this dataset provides a unique opportunity to explore multimodal prediction in mental health, considering both static (cross-sectional) and dynamic (longitudinal) aspects of risk assessment. The integration of various data types promises a more holistic and accurate prediction model, enhancing our ability to intervene early in individuals at risk of psychosis. The findings from this study will contribute significantly to the field of mental health diagnostics, offering insights into effective strategies for early identification and intervention in psychosis.

## 2.3 The Power of Longitudinal Data

### 2.3.1 What Longitudinal Data Offers

Longitudinal data, collected over time from the same subjects, provides unique insights into the progression and dynamics of mental health disorders. This type of data is invaluable for understanding the natural history of a disease, identifying risk factors, and evaluating the effectiveness of interventions. It allows for the analysis of changes within individuals, offering a deeper understanding of causal relationships and the potential for personalizing treatment strategies. Unlike cross-sectional data, which offers a snapshot at a single point in time, longitudinal data can uncover patterns and trends that only emerge over time. This can lead to the development of predictive models that are better able to forecast future outcomes for individuals based on their specific trajectories. By capturing the temporal dimension of data, researchers and clinicians can identify critical periods for intervention, track the effectiveness of treatments over time, and adjust strategies as needed to improve patient outcomes.

### 2.3.2 Historical Statistical Approaches

Historical statistical approaches provide a robust foundation for analyzing longitudinal data, particularly in healthcare and psychological research. These methods, developed over decades, are instrumental in understanding the dynamic nature of data collected over time from the same subjects. Here, we dive into some key statistical methodologies.

### Mixed Models

Mixed models, also known as multilevel models, incorporate both fixed effects (common to all individuals) and random effects (individual-specific variations) to analyze data collected across multiple time points (Bolker et al., 2009). This flexibility makes mixed models ideal for addressing within-subject correlations and analyzing the impact of time-invariant and time-varying covariates on the outcome of interest. The general form of a mixed model can be expressed as:

$$y_{ij} = X_{ij}\beta + Z_{ij}u_i + \epsilon_{ij} \tag{2.1}$$

where $y_{ij}$ is the response for the $i$th subject at time $j$, $X_{ij}$ and $Z_{ij}$ are the matrices of fixed and random effects covariates, respectively, $\beta$ is the vector of fixed effects coefficients, $u_i$ is the vector of random effects for the $i$th subject, and $\epsilon_{ij}$ is the error term.

For example, in the context of healthcare, mixed models have been used to analyze the effectiveness of treatments across different patient groups over time, allowing for personalized treatment plans. Cnaan, Laird, and Slasor, 1997 used mixed models to evaluate the longitudinal effects of medication on growth rates in children, demonstrating their applicability in monitoring and adjusting treatments.

### Growth Curve Modeling

Growth curve modeling, a subtype of mixed models, focuses on individual trajectories over time (Preacher, 2008). It enables the examination of how subjects grow or change across different stages of life or treatment phases. It is particularly adept at capturing both linear and nonlinear patterns of change, accommodating diverse developmental trajectories. The mathematical representation of a simple linear growth curve model is:

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \epsilon_{ij} \tag{2.2}$$

Here, $y_{ij}$ represents the outcome for the $i$th individual at time $j$, $t_{ij}$ is the time variable, $\beta_{0i}$ and $\beta_{1i}$ are the individual-specific intercept and slope (reflecting initial status and rate of change), and $\epsilon_{ij}$ is the error term. This model can be extended to include non-linear effects by adding higher-order terms of $t$.

It's particularly useful in pediatric psychology for tracking developmental progress and in gerontology for assessing cognitive decline. An example can be found in Biesanz et al., 2004, where growth curve modeling was applied to understand how academic and social experiences influence student development over time

### Time Series Analysis

Time series analysis, which includes techniques like autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA) models (Box and Pierce, 1970), is pivotal for forecasting future data points based on past observations. The ARIMA model, combining AR and MA components with differencing to achieve stationarity, is particularly versatile. The ARIMA model is defined by three parameters $(p, d, q)$ and can be expressed as:

$$(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d y_t = (1 + \sum_{i=1}^{q} \theta_i L^i)\epsilon_t \tag{2.3}$$

where $L$ is the lag operator, $p$ is the order of the AR part, $d$ is the degree of differencing, $q$ is the order of the MA part, $\phi_i$ are the AR coefficients, $\theta_i$ are the MA coefficients, and $\epsilon_t$ is white noise.

Time series analysis is particularly useful for analyzing and predicting trends, cycles, and seasonal variations in longitudinal data. It's widely used in epidemiology to forecast disease outbreaks and evaluate the impact of public health interventions. A notable application is the use of ARIMA models to predict influenza trends, as detailed by Box et al., 2015, guiding public health policies and vaccination strategies.

#### Survival Analysis

Survival analysis focuses on the time until an event of interest occurs, such as relapse or recovery in the context of health (Klein, Moeschberger, et al., 2003). It accounts for censoring, where the outcome may not be observed for all subjects within the study period. The Cox proportional hazards model, a semi-parametric approach, is widely used for its flexibility in assessing the impact of covariates on survival times without specifying the baseline hazard function (Cox, 1972). The model is given by:

$$h(t|X) = h_0(t) \exp(X\beta) \tag{2.4}$$

where $h(t|X)$ is the hazard function at time $t$ for an individual with covariates $X$, $h_0(t)$ is the baseline hazard function, and $\beta$ is the vector of coefficients estimating the effect of covariates.

Techniques like the Kaplan-Meier estimator and Cox proportional hazards model are key tools in survival analysis, allowing for the estimation of survival times and the assessment of factors contributing to the risk. It's crucial in clinical trials for comparing the efficacy of treatments over time. The use of Cox proportional hazards models to evaluate factors affecting survival rates in cancer patients is a prime example, allowing clinicians to tailor treatments based on individual risk profiles, highlighted in the original paper Cox, 1972.

These statistical approaches have played a foundational role in analyzing longitudinal data, enabling us to understand complex dynamics in health progression and treatment outcomes over time. But given the complex and dynamic nature of mental health disorders, which necessitates capturing non-linear interactions and temporal dependencies in the data, more advanced techniques such as Deep Learning and Liquid State Machines (LSMs) have emerged as the state-of-art. They offer the flexibility, scalability, and depth required to address these challenges effectively. Hence, our research leverages those cutting-edge methodologies to provide nuanced insights and predictions that traditional models may not fully capture. Deep Learning and LSM approaches are discussed in more detail in the subsequent sections.

### 2.3.3 Deep Learning Approaches

Deep Learning (DL) represents a paradigm shift in analyzing longitudinal data, offering unparalleled capabilities in handling complex, high-dimensional datasets. DL models, particularly those designed for sequence processing, such as Recurrent Neural Networks (RNNs) (Elman, 1991), Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997), and Transformers (Vaswani et al., 2017), excel in capturing temporal dependencies and patterns in data collected over time. These models have revolutionized the way we approach longitudinal data analysis

in healthcare, providing insights into patient trajectories, disease progression, and treatment outcomes with remarkable accuracy.

**Recurrent Neural Networks (RNNs)**

RNNs are designed to process sequential data, making them suitable for longitudinal data analysis. They uniquely capture temporal dependencies in data sequences by maintaining a hidden state that acts as a memory of all previously seen inputs. Mathematically, the hidden state at time $t$, denoted as $h_t$, is updated based on the previous hidden state $h_{t-1}$ and the current input $x_t$, as shown in the formula:

$$h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \tag{2.5}$$

where $f$ is a nonlinear activation function, $W_{hh}$ is the weight matrix for the hidden state, $W_{xh}$ is the weight matrix for the input, and $b_h$ is the bias. Despite their prowess, RNNs often struggle with long-term dependencies due to the vanishing gradient problem, limiting their effectiveness in capturing long-range temporal relationships.

RNNs have been applied in predicting patient outcomes, modeling disease progression, and personalizing treatment recommendations based on historical patient data.

**Long Short-Term Memory Networks (LSTMs)**

LSTMs, a special kind of RNN, address the vanishing gradient problem through the introduction of memory cells and gates that regulate the flow of information. These gates — the input, forget, and output gates — allow LSTMs to selectively remember and forget information, making them highly effective for analyzing longitudinal data where past information significantly influences future outcomes. The core equations governing an LSTM unit include:

$$\begin{aligned}
\textbf{\textit{Forget Gate}} &: f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
\textbf{\textit{Input Gate}} &: i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
\textbf{\textit{Candidate Cell State}} &: \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
\textbf{\textit{Current Cell State}} &: C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \\
\textbf{\textit{Output Gate}} &: o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
\textbf{\textit{Current Hidden State}} &: h_t = o_t * \tanh(C_t)
\end{aligned} \tag{2.6}$$

where $\sigma$ denotes the sigmoid activation function, and $*$ represents element-wise multiplication. These operations enable LSTMs to maintain a long-term memory, crucial for processing longitudinal health data.

This capability makes LSTMs well-suited for analyzing longitudinal health data, where understanding long-term patterns is essential for accurate prognosis and diagnosis. LSTMs have shown significant success in tasks such as predicting the onset of diseases, understanding patient risk factors, and forecasting hospital readmissions.

#### Transformers

Transformers revolutionized sequence processing through a self-attention mechanism, allowing the model to focus on different parts of the input sequence to predict each output element. Unlike RNNs and LSTMs, Transformers process all input data simultaneously, making them highly efficient and effective for handling large datasets. The key formula for the self-attention mechanism in Transformers is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.7}$$

where $Q$, $K$, and $V$ represent the queries, keys, and values matrices respectively, and $d_k$ is the dimension of the keys. This mechanism allows the model to dynamically weigh the significance of different parts of the input sequence, adapting to the varying importance of temporal features in longitudinal data.

Transformers, initially designed for natural language processing tasks, have demonstrated exceptional ability in handling sequence data. In healthcare, Transformers have been applied to predict disease progression, patient outcomes, and treatment responses with remarkable accuracy. Their ability to process and integrate diverse data types over time makes them particularly suited for analyzing the multifaceted nature of mental health disorders.

Deep Learning approaches offer significant advantages over traditional models, including the ability to automatically learn complex representations of data without the need for manual feature engineering. This automation, combined with their superior predictive performance, makes DL models particularly attractive for longitudinal data analysis in the healthcare domain. Their application has led to more accurate, timely, and personalized insights into patient health, advancing the field of predictive healthcare.

### 2.3.4 Neuromorphic Approaches

Neuromorphic computing, inspired by the biological processes of the human brain, represents a significant leap in the evolution of computational models, particularly in the context of processing complex temporal data like longitudinal health records. At the heart of neuromorphic computing are Spiking Neural Networks (SNNs), which are considered the third generation of neural network models, following perceptrons and traditional artificial neural networks. SNNs offer a more granular simulation of neural activity compared to their predecessors.

#### Spiking Neural Networks (SNNs)

Spiking Neural Networks (SNNs) stand at the forefront of this paradigm, mirroring the intricate mechanisms of biological neural networks. Unlike traditional neural networks that process information in a continuous flow, SNNs operate using discrete events called spikes, which are analogous to the action potentials in biological neurons. This spiking mechanism enables SNNs to process information in a way that closely resembles the dynamic and temporal nature of biological neural processing.

The core advantage of SNNs lies in their ability to capture the temporal dynamics of input signals. Each neuron within the network responds only when the incoming signal's strength exceeds a certain threshold, leading to a spike. This event-driven mechanism allows SNNs to be highly efficient, as computations are carried out only

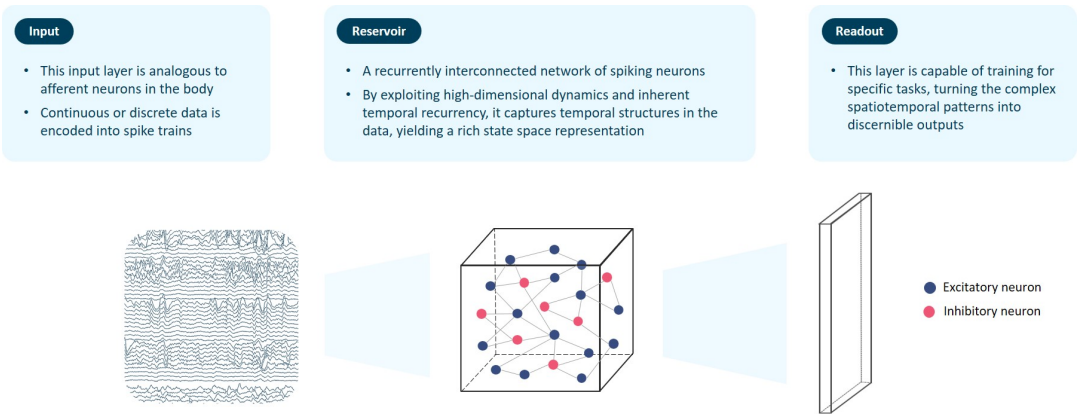


FIGURE 2.1: Liquid State Machine (LSM) Architecture

in response to stimuli, reducing the need for constant computation and enabling the network to handle time-based data more naturally.

Moreover, SNNs are inherently capable of learning temporal patterns through mechanisms such as spike-timing-dependent plasticity (STDP), a form of synaptic plasticity where the timing of spikes influences the strength of connections between neurons. This feature makes SNNs particularly suitable for tasks involving time-series data, offering a robust framework for understanding and predicting dynamic processes over time (Doborjeh et al., 2021b; Doborjeh et al., 2021a; Doborjeh et al., 2020; Kasabov et al., 2023).

### Liquid State Machines (LSMs)

Liquid State Machines (LSMs) (Maass, Natschläger, and Markram, 2002) introduce a novel approach to processing temporal patterns in longitudinal data, especially within the realm of neuroscience and cognitive computing. LSMs are part of the third generation of neural network models, emphasizing the dynamics of spiking neural networks. Their architecture enables the processing of information in a way that mimics the biological processes of the human brain, allowing for a more nuanced understanding of temporal and spatial patterns in data.

LSMs operate on the principle of creating a "liquid" or reservoir of sparsely connected neurons, through which input signals propagate and transform. This reservoir, typically a fixed recurrent neural network, retains a dynamic history of the input signals, enabling the machine to capture and compute complex patterns of activity over time. In the context of healthcare and mental health analysis, LSMs offer promising avenues for modeling and predicting neurological and psychiatric conditions by closely mimicking the way the human brain processes information.

The application of LSMs in analyzing longitudinal data stems from their ability to handle the variability and complexity inherent in such data, making them particularly suited for tasks requiring the analysis of intricate temporal sequences and the prediction of outcomes based on subtle changes in patterns over time. Their potential in advancing personalized medicine, through the precise modeling of individual patient data over extended periods, marks a significant step forward in the predictive analytics of mental health disorders.

In conclusion, neuromorphic approaches, particularly through the lens of Spiking

Neural Networks (SNNs) and Liquid State Machines (LSMs), represent a cutting-edge frontier in the analysis of longitudinal data. By harnessing the intricate dynamics of spiking mechanisms, these models offer a closer approximation to the way biological neural systems process information over time. The adoption of LSMs, with their unique reservoir computing capabilities, underscores a significant advancement in our ability to model and interpret complex temporal patterns in health-related data.

## 2.4 Conclusion

The field of mental health diagnostics and prognostics is undergoing a profound transformation, driven by advances in data collection, analysis techniques, and computational models. This chapter has explored the multifaceted landscape of current approaches and emerging technologies in this critical area of healthcare.

We have seen how the shift from unimodal to multimodal approaches is enhancing our understanding of mental health disorders. By integrating diverse data types - from clinical assessments and cognitive tests to genetic markers and neuroimaging - researchers and clinicians are developing a more comprehensive picture of mental health conditions. This holistic approach is particularly crucial in identifying and managing Ultra High Risk (UHR) states, where early intervention can significantly alter the trajectory of potential disorders.

The power of longitudinal data has emerged as a central theme in this review. The ability to track individuals over time, as exemplified by studies like LYRIKS, provides invaluable insights into the progression of mental health conditions and the factors influencing their development. This temporal dimension is essential for developing more accurate predictive models and personalized treatment strategies.

Advancements in analytical techniques have kept pace with the increasing complexity and volume of data. From traditional statistical methods to sophisticated machine learning approaches, the tools available for analyzing mental health data are becoming increasingly powerful and nuanced. Deep learning models, particularly those designed for sequence processing like RNNs, LSTMs, and Transformers, have shown remarkable potential in capturing the complex temporal dynamics. Neuromorphic approaches, such as Spiking Neural Networks and Liquid State Machines, represent the more exploratory cutting edge of this field. These models, inspired by the biological processes of the human brain, offer new possibilities for understanding and predicting mental health outcomes.

However, it is important to note that with these advancements come new challenges. Ethical considerations in data collection and use, the need for interpretable AI models in clinical settings, and the integration of these technologies into existing healthcare systems are all areas that require careful consideration and ongoing research.

In conclusion, the field of mental health diagnostics and prognostics stands at an exciting juncture. By leveraging the power of multimodal, longitudinal data and cutting-edge analytical techniques, we are poised to make significant strides in our understanding and management of mental health disorders. This progress offers hope for more effective, personalized, and timely interventions, ultimately improving the lives of millions affected by mental health conditions worldwide.

# Chapter 3

# Biomarker Discovery: Identifying Indicators of Health

## 3.1 Introduction

In the quest to unravel the mysteries of human health and disease, the discovery and utilization of biomarkers have emerged as a pivotal element. Biomarkers, or biological markers, are measurable indicators that can signify various health or disease states in the body. They serve as vital tools in medical research and practice, offering insights into the complex mechanisms of diseases, aiding in early diagnosis, monitoring disease progression, and guiding therapeutic interventions.

The accurate diagnosis and prognosis of mental health disorders rely on the identification of reliable biomarkers that can provide insights into the underlying biological mechanisms of these conditions. Biomarkers, particularly those derived from genetic data, hold significant promise for enhancing our understanding of mental health disorders and improving the accuracy of diagnostic and prognostic models. However, the discovery of effective biomarkers is challenging due to the complex and multifactorial nature of these disorders.

This chapter focuses on the development of a novel method for biomarker discovery using gene expression data, addressing the research objectives of identify robust early markers of mental health (Objective 1 and Question 1). Biomarkers discovered through genetic data provide critical pieces of the larger puzzle of mental health, facilitating the integration of genetic information with other data types such as clinical assessments, social functioning, and cognitive assessments. By doing so, it becomes possible to create a more holistic and robust diagnostic and prognostic model, aligning with Research Question 4: How can multimodal data sources be effectively integrated to improve the diagnosis and prognosis of mental health disorders?

Furthermore, understanding genetic biomarkers helps in tracking changes over time, providing a deeper insight into the temporal dynamics of mental health conditions. This is crucial for identifying early markers of mental health deterioration and improving the accuracy of predictive models, thus addressing Research Question 2: What are the key temporal patterns and early markers that can be identified using longitudinal data for predicting mental health outcomes?

In this chapter, the Filter and Wrapper Stacking Ensemble (FWSE) approach is developed and validated to identify significant genetic markers associated with mental health disorders. This method enhances the integration of gene expression data with other modalities and improves the predictive power of the models by incorporating longitudinal data analysis.

The chapter is structured as follows. First, a review of existing methods for biomarker discovery in mental health research is presented, highlighting their strengths

and limitations. Next, the chapter introduces the FWSE approach, detailing its components and the rationale behind its design. The methodology section describes the datasets used, the preprocessing steps, and the experimental setup for validating the FWSE approach. Results from the experiments are then presented and analyzed, demonstrating the effectiveness of the FWSE method in identifying significant genetic markers. Finally, the chapter discusses the implications of the findings, potential applications in clinical settings, and future directions for research in biomarker discovery.

By developing and validating a novel method for biomarker discovery, this chapter contributes to the overarching goal of improving the diagnosis and prognosis of mental health disorders. The insights gained from this research will inform subsequent chapters, where the identified biomarkers will be integrated into multimodal predictive models and further validated using real-world datasets.

## 3.2 Background and Context

The human body is an incredibly complex system that relies on the coordinated efforts of trillions of cells to function properly. Each cell contains thousands of genes, proteins, and other molecules that carry out specific roles and processes that are essential for life. For example, the heart cells contract rhythmically to pump blood, immune cells produce antibodies to fight infection, and neuronal cells transmit electrical signals between the brain and body. No single cell works in isolation - cells communicate with each other through signaling pathways, molecular interactions, and metabolic processes to generate the intricate physiology of our organs and tissues.

The intricate workings of the cell have long captivated scientists. What makes some cells thrive while others falter? How does a cell know to become part of the heart or the brain? What happens at the molecular level during disease? To find answers, researchers needed more than a glimpse into the cellular world - they required a comprehensive analysis of the countless molecules within cells.

To find answers, they needed a way to closely examine everything inside the cells – the genes, proteins, and all the tiny chemical reactions. This is where the power of omics technologies comes into play. Omics is a suite of advanced scientific methods that allow researchers to look at thousands of these tiny components all at once. Omics technologies empower the discovery of biomarkers via comparative analyses. By comparing the set of tiny components present in one state e.g. healthy individuals, against another state e.g., individuals suffering from disease, we can identify differential components that has the potential to serve as biomarkers.

### 3.2.1 Understanding Omics Data

The term 'omics' encompasses a suite of disciplines in biological sciences, each focusing on the study of a specific type of molecular data. At the forefront is genomics, which involves the exhaustive study of an organism's complete set of DNA. This includes analyzing the structure and function of genes, as well as understanding genetic variations and mutations. For instance, genomics has been instrumental in mapping the human genome, leading to breakthroughs in identifying genetic markers for diseases like cystic fibrosis (Drumm et al., 2005) or breast cancer (Barzaman et al., 2020).

Another important branch is transcriptomics, which focuses on RNA and gene expression. It examines how genes are transcribed into RNA, and how this expression changes under different conditions. Transcriptomics has been crucial in understanding processes like how cells respond to stress (Tang et al., 2018) or how cancer cells differ from normal cells (Chen et al., 2022).

Proteomics studies the functions, structures, and interactions of proteins. Proteins are the workhorses of the cell, carrying out the instructions encoded in genes. Proteomics can reveal how protein modifications affect cellular signaling pathways (Kolch and Pitt, 2010) or how protein interactions change in neurodegenerative diseases (Zhang et al., 2021). Metabolomics examines the metabolic processes by studying metabolites, the small molecules involved in metabolism. Metabolomics can provide insights into an organism's physiological state, such as identifying metabolic biomarkers for diseases like diabetes (Roberts, Koulman, and Griffin, 2014) or understanding the metabolic changes that occur during drug treatment (Clarke and Haselden, 2008).

Emerging omics branches like lipidomics and epigenomics further extend our understanding. Lipidomics studies lipids, the fat molecules crucial for cell structure and energy storage (Wenk, 2005), while epigenomics focuses on epigenetic changes, modifications on the DNA that affect gene expression without altering the DNA sequence itself (Suzuki and Bird, 2008). These branches collectively offer an integrated perspective of an organism's biological systems and processes.

Advances in high-throughput technologies have revolutionized the collection of omics data. Next-Generation Sequencing (NGS) has significantly accelerated genomic and transcriptomic research by allowing for the rapid sequencing of DNA and RNA at lower costs (Goodwin, McPherson, and McCombie, 2016). Similarly, mass spectrometry and advanced chromatography techniques have advanced the fields of proteomics and metabolomics, enabling the detailed analysis of proteins and metabolites (De Hoffmann and Stroobant, 2007).

More recently, the advent of single cell omics technologies alongside spatial profiling technologies allows us to glimpse into the workings of individual cells, or distribution of distinct cellular populations across two and three dimensional space. These high resolution techniques not only provide opportunity to discover distinct cell populations associated with disease, but also has the potential to understand each person's disease trajectory and mechanism in a personalized manner

Omics data has vast potential in medical research, particularly in understanding the molecular basis of diseases. Amongst the various branches, the high throughput study of genes or genomics has been instrumental in identifying genetic markers associated with various diseases, enabling personalized medicine approaches (Tremblay and Hamet, 2013). Additionally, proteomics and metabolomics provide insights into disease mechanisms and potential therapeutic targets (Butterfield, Boyd-Kimball, and Castegna, 2003; Rabinowitz et al., 2011). However, each omics only provides insight into a particular branch or molecular entity. Biological systems are far more complex, where different molecular entities work collectively to achieve various purposes. Hence, the future of omics research lies in the integration of different omics data types, known as multi-omics. This approach promises a more comprehensive understanding of biological systems and disease etiology (Hasin, Seldin, and Lusis, 2017).

### 3.2.2 What are Biomarkers?

Biomarkers, short for biological markers, are measurable indicators of some biological state or condition. They are crucial in understanding the interaction between biological systems and can be a physical, chemical, or molecular indicator of health or pathogenic processes (Strimbu and Tavel, 2010). Biomarkers can be derived from bodily fluids, tissues, or as a measure of biological processes at the molecular, cellular, or physiological level (Jain and Jain, 2010).

Given their diverse origins and functions, biomarkers are classified into various types, each serving distinct roles based on their application and biological characteristics. These classifications underline the versatility and critical importance of biomarkers in multiple domains of healthcare and medical research. The key use cases for biomarkers include:

- **Disease Detection and Diagnosis**: Biomarkers serve a crucial role in the early detection and diagnosis of diseases, often identifying pathological changes before clinical symptoms become apparent. For instance, the elevation of carcinoembryonic antigen (CEA) can suggest colorectal cancer (Duffy, 2001), while high levels of alpha-fetoprotein (AFP) are associated with liver cancer (Galle et al., 2019). These biomarkers facilitate earlier intervention and potentially more effective treatment outcomes.

- **Monitoring Disease Progression**: In chronic conditions, biomarkers are instrumental in monitoring disease progression. Glycated hemoglobin (HbA1c), for example, provides an overview of the average blood sugar control over the past two to three months in diabetes patients, playing a key role in tailoring treatment regimens and lifestyle modifications (Rohlfing et al., 2002).

- **Predicting Disease Susceptibility**: Genetic biomarkers can reveal an individual's predisposition to certain diseases, allowing for preemptive measures and personalized surveillance strategies. For example, mutations in the BRCA1 and BRCA2 genes are well-established markers for an increased risk of developing breast and ovarian cancers (Petrucelli, Daly, and Feldman, 2010), often leading to more stringent screening protocols and preventative strategies.

- **Evaluating Treatment Responses**: Biomarkers are pivotal for assessing the effectiveness of treatment regimens. HER2/neu, a protein overexpressed in some breast cancer patients, not only guides the use of targeted therapies like trastuzumab (Piccart-Gebhart et al., 2005) but also serves to monitor the patient's response to these treatments, thereby playing a significant role in personalized cancer care.

- **Drug Development and Clinical Trials**: Biomarkers are integral in the drug development process, especially in clinical trials where they are used to assess the safety and efficacy of novel therapeutic agents. They provide critical data points for the evaluation of pharmacodynamics and pharmacokinetics, helping to streamline the drug development process by identifying potentially successful therapeutic candidates early on (Frank and Hargreaves, 2003).

- **Personalized Medicine**: Personalized medicine, the tailoring of medical treatment to individual patient characteristics, relies heavily on biomarkers (La Thangue and Kerr, 2011). Pharmacogenomic biomarkers, for example, can predict an individual's response to antidepressants (Leuchter et al., 2010), allowing clinicians to select medications that are more likely to be effective and

have fewer side effects, thereby improving patient outcomes and treatment satisfaction.

The advent of omics technologies has expanded the potential for biomarker discovery. Genomic, proteomic, metabolomic, and transcriptomic studies have identified numerous biomarkers that provide insights into the molecular mechanisms underlying various diseases (Horgan and Kenny, 2011). This omics-driven approach is enhancing our understanding of complex diseases like cancer, cardiovascular diseases, and neurological disorders (Sun and Hu, 2016).

Despite the potential, there are challenges in biomarker research, discussed in the next section.

### 3.2.3 Challenges in Biomarker Discovery

The discovery of biomarkers through omics data analysis presents unique challenges, primarily due to the nature and complexity of the data. These challenges significantly impact the reliability and applicability of the discovered biomarkers in clinical settings. Some of these challenges are discussed below.

- **Dimensionality versus Sample Size**: One of the most significant challenges in omics data analysis is the high dimensionality of the data relative to the sample size. This is commonly known as the Curse-of-dimensionality problem. Omics datasets often contain tens of thousands of features (genes, proteins, metabolites, etc.), while the number of samples (patients, experimental conditions) is comparatively smaller, often at least 2-3 orders of magnitude lower. This imbalance leads to issues in feature selection, as it increases the risk of overfitting the model to the data and finding spurious correlations that do not generalize well to new data sets (Loscalzo, Yu, and Ding, 2009).

- **Variability Across Methods**: The choice of feature selection method can significantly impact the results in omics data analysis. Different feature selection algorithms may identify different sets of features as being the most important, leading to inconsistencies in the biomarkers identified. This method-dependence is a substantial challenge, complicating the process of distinguishing genuinely relevant biomarkers from those identified due to methodological biases (Ioannidis et al., 2005).

- **Inconsistency Across Data Perturbations**: Another issue in biomarker discovery is the stability of feature selection. The same algorithm, when applied to slightly different variations of the same data (perturbations), may not consistently select the same features as the most important. This poses a significant challenge to the reliability of the feature selection algorithm, as it questions the reproducibility of the selected features within a given dataset (He and Yu, 2010).

- **Non-Reproducibility Across Different Datasets**: A major hurdle in omics data analysis is the reproducibility of findings across different datasets. Often, biomarkers identified in one study may not be replicated in subsequent studies using different datasets. This lack of reproducibility is a significant impediment to the clinical application of identified biomarkers (McDermott et al., 2013). It underscores the need for robust feature selection methods that can identify biomarkers consistently across diverse datasets.

These challenges highlight the need for sophisticated data analysis methods in omics research. Improving the reproducibility and reliability of biomarker discovery requires advanced analytical techniques, careful data preprocessing, and validation of findings across multiple independent datasets. Biomarker development is an inter-disciplinary enterprise. Hence, the integration of computational and biological expertise is essential to overcome these hurdles and to harness the full potential of omics data in biomarker discovery and clinical applications.

## 3.3 Methods for Biomarker Discovery

### 3.3.1 Fundamentals of Feature Selection

Feature selection involves identifying the most relevant features (such as genes, proteins, or metabolites) from vast datasets, which can then be used for accurate disease diagnosis, prognosis, or understanding the disease's molecular mechanisms. The primary benefit of performing feature selection is to reduce data dimensionality, therefore eliminating irrelevant or redundant data and allowing us to focus on a prioritized subset of candidate features (e.g. genes, proteins) that may be suitable as biomarkers. Selected features can be assayed for real-world deployability as biomarkers by incorporating them into models. As opposed to simply inputting all features, feature selection reduces the input feature set to only informative ones, which often improves model performance, thereby enhancing the predictive power and interpretability of the model. Particularly for omics data, which often contains tens of thousands of features, effective feature selection is essential for handling the 'curse of dimensionality' and uncovering meaningful biological insights.

From a selection strategy perspective, feature selection methods can be subdivided into filter, wrapper, and embedded methods (Zhang, Jonassen, and Goksøyr, 2021). These are discussed below:

#### Filter Methods

Filter methods assess the importance of features based on intrinsic properties of the data, independent of any learning algorithm. They usually involve statistical tests for assessing the correlation or difference between features and the target variable. Filter methods are computationally efficient and generally have lower complexity. They are good at removing irrelevant features and are not prone to overfitting. Filter methods might overlook feature dependencies and interactions since they evaluate each feature in isolation. They may also ignore the model's biases and preferences, leading to suboptimal feature selection for specific predictive models. Common filter methods include the Chi-square test, information gain, and correlation coefficient metrics.

#### Wrapper Methods

Wrapper methods assess subsets of features based on their usefulness to a given predictive model. They use a predictive model to evaluate the combination of features and determine which set produces the best performance for that particular model. These methods can find the best feature set for a given model, considering feature interactions. They are generally more accurate than filter methods as they tailor the feature set to the model's needs. Wrapper methods are computationally intensive

and have a higher risk of overfitting, especially with smaller datasets. They also depend heavily on the choice of the learning algorithm. Recursive Feature Elimination (RFE) and forward feature selection are typical examples of wrapper methods.

#### Embedded Methods

Embedded methods incorporate feature selection as part of the model training process. They perform feature selection during the model training and are specific to given learning algorithms. These methods can capture feature interactions while being computationally more efficient than wrapper methods. They balance between the filter and wrapper methods, providing good model performance with reasonable computational cost. The main limitation is that they are tied to specific models, and the selected features may not be optimal for other types of models. Lasso (Least Absolute Shrinkage and Selection Operator) and Decision Trees are examples of embedded methods, where feature selection is inherent to the model training process.

Some approaches combine multiple feature selection methods from these different categories, forming a fourth category known as 'integrated' or 'ensemble' methods. These are discussed in more detail in Section 3.4.

In summary, feature selection is a pivotal step in biomarker discovery, particularly in the analysis of high-dimensional omics data. Understanding the strengths and limitations of different feature selection methods is crucial for selecting the most appropriate method for a given study and ensuring the reliability and validity of the selected features.

### 3.3.2 Traditional Feature Selection Methods

In this section, we review popular feature selection methods across the filter, wrapper and embedded categories frequently used for tasks such as biomarker discovery. Examining the mathematical underpinnings and applications of these fundamental approaches is essential for building more reliable and robust methods.

#### F-Statistic

The F-Statistic is the result of an Analysis of Variance (ANOVA) F-test for a feature across target classes. It measures the ratio of the intra-group variance to the inter-group variance, where the groups correspond to different target classes. The F-Statistic is effective for identifying features that show statistically significant differences between groups. However, it assumes normal distribution of data and equal variances across groups, which may not always hold in real-world datasets.

#### Signal-to-Noise Ratio (SNR)

SNR is a measure of the separation between the means of two classes relative to the variability within each class. It is mathematically defined as:

$$SNR_i = \frac{\left| M_i^{(\text{class 1})} - M_i^{(\text{class 2})} \right|}{Std_i^{(\text{class 1})} + Std_i^{(\text{class 2})}} \tag{3.1}$$

where $M_i^{(\text{class})}$ and $Std_i^{(\text{class})}$ are the mean and standard deviation of the $i^{th}$ feature for a given class.

SNR is robust for assessing feature importance, especially when distinguishing between two classes (Kasabov, 2007a). However, it assumes feature independence, which may not be valid in high-dimensional omics data.

#### Lasso Regression

Lasso Regression is a variant of logistic regression that incorporates L1 penalization. This penalization leads to sparse solutions, effectively reducing the number of features by setting many coefficients to zero. Lasso is powerful for feature selection, particularly in contexts with many features, as it can identify a smaller, more relevant set (Tibshirani, 1996). However, it may be sensitive to outliers and less effective when the number of features significantly exceeds the number of samples.

#### Random Forest

Random Forest is an ensemble method that uses multiple decision trees for classification or regression. Feature importance is derived from the decrease in impurity at each node of the trees, with more important features appearing earlier in the trees (Breiman, 2001). Random Forest is robust and versatile, suitable for various data types and complex relationships. However, it can be computationally demanding, especially with large, high-dimensional datasets.

#### Recursive Feature Elimination (RFE)

RFE uses an external estimator to assign importance to features. It systematically removes the least important features, re-evaluating importance at each step, until a specified number of features remains. RFE, particularly when combined with models like Linear SVMs, Logistic Regression, and Random Forests, is effective in identifying significant features in high-dimensional datasets (Guyon et al., 2002). Its main limitation is the potential computational intensity, depending on the size of the dataset and the complexity of the model used.

Each of these approaches has been widely applied in field of biomarker discovery. For example, elastic net has been used for diagnosing papillary thyroid carcinoma (Park et al., 2021), Random Forest for tracking prostate cancer progression (Toth et al., 2019), and Lasso regression and RFE using Support Vector Machines (RFE-SVM) for identifying therapeutic targets in ferroptosis from coronary artery disease (Wu et al., 2022) and validating biomarkers for Alzheimer's Disease (Liu, Li, and Pan, 2021). No universally optimal feature selection method for all conditions exists. Hence, it is important to carefully consider the respective strengths and limitations of the different methods, as well as the specific characteristics of the dataset being analyzed.

### 3.3.3 Differential Expression Analysis

Differential expression analysis stands as a popular and fundamental approach in biomarker discovery, in the analysis of omics data. This approach is favored for its simplicity and interpretability, as it focuses on identifying genes, proteins, or other molecular entities that exhibit significant differences in expression between various conditions or classes. Hence, differential expression analysis is also known as comparative analysis. As an approach, differential expression analysis requires identification of an appropriate feature selection method. The ANOVA f-test, Student's

t-test, DESeq2 for RNA-Seq data, and limma are popular methods used for differential expression analysis.

Commonly used differential expression tests like the Student's t-test, assume adherence to normality (i.e., Gaussian distribution) in the data. However, this assumption is often not met in gene expression and other omics datasets (Lyons-Weiler, Patel, and Bhattacharya, 2003). The selection of an appropriate p-value threshold is crucial for the interpretability of the test's results (Dalman et al., 2012). Recent debates emphasize the risks of misusing p-values, affecting the reliability of conclusions drawn from these analyses (Amrhein, Greenland, and McShane, 2019; Halsey et al., 2015).

Another fundamental assumption that is often violated in omics data is feature independence. This is a significant limitation, as genes, proteins, and other molecules function in interconnected networks and pathways (Dix et al., 2016). The reality of interconnected gene and protein networks challenge the assumption of feature independence, necessitating more complex models or approaches to accurately select the best biomarkers. The erroneous assumption of independence amongst individual genes or proteins also means that multiple-testing corrections (MTCs) overcorrect, thus, lowering sensitivity (Goh and Wong, 2019).

Differential expression analysis is a key technique in the discovery of biomarkers from omics data, offering straightforward interpretability and ease of application. However, challenges such as non-Gaussian data distributions, the misuse of p-values, and the assumption of feature independence require careful consideration. Addressing these challenges is crucial for leveraging this approach effectively, in the pursuit of reliable and meaningful biomarkers in omics research.

### 3.3.4 Evaluating Biomarker Discovery Methods

The process of biomarker discovery involves identifying a minimal yet impactful set of features (biomarkers) that effectively differentiate between two phenotypic classes, such as diseased and healthy individuals. These selected sets, known as 'signatures', must not only be statistically robust but also biologically relevant and reproducible across different data sets (Goh and Wong, 2016). Each of these key aspects is described below in detail.

#### Feature Stability

Feature stability refers to the consistency with which a feature selection method identifies the same features as important across different datasets or under varying conditions. High stability is crucial for the reproducibility of discovered biomarkers. Methods to assess feature stability include:

- **Cross-Validation Techniques**: Using techniques like k-fold cross-validation to assess whether the same features are consistently selected across different subsets of the data.
- **Perturbation Analysis**: Introducing small perturbations in the data (e.g., adding noise) and observing the impact on feature selection, which can reveal the sensitivity of the method to data variability.
- **Bootstrap Resampling**: Analyzing the selection frequency of features across multiple bootstrap samples to determine the robustness of the feature selection process.

For measuring stability, the Jaccard Index is particularly useful, as it compares the similarity between two sets of selected features (top $n$ features), disregarding their internal ranking. The Jaccard formula is:

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{3.2}$$

Here, $A$ and $B$ represent different sets of selected features. When comparing more than two feature sets, the overall stability is computed by aggregating the pairwise Jaccard Indices for all combinations of feature sets:

$$Stability = \sum_{i=1}^{N} \sum_{j=i+1}^{N} Jaccard(R_i, R_j) \tag{3.3}$$

In this formula, $N$ represents the total number of feature sets being compared, and $R_i$ is the $i^{th}$ feature set.

#### Statistical Performance

True biomarkers are intrinsically associated with the characteristics of the phenotypic classes under consideration (e.g., diseased vs. healthy). Therefore, a key aspect of evaluating biomarker discovery methods is assessing the ability of the biomarker signatures identified by the methods in discriminating these classes effectively. This predictive performance is measured using statistical metrics such as classification accuracy, sensitivity, specificity, and area under the ROC curve (AUC-ROC) on independent test sets not used during the discovery phase. High accuracy in these metrics indicates that the biomarkers are indeed capturing essential characteristics of either class and are thus valuable in distinguishing between them.

#### Biological Relevance

The biological relevance of discovered biomarkers is essential to ensure that they have practical significance in understanding and treating diseases. This involves:

- **Enrichment Analysis**: Utilizing computational methods to assess whether identified biomarkers are over-represented in certain biological categories or pathways, more than would be expected by chance. This helps in finding potential associations between biological functions, pathways, or categories that are significantly 'enriched' and the biological or disease processes under study.
- **Pathway and Network Analysis**: Integrating biomarkers into known biological pathways and networks to understand how biomarkers interact within that network. This reveals their role in different biological processes and disease mechanisms.
- **Functional Validation**: Conducting experimental studies, such as knockdown or overexpression experiments, to validate the functional role of the biomarkers in the biological context.
- **Literature Corroboration**: Reviewing scientific literature to find supporting evidence for the association between the discovered biomarkers and the biological states or conditions they are purported to indicate.

Evaluating biomarker discovery methods is a multifaceted process that includes assessing the stability, statistical performance, and biological relevance of the selected biomarkers. Feature stability ensures reproducibility, while statistical performance, such as accuracy in discriminating between phenotypic classes, ensures the biomarkers' effectiveness. Biological relevance validates the practical significance of the biomarkers in a biological context. Evaluating all three aspects is crucial for ensuring that a biomarker discovery method identifies robust, reliable, and meaningful biomarkers.

## 3.4 Ensemble Feature Selection

### 3.4.1 Concept and Advantages

Ensemble feature selection has emerged as a robust approach in biomarker discovery. This method gains its strength by integrating multiple feature selection techniques, enhancing performance and reliability across diverse datasets.

At its core, ensemble feature selection combines different feature selection strategies, including filter, wrapper, and embedded methods. This integration ensures a more balanced and comprehensive approach to feature selection, reducing the dependence on any single method's specific characteristics.

- **Adaptability to Dataset Variability**: Previous studies by Brahim and Limam, 2013, IJzendoorn et al., 2019, and Seijo-Pardo et al., 2017 underscore the variability in the efficacy of individual feature selection methods across diverse datasets. Ensemble feature selection addresses this issue by combining results from multiple methods, thereby offering a more adaptable solution tailored to the unique aspects of each dataset.
- **Reduced Overfitting Risk**: A significant benefit of ensemble methods is their natural inclination to resist overfitting. The collaborative use of multiple methodologies offers a broader perspective, reducing biases that might arise from the idiosyncrasies of a single dataset.
- **Enhanced Generalizability**: By pooling feature sets from diverse methods, ensemble feature selection achieves greater generalizability and versatility. This is particularly beneficial for consistent identification of biomarkers across various conditions and datasets.
- **Effective in Complex Data Analysis**: The intricate nature of omics data, characterized by complex interactions and dependencies, is aptly handled by the ensemble approach. By integrating diverse perspectives, ensemble feature selection can potentially identify these complex relationships, which single-method strategies might miss.

In essence, ensemble methods empower more balanced, adaptable, and generalized feature selection by consolidating diverse insights on biomarker relevance and interactions within the intricate biology. The effectiveness of ensemble feature selection hinges on the strategic combination of different methods, tailored to the dataset's characteristics and the study's objectives. Aggregation of different feature selection methods can be done through various techniques, detailed in the next section.

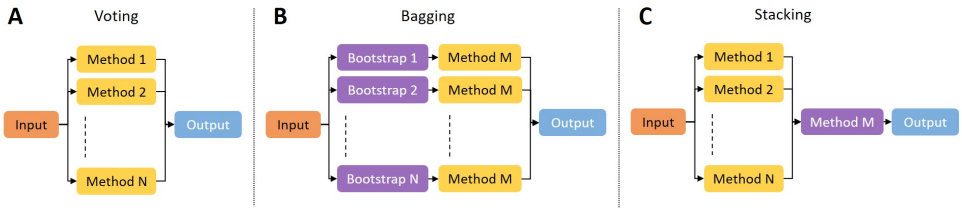


FIGURE 3.1: Ensemble learning techniques for feature selection. (A) Voting. (B) Bagging. (C) Stacking.

### 3.4.2 Types of Ensembling Techniques

Ensemble learning, a concept widely adopted in various domains of data science, involves the strategic combination of multiple algorithms to achieve better predictive performance than any single algorithm could on its own. This section delves into the core types of ensemble learning techniques: Voting, Bagging, and Stacking. While these methods have broad applications across numerous fields, their significance is particularly pronounced in the domain of biomarker discovery to overcome method-dependence. Each technique has its own strengths and specific ways of enhancing data analysis discussed below.

#### Voting

Voting is a simple ensemble approach in which results from different methods, trained on the same data, are aggregated to arrive at a consensus decision. The consensus can be based on majority for classification tasks or by taking the average for regression tasks. The consensus decision can also be weighted, where the output of each method is multiplied by a weight before combining. The approach is illustrated in Fig.3.1A. In the experiments presented in the following sections, rank aggregation (Dwork et al., 2001) has been used to combine the multiple feature rankings. In rank aggregation, the feature importance results from the methods are converted to feature rankings, where a smaller rank is allotted to a feature with higher importance. Then all the feature rankings are simply added up and the resulting array is sorted to obtain the final rankings. The features with smaller values in the summed vector end up with smaller ranks after sorting and are more important.

#### Bagging

Bagging, short for bootstrap aggregation, is an ensemble learning method that involves training multiple instances of the same model, on different subsets of the training data. There exist many approaches for creating the subsets, which involve choosing a subset of samples at random, and may also involve choosing a subset of features at random (Breiman, 1996; Breiman, 1999; Ho, 1998; Louppe and Geurts, 2012). However, for the experiments, the original approach of creating bootstrap samples has been used, which involves choosing random samples with replacement, to form the subsets (Efron and Tibshirani, 1994). This approach has been shown to decrease the variance of the model while maintaining bias (Bauer and Kohavi, 1999), thus, reducing overfitting. The approach is illustrated in Fig.3.1B.

#### Stacking

Stacking is a multi-layer approach in which outputs of the methods in the previous layer act as input to the methods in the next layer. The methods in the first layer are trained on the data whereas the methods in all the succeeding layers are trained on the outputs of the methods in the previous layer. Fig.3.1C illustrates one such stack. Stacking has been shown to outperform any of the single models used in the stack. (Wolpert, 1992).

### 3.4.3 Established Ensemble Biomarker Discovery Methods

In the realm of biomarker discovery, ensemble feature selection methods have gained prominence. Below are some established ensemble feature selection methods used in biomarker discovery:

#### Variable-Selection Oriented Lasso Bagging (VSOLassoBag)

VSOLassoBag is an ensemble method that integrates multiple Lasso regression models within a bagging framework. It creates multiple subsets of data, filters out features based on their correlation with the outcome, and employs cross-validated Lasso regression to rank the remaining features. Features with non-zero coefficients in Lasso regression are considered 'selected'. The process is iterated across all data subsets. Features that are consistently selected across these iterations are deemed more important (Liang et al., 2023).

#### Multi-Criterion Fusion-based Recursive Feature Elimination (MCF-RFE)

MCF-RFE is a wrapper-based ensemble feature selection method. It recursively eliminates features based on a voting ensemble comprising multiple feature selection methods. This method capitalizes on the strengths of various feature selection techniques. However, it can be computationally demanding and may not perform optimally if the underlying feature selection methods are not well-suited to the specific data (Du et al., 2016; Khaire and Dhanalakshmi, 2019).

#### Ensemble Support Vector Machine based Recursive Feature Elimination (ESVM-RFE)

ESVM-RFE is an adaptation of the RFE method. Instead of using a single SVM classifier, it employs a bagging ensemble of SVM classifiers to rank features during each step of the elimination process (Anaissi et al., 2016).

#### Voting Ensemble using Borda Count (E-Borda)

E-Borda addresses the challenge of varying optimal feature selection methods across different datasets or tasks by employing a voting ensemble of multiple feature selection algorithms. The integration of feature rankings from these algorithms is executed using the Borda Count method. This approach, however, may be sensitive to the choice of base feature selection methods (Shi et al., 2020; Drotár, Gazda, and Vokorokos, 2019).

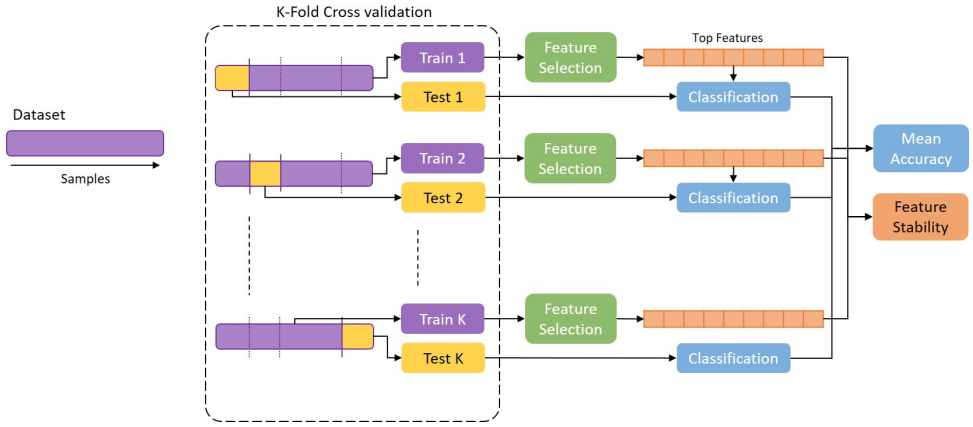


FIGURE 3.2: Evaluation of a feature selection method on a dataset.

In conclusion, these ensemble feature selection methods represent advanced strategies in the field of biomarker discovery. They offer improved performance by combining the advantages of various feature selection techniques, though considerations around computational intensity and the suitability of base methods to specific datasets must be taken into account.

### 3.4.4 Effect of Ensembling on Biomarker Signature

This section analyzes the effects of the different types of ensembling techniques on the resulting set of biomarkers selected. Six feature selection methods - ANOVA, SNR, Random Forests, Lasso Regression, RFE-SVM and RFE-LR were run unmodified (also referred to as vanilla) on the LYRIKS dataset. K-fold cross validation with k set to five, was used to evaluate the performance of the methods. In each fold, the feature selection methods were run on the training set and the identified features were used to train 4 different classifiers viz. Support Vector Machine (SVM) (Cortes and Vapnik, 1995), K-Nearest Neighbours (KNN) (Fix and Hodges, 1989), Multilayer Perceptron (MLP) (Rosenblatt, 1961; Rumelhart, Hinton, and Williams, 1985) and Gradient-Boosted Decision Tress (GBDT) (Friedman, 2002), for classification of UHR and control groups. Four different types of classifiers were used to ensure the selected features perform well across any type of classification method. The accuracy of these classifiers was evaluated on the test set. Fig. 3.3A shows the accuracy of these classifiers with increasing number of selected features.

#### Vanilla Feature Selection Results

The highest accuracies have been achieved by the wrapper methods RFE-SVM and RFE-LR. Lasso performs the next best while Random Forest gives lowest accuracies over the top 20 features. For the filter methods ANOVA and SNR, after the first five features, the accuracy remains nearly constant with increasing number of features.

The stability of these traditional algorithms with increasing number of features is shown in Fig. 3.3B. It is evident that the filter methods ANOVA and SNR are more stable relative to other methods. ANOVA achieves high top 5 feature consistency. RFE-LR achieves higher stability than filter methods after the top 40 features, whereas its companion wrapper method RFE-SVM, achieves much lower stability comparatively. Random Forest performs the worst in terms of stability because of the intrinsic randomness involved in the algorithm (Wang, Yang, and Luo, 2016).

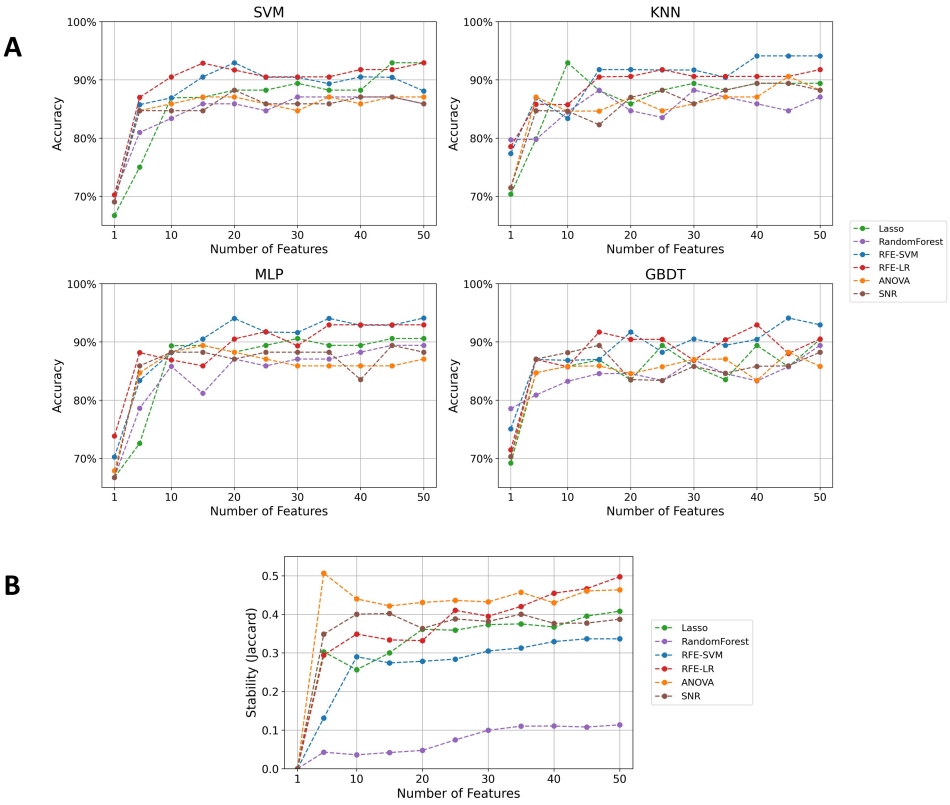


FIGURE 3.3: (A) Comparison of the classification accuracies of different classifiers using the top features selected by various feature selection algorithms. The accuracy represents the average accuracy of the classifier across the K folds of cross-validation. (B) Comparison of stability of the features selected by the various feature selection algorithms across the K folds. The stability score is calculated using Jaccard Index (defined in Eq. 3.2).

In the following, the effect of the three ensemble techniques is analysed (illustrated in Fig. 3.1A) for feature selection.

### Bagging

The first ensemble technique explored in the study is bagging. In bagging, in each fold of the cross-validation, 10 bootstrap samples were created, each bootstrap sample containing a subset of available samples in the training set, chosen at random with replacement. The feature selection algorithms were run on these bootstrap samples, and the feature rankings generated by each algorithm were aggregated to obtain one final ranking per algorithm.

The mean accuracies of the feature selection algorithms across the 4 classifiers viz. SVM, KNN, MLP and GBDT, with and without bagging, are compared in Fig. 3.4A. The mean accuracy does not change much with bagging, across all types of feature selection methods. Fig. 3.4B records the stability of the features selected by the algorithms. Significant improvement in stability can be seen for algorithms that involve some randomness in initialization like Random Forests. The stability does not improve for filter methods SNR and ANOVA partially because there are no hyperparameters whose values affect the calculation of feature importance.

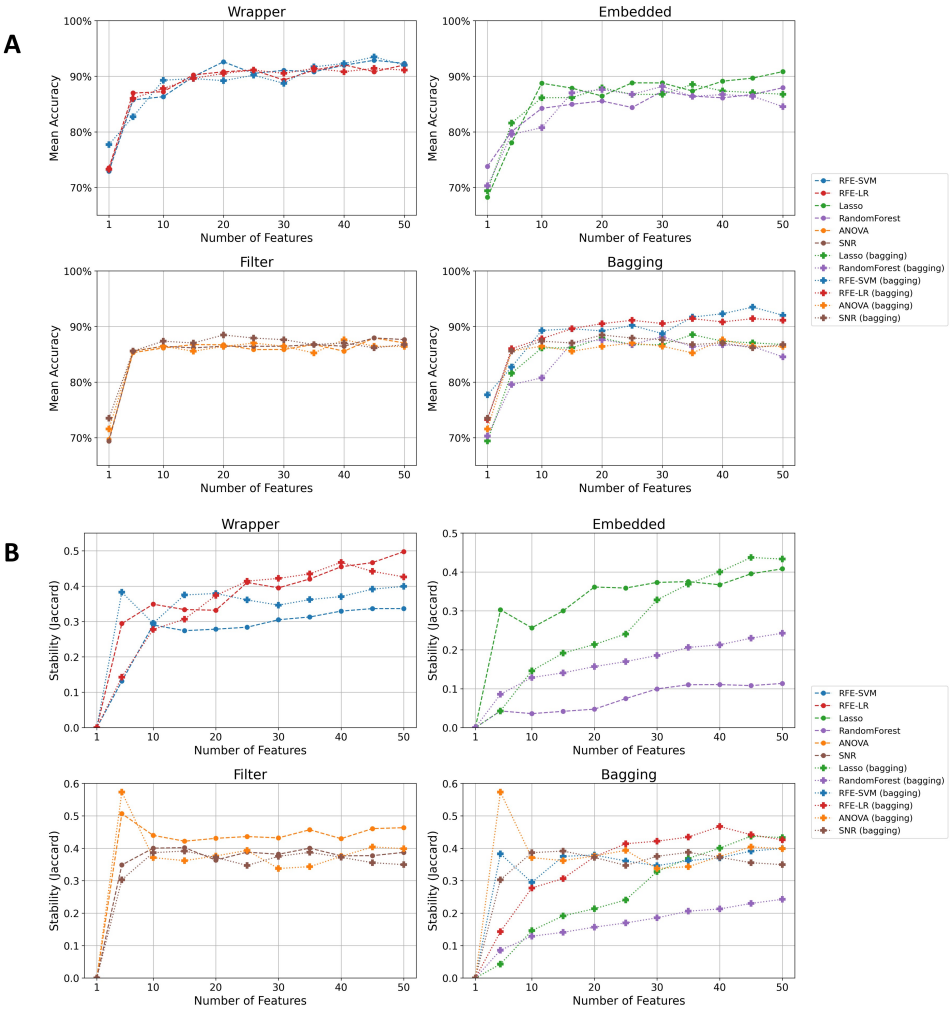


FIGURE 3.4: Comparison of feature selection methods and their bagging variants. (A) The accuracy of all wrapper, embedded, and filter methods displays negligible changes after implementing bagging. (B) Bagging ensembles substantially enhance stability, especially for methods where initialization involves an element of randomness.

### Voting

In Voting, rankings from different algorithms are combined using rank aggregation to obtain one ensemble ranking. To analyse the effect of voting, the vanilla rankings of SNR and ANOVA were combined into a filter ensemble, RFE-SVM and RFE-LR into a wrapper ensemble, Lasso and Random Forest into an embedded ensemble and all the six algorithms in one ensemble referred to as 'All'. Fig. 3.5A and Fig. 3.5B show the mean accuracies and stability scores of these ensembles compared with the vanilla algorithms used to create the ensemble.

From the results, it can be seen that in all cases, the accuracy and stability of an ensemble ranking are nearly an average of the vanilla rankings used to create them. For example, the stability results of the filter ensemble lie midway between the stability of ANOVA and SNR and similarly, the mean accuracy of the filter ensemble lies between the accuracy of ANOVA and SNR.

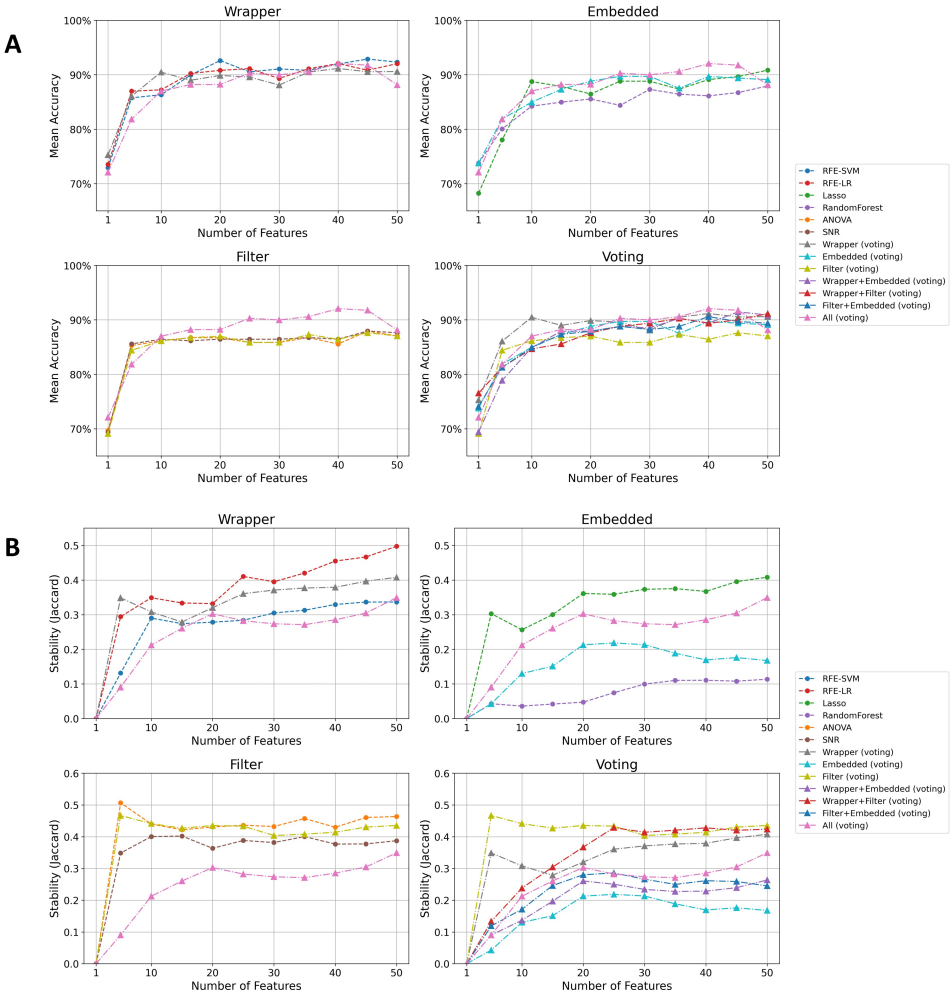


FIGURE 3.5: Comparison of feature selection methods to their voting ensembles. (A) The accuracy of voting ensembles tends to reflect the mean performance of the individual methods included in the ensemble. (B) Similarly, the stability of voting ensembles typically approximates the average stability of the individual algorithms within the ensemble.

### Stacking

Stacking is an ensembling technique in which the output of one algorithm is provided as input to another algorithm. In the feature selection context, this means the features selected by the first algorithm are passed to the second algorithm to further select a smaller subset of important features. In the analysis, one feature selection technique was used from each category, viz. ANOVA from filter, RFE-SVM from wrapper and Lasso from embedded and basic stacking ensembles were explored that can be created using two out of the three algorithms. The first algorithm was used to select the top 50% of the features and the second algorithm ranked the remaining features to achieve the final ranking.

Fig. 3.6A shows the mean accuracy of the stacking ensembles compared to the vanilla algorithms used in them. The only stack that significantly outperforms its vanilla algorithms is the ANOVA+RFE-SVM stack, where ANOVA is the first algorithm that is used to select the top 50% of features and RFE-SVM is the second and final algorithm used to rank the selected top 50% features. RFE-SVM is the closest to

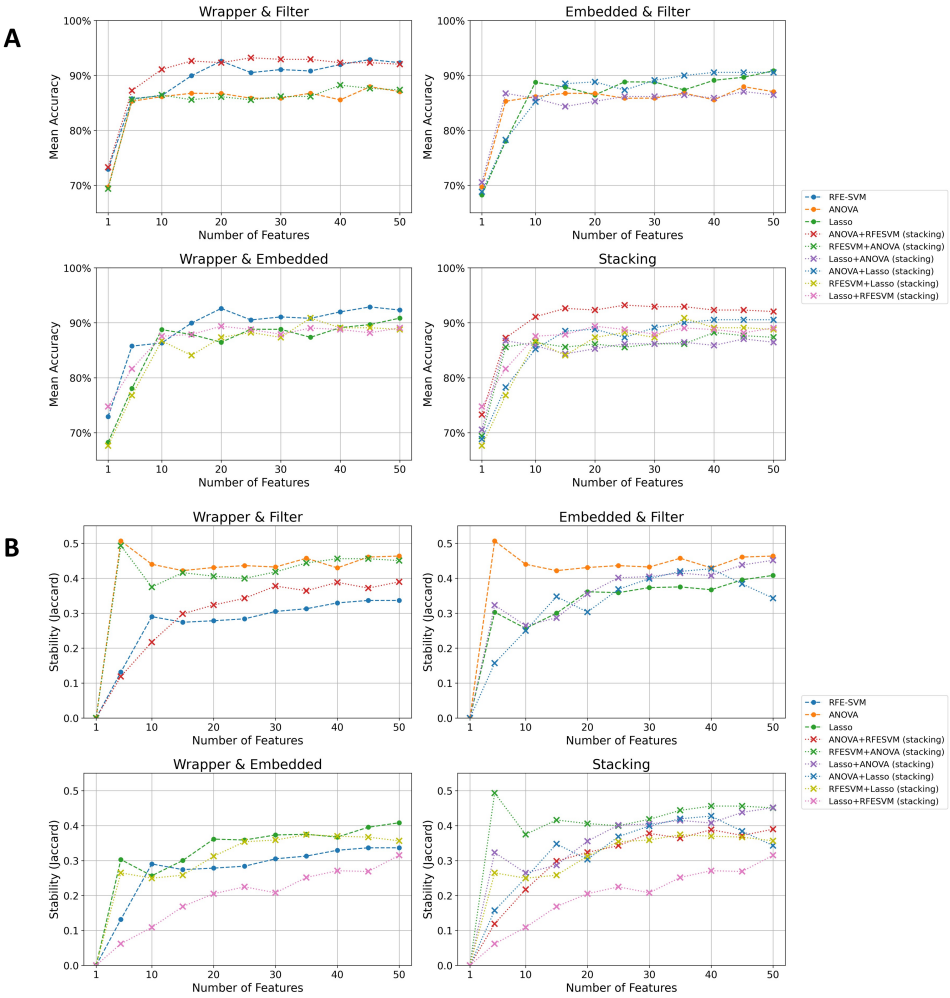


FIGURE 3.6: Comparison of feature selection methods and their stacking combinations. (A) The combination of Filter and Wrapper methods (ANOVA+RFE-SVM stack) demonstrates enhanced accuracy relative to their individual performances. (B) Furthermore, the ANOVA+RFE-SVM stack exhibits superior stability compared to the standalone Wrapper method RFE-SVM.

the stack in terms of mean accuracy, but as Fig. 3.6B illustrates, the ANOVA+RFE-SVM stack achieves higher stability than vanilla RFE-SVM.

Based on the insights gathered from these ensembling analyses, the FWSE algorithm was designed.

## 3.5 FWSE: A Novel Paradigm for Biomarker Discovery

### 3.5.1 Concept and Design

The Filter and Wrapper Stacking Ensemble (FWSE) is an innovative approach in biomarker discovery, strategically designed to balance accuracy and stability. This method, as depicted in Figure 3.7, integrates several key steps to enhance the reliability and effectiveness of the features selected.

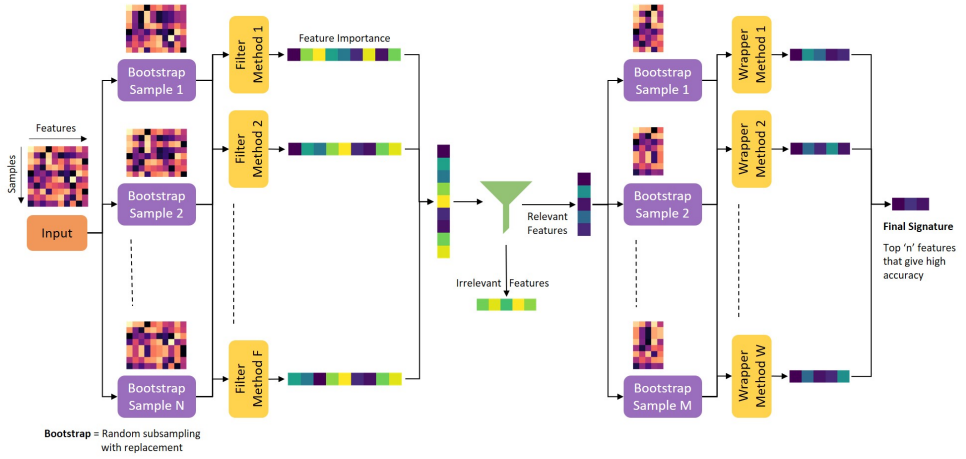


FIGURE 3.7: The architecture of FWSE.

- **Generating Bootstrap Samples**: FWSE begins by creating multiple bootstrap samples from the dataset. This technique captures varied data representations, crucial for a robust feature selection process.
- **Employing Filter Methods**: On these bootstrap samples, filter feature selection methods are applied. Known for their efficiency in dealing with high-dimensional datasets, these methods effectively remove irrelevant or redundant features.
- **Unified Feature Ranking via Aggregation**: The output feature rankings of the filter methods are then merged through rank aggregation. This essential step combines insights from various filter methods to produce a unified feature ranking.
- **Feature Pruning with the Pruning Factor**: A key element of FWSE is the 'pruning factor', a parameter that governs the elimination of the least significant features. It is set to 0.5 for the scope of the analyses presented in this section. This means 50% of the features are eliminated, thereby reducing the dimensionality of the feature space and enhancing the efficiency of subsequent steps.
- **Creating New Samples with Pruned Features**: Post-pruning, new bootstrap samples are generated using the reduced feature set. The subsequent steps are focused on the most relevant features, enhancing the probability of identifying accurate biomarkers.
- **Applying Wrapper Methods for Ranking**: Next, wrapper methods are used on these new bootstrap samples. These methods are adept at discerning statistically significant feature groups, assessing subsets based on their predictive capacity and inter-feature interactions.
- **Final Feature Ranking via Aggregation**: The final step involves aggregating the rankings from the wrapper methods to form the final feature ranking, highlighting the most promising biomarkers identified through this comprehensive and multifaceted procedure.

FWSE is grounded in the empirical findings related to the effect of ensembling techniques on biomarker discovery, discussed in Section 3.4.4. It utilizes bootstrap

sampling and multiple filter/wrapper feature selection algorithms to enhance stability, and the stacked ensemble approach of filter and wrapper methods to improve accuracy. This method is a deliberate integration of various feature selection paradigms, capitalizing on the strengths of both filter and wrapper methods while addressing their individual limitations. The incorporation of the pruning factor adds a strategic layer of control, refining the selection process. Detailed analysis of the impact of varying the pruning factor on the model's accuracy and stability can be found in Appendix A.

### 3.5.2 Datasets for Experiments

TABLE 3.1: Summary of datasets used to evaluate FWSE

| Dataset | Profiling Technique | # of features | # of samples | Clinical Factor | Case | Control | Reference |
|---|---|---|---|---|---|---|---|
| LYRIKS | Microarray | 34928 | 84 | Total<br><br>Age<br>Female sex | 56 (66.7%)<br>22.1<br>14 (25.0%) | 28 (33.3%)<br>22.5<br>7 (25.0%) | Lee et al., 2013 |
| Bipolar | RNASeq | 20581 | 480 | Total<br><br>Age<br>Female sex | 240 (50%)<br>50.3<br>131 (54.6%) | 240 (50%)<br>43.4<br>119 (49.6%) | Krebs et al., 2020 [↓] |
| LUAD | RNASeq | 20531 | 569 | Total<br><br>Age<br>Female sex | 510 (89.6%)<br>65.3<br>274 (53.9%) | 59 (10.4%)<br>66.0<br>32 (56.1%) | Liu et al., 2018a [↓] |
| PDAC | Proteome | 11662 | 215 | Total<br><br>Age<br>Female sex | 140 (65.1%)<br>64.3<br>66 (47.1%) | 75 (34.9%)<br>64.5<br>33 (49.2%) | Cao et al., 2021 [↓] |

In the evaluation of the FWSE method for biomarker discovery, four distinct datasets were employed, each offering a unique perspective in the biomedical field. These datasets enabled a comprehensive comparison of FWSE with both traditional feature selection methods and established ensemble feature selection methods.

The first dataset used was the Longitudinal Youth at Risk Study (LYRIKS), sourced from the Institute for Mental Health (IMH) in Singapore (Lee et al., 2013). This dataset includes gene expression values for 84 participants, comprising 56 individuals identified as Ultra-High Risk (UHR) for psychosis (Yung et al., 2003) and 28 control participants, with a total of 34,928 genes analyzed. The UHR participants were assessed using the Comprehensive Assessment of At Risk Mental States (CAARMS) (Yung et al., 2005). Gene expression profiling in this dataset was conducted using peripheral blood samples and analyzed on Illumina HumanHT-12 v4 Expression

BeadChip arrays. Notably, a previous study by Goh et al., 2017 identified a 12-gene signature in this dataset that achieved 90% accuracy in identifying individuals at UHR.

The second dataset focused on Bipolar Disorder and was obtained from the University Medical Center Utrecht in the Netherlands (Krebs et al., 2020). It comprises 20,581 gene expression values from 480 participants, equally divided between individuals diagnosed with bipolar disorder and control subjects. Bipolar Disorder is characterized by significant mood instability and has a high level of heritability (Grande et al., 2016). Diagnoses were confirmed using the Structured Clinical Interview for DSM-IV (SCID) (First and Gibbon, 2004). The gene expression values were derived from RNA sequencing of peripheral whole blood.

The Lung Adenocarcinoma (LUAD) dataset, part of The Cancer Genome Atlas (TCGA) PanCancer Atlas study, was also utilized (Liu et al., 2018a). This dataset contains 20,531 gene expression values from 569 participants, including 510 individuals with cancerous conditions and 59 controls. Lung adenocarcinoma represents a major form of non-small cell lung cancer and is especially prevalent among non-smokers. It remains one of the leading causes of cancer-related deaths worldwide (Sung et al., 2021). The gene expression data for this study were obtained using the Illumina HiSeq platform.

Lastly, the Pancreatic Ductal Adenocarcinoma (PDAC) dataset was included, featuring data in accordance with the Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines (Cao et al., 2021). This dataset is composed of eight types of omics data from 140 pancreatic tumor tissues, 67 paired normal adjacent tissues, and 9 normal pancreatic duct tissues. PDAC is known for its aggressive nature and poor survival rates, often being diagnosed at advanced stages. The dataset focused on proteomic expression, quantifying 11,662 proteins. PDAC is projected to become a leading cause of cancer death by 2030, highlighting the importance of effective biomarker discovery in this area (Quante et al., 2016).

The inclusion of cancer datasets, despite the primary focus of the thesis being on mental health, is strategic as it can also be used to demonstrate the generalizability of the FWSE approach. Moreover, cancer biomarkers are more extensively researched, offering a rich field for benchmarking the performance of biomarker discovery methods. This approach facilitates a more straightforward assessment of the biological relevance of the selected markers, providing a robust context for evaluating and validating the FWSE method's effectiveness.

### 3.5.3 Comparative Analysis of FWSE's Accuracy and Stability

This section presents a comparative analysis of the accuracy and stability of the Filter and Wrapper Stacking Ensemble (FWSE) method against both traditional feature selection methods and established ensemble feature selection methods. Employing four distinct datasets, this analysis aims to demonstrate the efficacy of FWSE in identifying biomarkers across varied biomedical contexts. The experimental design was the same as used for the ensembling experiments, detailed in 3.4.4.

#### Case Study #1: LYRIKS Data

The LYRIKS dataset, which deals with Ultra-High Risk (UHR) individuals, presents a unique landscape for biomarker discovery. For this dataset, in FWSE, we employed ANOVA and SNR as filter methods, and RFE-SVM and RFE-LR as wrapper

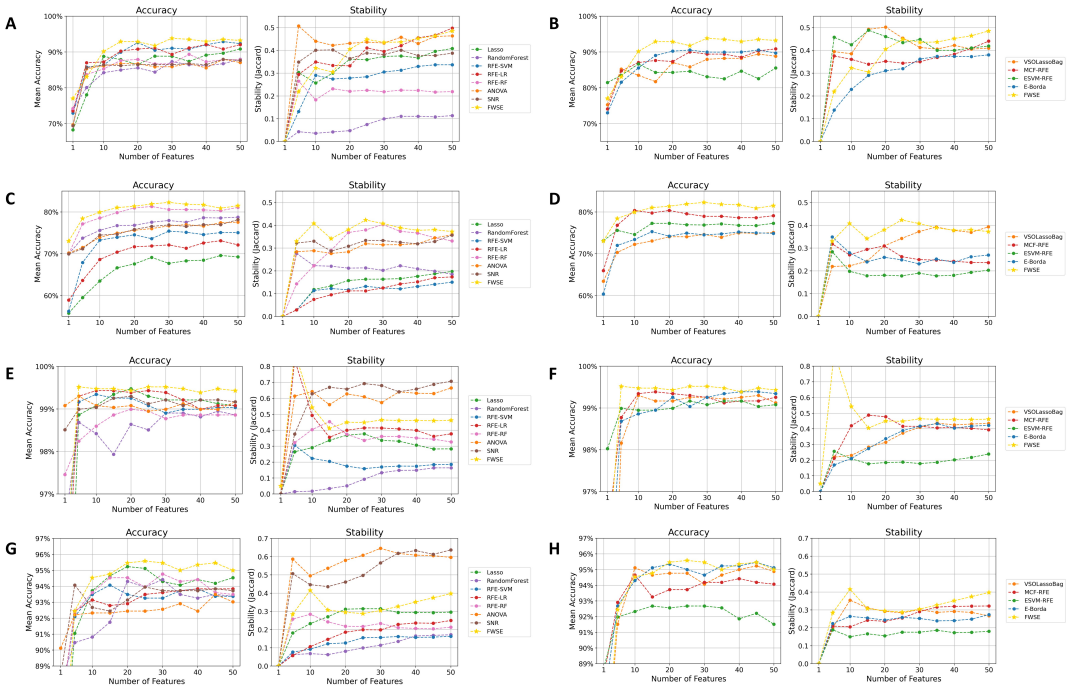


FIGURE 3.8: Comparison of mean accuracy and stability of FWSE to other feature selection algorithms. (A) Comparison against traditional feature selection methods on the LYRIKS data. (B) Comparison against ensemble feature selection methods on the LYRIKS data. (C) Comparison against traditional feature selection methods on the Bipolar data. (D) Comparison against ensemble feature selection methods on the Bipolar data. (E) Comparison against traditional feature selection methods on the LUAD data. (F) Comparison against ensemble feature selection methods on the LUAD data. (G) Comparison against traditional feature selection methods on the PDAC data. (H) Comparison against ensemble feature selection methods on the PDAC data.

methods, creating 10 bootstrap subsets at both stages, each subset equal in size to the original dataset.

As depicted in Fig. 3.8A, FWSE outperforms traditional feature selection algorithms in terms of accuracy. The filter methods in FWSE effectively eliminate irrelevant features, while the wrapper methods rank the remaining relevant features to identify a minimal subset that maximizes accuracy. The stability of FWSE is also comparable to the most stable traditional algorithms, demonstrating the robustness of our ensemble approach. The usage of bagging and multiple filter/wrapper methods greatly improves the stability of the proposed ensemble compared to a simple one filter and one wrapper method stack. ANOVA and RFE-LR are closest to FWSE's stability but FWSE heavily outperforms them in accuracy, achieving $\approx 93\%$ mean accuracy in top 15 features and $\approx 95\%$ mean accuracy in top 30 features. RFE-SVM is the closest to FWSE in terms of accuracy, but the ensemble significantly outperforms RFE-SVM in stability.

When compared against ensemble biomarker discovery methods (Fig 3.8B), FWSE exhibits superior accuracy and stability on the LYRIKS dataset. While E-Borda comes close in terms of accuracy, FWSE achieves much higher stability, demonstrating its robustness in the face of data perturbations. Similarly, ESVM-RFE and VSOLassoBag initially give better stability than FWSE when the number of selected features is less

than 30, but FWSE significantly outperforms both in terms of accuracy ($\approx$ 8% on average).

#### Case Study #2: Bipolar Data

The Bipolar dataset, with its focus on a complex mental disorder, presents a challenging testbed for our proposed FWSE method due to low separability between the target classes. For this dataset, ANOVA and SNR were employed as filter methods and RFE-RF was employed as the wrapper method in FWSE.

As shown in Fig. 3.8C, FWSE outperforms traditional methods in terms of both accuracy and stability. RFE-RF comes close in terms of accuracy, but FWSE demonstrates higher stability, especially when the number of selected genes is less than 20.

When compared to ensemble biomarker discovery methods (Fig 3.8D), FWSE again outperforms the competition in terms of both accuracy and stability, demonstrating its robustness across different datasets and conditions. MCF-RFE is closest to FWSE in terms of accuracy and VSOLassoBag is closest in terms of stability but when accuracy and stability are considered together, FWSE significantly outperforms all the ensemble biomarker discovery methods. This demonstrates the robustness of FWSE, even in the face of data perturbations on a dataset with low separability, making it a promising tool for biomarker discovery in mental health research.

#### Case Study #3: LUAD Data

The LUAD dataset, which focuses on Lung Adenocarcinoma, presents a different set of challenges due to very high separability. Based on the performance of the traditional feature selection methods, in FWSE, ANOVA and SNR were employed as the filter methods and RFE-SVM and RFE-LR were employed as the wrapper methods.

As shown in Fig. 3.8E, FWSE achieves the highest accuracy among traditional feature selection methods, although its stability is lower than filter methods ANOVA and SNR. This is likely due to the high separability of the LUAD dataset, which contains many groups of genes that can achieve similar high levels of separability. FWSE does outperform other wrapper and embedded methods in terms of stability. RFE-LR has slightly lower accuracy and stability than FWSE. Lasso comes close in terms of accuracy only around the top 20 features mark but has lower stability than FWSE throughout.

When compared to ensemble biomarker discovery methods, FWSE outperforms all other approaches in terms of both accuracy and stability (Fig. 3.8F). MCF-RFE, while slightly lower in accuracy and slightly higher in stability around top 15-20 features, eventually falls behind in both accuracy and stability. E-borda also comes close to FWSE after the top 35 features, but FWSE maintains higher accuracy and stability.

#### Case Study #4: PDAC Data

The PDAC dataset also exhibits good separability, like the LUAD dataset. For PDAC, in FWSE, ANOVA and SNR were employed as the filter methods, and RFE-RF and RFE-LR (with L1 norm) were employed as the wrapper methods. Just like in previous case studies, 10 bootstrap subsets were created at each stage, each subset equal in size to the original dataset.

As depicted in Fig. 3.8G, FWSE stands out as the most accurate among traditional machine learning feature selection methods. While Lasso displays slightly higher stability in the top 20-30 features, FWSE surpasses it considerably in terms of accuracy. The stability of SNR and ANOVA on this dataset further confirm this trend that in datasets with high separability, filter methods tend to achieve the highest stability. However, FWSE outperforms these methods in accuracy, making it a superior choice when prioritizing accuracy.

When pitted against ensemble feature selection methods, FWSE again outperforms in terms of accuracy and stability. E-borda demonstrates comparable accuracy but falls significantly short in stability. VSOLassoBag exhibits equivalent stability in the top 15-30 features but does not match FWSE's accuracy. ESVM-RFE shows subpar performance in both accuracy and stability.

These results underscore the versatility of FWSE across different datasets and conditions. An interesting trend to note is the relationship between class separability and the stability of FWSE. In the Bipolar dataset, where class separability is the lowest, FWSE surpasses all other methods in terms of stability, demonstrating its resilience in challenging conditions. In the LYRIKS dataset, FWSE's stability is on par with the top-performing approach, further attesting to its robustness. However, in the LUAD and PDAC datasets, where class separability is very high, the stability of FWSE is slightly lower than that of filter methods. This is potentially because, in high separability scenarios, many groups of genes can achieve the same level of accuracy. Nonetheless, the high accuracy and overall performance of FWSE across all datasets underscores its potential as a reliable and versatile tool for biomarker discovery from high-dimensional omics data.

TABLE 3.2: Potential biomarkers identified for the different datasets using FWSE

| **Dataset** | **Potential Biomarkers** |
|---|---|
| LYRIKS | LDLRAD1, CYP8B1, CDH11, TMEM225, HS.170946, LOC100134138, HS.537754, C2CD3, LRRTM2, OR4A15, LOC100127888, LOC100129002, LOC100131961, PLA2G5, LOC389118, LOC100134413, LOC641964, SVOPL, LOC100133959, SNORD113-6, OR56B4, LOC653113, SRD5A2, HS.146184, CNTNAP5 |
| Bipolar | TSPAN2, TAGLN2, FAR2, CXCL8, PFKFB2, LINC01765, CD300A, TAGLN2P1, MIR23AHG, FLT3, PIGB, SLC31A2, IVNS1ABP, AKAP12, CNTNAP3, LINC00877, YIPF4, LILRA4, RFX2, SYTL3, PGM5, CFAP45, MAK, DNASE1L3, CASP10 |
| LUAD | ALDH18A1, SRPK1, SMYD5, KIAA0907, MYO7A, CSTF2, SERINC2, ZNF207, ABCC3, GMPPA, C1orf63, C1orf131, UBFD1, DLG3, P4HB, GYG2, SKIV2L, TXNDC5, PVRL4, NEK6 |
| PDAC | C19orf33, GLA, S100A14, MISP, SCEL, IGF2BP3, SDCBP2, GALNT7, HEPH, SFN, SULF1, SAMD9, SERPINB5, PGM2L1, LMO7, MDK, REG4, STON1, HK2, GSDMB, ARPC1B, MYO1E, SDR16C5, S100A16, ACTN1 |

### 3.5.4 Biological significance of Identified Biomarkers

Beyond the statistical performance and feature stability, the biological relevance of the biomarkers identified by FWSE was further evaluated. A literature survey was conducted to confirm that the biomarkers identified by FWSE in benchmark cancer datasets aligned with previous findings in the field. A Gene Ontology (GO) enrichment analysis was also employed to interpret and understand the overarching biological processes associated with the biomarker signatures identified by FWSE.

The resulting features selected from each dataset using FWSE are summarized in Table 3.2. For each dataset, the features are listed in descending order of their contribution to model performance, as determined during the iterative backward-selection process. Features appearing earlier in each list yielded greater improvements in the evaluation metric when added, reflecting their relative predictive importance within that dataset.

#### Literature Survey

The genes selected on LUAD, including ALDH18A1, CSTF2, MYO7A, SMYD5, SRPK1, C1orf63, GMPPA, ZNF207, and ABCC3, have demonstrated associations with lung cancer and other cancers in the literature. For instance, ALDH18A1 has shown overexpression in lung cancer (Ye et al., 2021). CSTF2 has been identified as an independent prognostic factor in non-small cell lung cancer (NSCLC) and its suppression has been linked to inhibited lung cancer cell growth (Aragaki et al., 2011). MYO7A, although primarily studied in the context of melanoma, has demonstrated roles in cell growth and migration, suggesting potential significance in lung cancer (Liu et al., 2018b). SMYD5 and SRPK1, when depleted, have been associated with increased tumor growth (Kidder et al., 2017; Liu et al., 2016). C1orf63 has been found overexpressed in several cancers including lung cancer (Hong et al., 2015). Mutations in GMPPA have shown a significant association with patient mortality in LUAD (Cho et al., 2018). SERINC2 plays a critical role in LUAD, with SERINC2 knockdown shown to inhibit proliferation, migration, and invasion in this cancer type (Zeng et al., 2018). ABCC3 has been identified as a marker for multiple drug resistance and predictor for poor clinical outcome in NSCLC, indicating its critical role in lung cancer pathology (Zhao et al., 2013).

On PDAC, FWSE identified some proteins that could be key potential biomarkers such as S100A14, MISP, SFN, SULF1, SAMD9, and SERPINB5. S100A14 is not only an indicator of PDAC progression but also contributes to gemcitabine resistance, making it a potential therapeutic target (Zhu et al., 2021). MISP's upregulation is linked to poor patient outcomes and is instrumental in immune system alterations in PDAC (Huang et al., 2022). SFN has been validated as a stromal marker with prognostic significance, specifically affecting both overall and disease-free survival (Robin et al., 2020). SERPINB5 stands out for its capacity to differentiate PDAC from pancreatitis, owing to its promoter hypomethylation (Mardin et al., 2016). SAMD9 and SULF1, although not as extensively studied, also show associations with PDAC pathology and are candidates for future investigations (Tan, Muckadell, and Joergensen, 2020). IGF2BP3 plays a role in PDAC malignancy by affecting cell invasiveness and modulating miRNA-mRNA interactions (Ennajdaoui et al., 2016).

#### Gene Ontology (GO) Enrichment Analysis

Gene Ontology is a standardized vocabulary (ontology) that describes genes and gene products in terms of their associated biological processes (such as "DNA repair"

or "cellular respiration"), cellular components (such as "nucleus", "mitochondrion", or "membrane"), and molecular functions (such as "DNA binding" or "enzyme activity") in a species-independent manner. In the context of biomarker discovery, GO enrichment analysis involves determining whether certain GO terms are over-represented in the set of identified biomarkers compared to a reference set. The reference set used in this analysis was the entire set of known genes or gene products for the human species (Homo sapiens).

The GO enrichment analysis helps in understanding the biological implications of the biomarkers. For instance, if a significant number of identified biomarkers are associated with the GO term "immune response", it suggests that these biomarkers are collectively involved in or related to immune system processes. The enrichment analysis provides a statistical measure to determine if the observed over-representation is non-random. A lower p-value in this context indicates that the association between the biomarkers and the GO term is strong and not due to chance. In this analysis, p-values less than 0.01 were considered significant.

- **LYRIKS dataset**: The LYRIKS signature is enriched in GO terms related to integral components of the membrane, plasma membrane, oxygen binding, heme binding, and endoplasmic reticulum membrane. The plasma membrane and endoplasmic reticulum are integral to the immune response, facilitating the recognition of antigens and the production of immune-related proteins. This aligns with the widely reported observation of immune dysfunction in individuals at Ultra-High Risk (UHR) for psychosis (Radhakrishnan, Kaser, and Guloksuz, 2017).

- **Bipolar dataset**: In the context of bipolar disorder, FWSE pinpointed genes enriched in GO terms such as carbohydrate binding, inflammatory response, apoptotic process, and cellular response to lipopolysaccharide. These terms are closely tied to immune function as well. For instance, carbohydrate binding is involved in cell-cell recognition, a crucial aspect of immune response, while inflammatory response and cellular response to lipopolysaccharide are directly linked to immune activation. This is consistent with the growing body of evidence suggesting a role for immune dysfunction in the pathophysiology of bipolar disorder (Rosenblat and McIntyre, 2017).

- **LUAD dataset**: The biomarker signature identified in the LUAD dataset exhibited enrichment in GO terms related to the endoplasmic reticulum membrane, ER to Golgi vesicle-mediated transport, mitochondrion, and ATP binding. The ER is involved in protein folding and transport, lipid metabolism, and calcium homeostasis, disruptions in which can lead to ER stress, a condition implicated in various diseases, including cancer (Hu et al., 2019). The ER also plays a role in vesicle-mediated transport to the Golgi apparatus, a pathway crucial for protein secretion (Lee et al., 2020). Mitochondria, known for their role in energy production through ATP synthesis, also play key roles in apoptosis and reactive oxygen species (ROS) production, critical processes in cancer development (Ghemrawi and Khair, 2020).

- **PDAC dataset**: Lastly, the biomarker signature on the PDAC dataset was enriched in GO terms associated with actin binding, calcium ion binding, extracellular space, and cytosol. Actin binding is relevant for cellular structure and motility, potentially contributing to cancer cell invasiveness. Calcium ion binding is involved in various cellular processes, including signal transduction pathways that could be altered in cancerous cells. The extracellular space

is key for cell-to-cell communication, often dysregulated in cancer, and the cytosol is involved in numerous metabolic and signaling pathways. These terms complement previous findings on GO terms related to extracellular structure and binding properties being associated with PDAC (Tan, Muckadell, and Joergensen, 2020).

In summary, the identified biomarker signature's enrichment in these GO terms aligns with existing literature on UHR, Bipolar, LUAD, and PDAC, further validating the biological relevance of the identified biomarkers. Altogether, these findings underscore the potential of FWSE in identifying biologically relevant biomarkers across diverse disease contexts and emphasize the importance of considering the collective action of these biomarkers in disease pathology.

## 3.6 Conclusion

In biomedical research involving the analysis of high-dimensional omics data, the 'curse of dimensionality' presents a formidable challenge. Traditional statistical tests, commonly employed for their simplicity in identifying differentially expressed genes, often operate under the assumption of feature independence. However, this assumption is frequently invalidated by the interconnected nature of genes functioning in complex networks and pathways. The inadequacy of these methods to capture the intricate dependencies among features significantly undermines their effectiveness in biomarker discovery.

Ensemble methods have been recognized for their ability to overcome the inherent limitations of individual feature selection strategies, offering a more robust and consensus-driven approach. Through a comparative analysis of filter, wrapper, and embedded methods, as well as their combinations using popular ensemble techniques like voting, bagging, and stacking, this chapter highlights the distinct advantages of these approaches. Filter methods provide stable feature selection, less prone to overfitting, but often at the cost of predictive accuracy. Conversely, wrapper and embedded methods, while achieving higher accuracy with their respective classifiers, can often succumb to overfitting, as reflected by their lower stability in certain contexts.

The central component of this chapter is the introduction of the Filter and Wrapper Stacking Ensemble (FWSE) method. This innovative approach synergistically combines the strengths of filter and wrapper methods, addressing both the accuracy and stability issues prevalent in traditional biomarker discovery techniques. The use of bootstrapping within FWSE enhances the stability and counteracts the tendency for overfitting. The stacking strategy employed allows for the effective elimination of non-differentially expressed genes through filter methods, followed by a collective evaluation of the remaining features by wrapper methods. This results in a robust set of genes that provide high group separability, even though each gene may be a weak biomarker individually.

The superior performance of the FWSE method has been demonstrated across various datasets, including LYRIKS, LUAD, and PDAC. Not only does FWSE excel in terms of accuracy and stability, but it also ensures the biological relevance of the identified biomarkers. The biomarkers discovered using FWSE for the LYRIKS data, for instance, outperform previous works in predicting UHR criteria, and those identified for LUAD and PDAC are significantly associated with the respective cancers. This alignment with known biological processes and disease mechanisms, validated

through Gene Ontology (GO) enrichment analysis, underscores the practical significance of FWSE in uncovering meaningful biomarkers.

Despite its strengths, FWSE is not without limitations. The method is computationally intensive, primarily due to the repeated application of wrapper feature selection techniques like RFE on multiple bootstrap samples. Future work will be directed towards enhancing the computational efficiency and biological relevance of FWSE, potentially through the incorporation of more constraints into its architecture. Such improvements will further cement the utility of FWSE in the rapidly evolving field of biomarker discovery.

In conclusion, this chapter has systematically analyzed the effects of ensemble feature selection methods and provided a novel and effective approach for identifying stable and biologically relevant biomarkers in high-dimensional medical data. The findings underscore the potential of genetic biomarkers to enhance the understanding of the biological underpinnings of mental health disorders. By integrating these genetic markers with other data types, such as clinical assessments and neuroimaging, the FWSE approach contributes to creating a more holistic and robust diagnostic and prognostic model. The insights gained from this research will inform the subsequent chapters, where these biomarkers will be further integrated and validated using real-world datasets. By enhancing the predictive power and accuracy of diagnostic models, this work contributes to the overall goal of improving mental health care and outcomes, with potential applications extending from disease diagnosis and prognosis to the development of therapeutic interventions. The research presented in this chapter has been published in Briefings in Bioinformatics (Budhraja et al., 2023a).

# Chapter 4

# Personalization and Explainability: Tailoring to Individual Needs and Enabling Transparency

## 4.1 Introduction

The field of mental healthcare is undergoing a significant transformation, driven by the integration of artificial intelligence (AI) and the growing recognition of the need for personalized, patient-centric approaches. This chapter explores two critical aspects of this transformation: personalized modelling and explainable AI. By leveraging these advanced techniques, we aim to address the complexities of mental health disorders, improve diagnostic accuracy, and facilitate the development of individualized interventions tailored to each patient's specific needs. This chapter specifically addresses the research objective of making longitudinal and explainable AI models (Objective 3 and Question 3), and enhance predictive accuracy using longitudinal data (Objective 2 and Question 2).

Personalized modelling is a fundamental component of this paradigm shift, acknowledging the heterogeneous nature of mental health disorders and the intricate interplay of biological, psychological, and social factors that contribute to their manifestation. Conventional approaches that adopt a one-size-fits-all perspective often fail to capture the individual variations and subtle nuances that characterize these disorders. This chapter investigates the potential of personalized modelling techniques, with a specific focus on transductive learning, in adapting to individual-specific data distributions and identifying the underlying patterns that shape each patient's unique mental health trajectory.

Equally essential to the advancement of mental healthcare is the explainability of AI models. As AI systems become increasingly complex, their decision-making processes often remain opaque, hindering trust and acceptance among clinicians and patients. Explainable AI (XAI) techniques aim to address this challenge by providing transparent and interpretable explanations of model predictions. This chapter emphasizes the significance of explainability in mental health AI, discussing the obstacles and approaches for developing models that not only exhibit strong performance but also offer clear insights into their reasoning process.

Central to this chapter is the introduction of a novel methodology, the Dynamic Attention Gateway (DAG) for Liquid State Machines (LSMs). This methodology is designed to provide meaningful explanations for model predictions while maintaining high predictive accuracy. Through a series of empirical analyses on the LYRIKS and ADNI (TADPOLE) datasets, we demonstrate the effectiveness of this approach

in predicting Ultra-High Risk (UHR) for psychosis and cognitive impairment, respectively. The chapter not only highlights the predictive performance of DAG and transductive methods but also showcases their inherent explainability capabilities, facilitating a deeper understanding of the factors influencing model decisions.

The chapter is structured as follows: Section 2 establishes the significance of patient-centric healthcare and the need for personalized modelling in mental health. Section 3 introduces the key concepts and methods of transductive learning, followed by a discussion on the importance and challenges of explainability in AI for mental health in Section 4. The Dynamic Attention Gateway (DAG) is then presented in Section 5 as a novel solution for achieving explainable and accurate readout in LSMs. Comparative analyses and case studies are provided throughout the chapter to illustrate the practical applications and benefits of these approaches.

By advancing the frontiers of personalized and explainable AI in mental health, this chapter aims to contribute to the development of precise, reliable, and transparent AI systems that can revolutionize mental healthcare. The integration of transductive learning and DAG into the mental health AI toolkit holds the potential to uncover new insights, support clinical decision-making, and ultimately enhance the outcomes for individuals affected by mental health disorders. As we navigate this transformative landscape, the synergy between personalization and explainability will be crucial in realizing the full potential of AI in advancing mental health research and practice.

## 4.2 Personalized Modelling

### 4.2.1 Significance of Patient-Centric Healthcare

Mental health disorders are complex, multifaceted conditions that are influenced by a variety of factors such as cognitive, behavioural, environmental, and hereditary. The traditional "one-size-fits-all" approach for diagnosing and treating health disorders falls short for mental health conditions due to the diversity and individualized nature of how mental health conditions manifest and affect each person. Not only manifestation, but individuals can also respond differently to the same treatment due to their unique biological and psycho-social profiles. Personalized modelling in mental health aims to tailor diagnostic procedures, treatment selection, and disease management strategies to the specific characteristics and needs of each individual patient.

The significance of patient-centric healthcare in mental health is far-reaching, with the potential to completely change the way we approach mental health care delivery and improve outcomes for those living with mental health conditions. By accounting for individual variability, patient-centric personalized healthcare can:

1. **Enhance diagnostic precision**: Since mental health conditions can be caused due to a variety of factors, integrating the modalities related to the factors, such as genomics, neuroimaging, cognitive assessments, and clinical tests, can facilitate the identification of personalized factors contributing to mental health disorders. Grouping the personalized factors can help in identifying different subtypes or endophenotypes of disorders. Improved precision in detecting the underlying causes facilitates more accurate diagnoses, moving away from the current symptom-based diagnostic approach, which often fails to capture the underlying biological and environmental factors contributing to mental illness.

2. **Optimize treatment selection**: By considering an individual's genetic profile, neurobiological markers (e.g., brain structure and function), environmental exposures, lifestyle factors, and clinical characteristics, personalized medicine can guide the selection of pharmacological and psychological treatments most likely to be effective for that individual. This approach can minimize trial-and-error prescribing and increase the likelihood of treatment response, reducing the burden of ineffective treatments and associated adverse effects.

3. **Facilitate early intervention and prevention**: Identifying individuals at high risk for developing mental health disorders based on genetic, neurobiological, and environmental risk factors can enable early preventive interventions. These interventions may include lifestyle modifications, psychological therapies, or targeted pharmacological interventions, potentially delaying or preventing the onset of illness and reducing the associated personal and societal costs.

4. **Improve patient engagement and adherence**: Personalized treatment plans tailored to an individual's specific needs, preferences, and circumstances can increase patient understanding, motivation, and adherence. By involving patients in the decision-making process and providing personalized education and support, personalized healthcare can empower individuals to take an active role in their mental health care, leading to better treatment outcomes and reduced relapse rates.

5. **Reduce healthcare costs and societal burden**: Mental health disorders pose a significant economic burden on healthcare systems and society, stemming from direct costs associated with treatment and indirect costs related to lost productivity, disability, and social support services. By minimizing ineffective treatments, adverse events, and chronicity of mental health conditions, personalized healthcare can potentially reduce these costs, improve resource allocation within healthcare systems, and mitigate strain on caregivers and support networks.

6. **Advance mental health research and knowledge**: The implementation of personalized healthcare in mental health necessitates the collection and integration of large-scale, multimodal data from diverse patient populations. This data can facilitate the identification of novel biomarkers, risk factors, and disease mechanisms, advancing our understanding of mental health disorders and informing the development of more effective diagnostic tools, treatments, and preventive strategies.

The integration of personalized healthcare approaches in mental health, facilitated by advances in omics technologies, neuroimaging, digital phenotyping, and computational methods for analyzing large-scale, multimodal data, holds immense promise for transforming mental health care delivery and improving the lives of those affected by mental health disorders. However, realizing the full potential of personalized mental healthcare will require interdisciplinary collaboration, robust data infrastructure, and a paradigm shift in how we approach mental health research, clinical practice, and healthcare delivery.

### 4.2.2 Personalized and Local Modelling Methods

To enable truly personalized and localized predictions in mental healthcare, machine learning techniques must go beyond global models that generalize across an entire population. Personalized and local modelling techniques aim to capture individual-level patterns and dynamics, tailoring predictions to the specific characteristics and circumstances of each patient (Kasabov, 2007b; Doborjeh et al., 2022). Several machine learning approaches have shown promise in this regard:

- **Memory-based Learning and Instance-based Models**: These techniques, such as k-Nearest Neighbors (kNN) (Cover and Hart, 1967), make predictions based on the similarity of a new instance to previously observed cases or instances in the training data. By identifying the most relevant "neighbors" or cases for a specific individual, these models can provide localized predictions that are highly personalized.

- **Gaussian Processes for Localized Kernel Regression**: Gaussian processes (GPs) are non-parametric probabilistic models that can be used for localized kernel regression tasks (Rasmussen, 2003). By learning a kernel function that captures the similarities between datapoints, GPs can make highly localized predictions, adapting their behavior based on the characteristics of each individual case.

- **Transductive Learning**: Transductive learning methods, like transductive support vector machines and graph-based semi-supervised learning, leverage information from both labeled and unlabeled data points during the training process (Vapnik, Vapnik, et al., 1998). This allows the model to learn from the specific data distribution of the target domain or individual, enabling more personalized and localized predictions.

- **Multi-Task Learning and Personalized Neural Networks**: Multi-task learning frameworks (Caruana, 1997) and personalized neural network architectures, such as attention-based models, can learn personalized representations and mappings for each individual. These models can effectively capture idiosyncratic patterns and personal characteristics, enabling individualized predictions and recommendations.

- **Meta-Learning and Few-Shot Learning**: Meta-learning (Finn, Abbeel, and Levine, 2017) and few-shot learning techniques aim to rapidly adapt models to new individuals or scenarios with limited data. By learning to learn from previous experiences, these methods can quickly personalize models to new individuals, leveraging shared knowledge while accounting for individual-specific characteristics.

- **Causal Inference and Counterfactual Reasoning**: Causal inference and counterfactual reasoning methods (Pearl et al., 2000), such as structural causal models and potential outcomes frameworks, can help disentangle the complex causal relationships between individual characteristics, environmental factors, and mental health outcomes. These approaches can facilitate personalized what-if analyses and inform individualized intervention strategies.

- **Local Model Interpretations and Case-Based Reasoning**: Methods like locally interpretable model-agnostic explanations (LIME) (Ribeiro, Singh, and Guestrin, 2016) can provide local explanations for individual predictions made

by complex global models. These techniques can illustrate how a model's predictions change with variations in the input data, facilitating personalized interpretations and insights.

Successful implementation of these personalized and local modeling techniques requires careful consideration of data quality, feature engineering, and model interpretation. Close collaboration between mental health professionals, data scientists, and domain experts is essential to ensure that personalized models align with clinical needs and provide actionable insights for personalized mental healthcare. Additionally, robust evaluation frameworks, including prospective clinical studies and real-world deployment, are crucial to validate the effectiveness and generalizability of these personalized modeling approaches in mental healthcare settings (Weng et al., 2017).

### 4.2.3 Transductive Learning

Transductive learning is a powerful paradigm for developing personalized models tailored to the unique characteristics and data distributions of specific individuals or target domains (Gammerman, Vovk, and Vapnik, 2013). In contrast to traditional inductive learning, which aims to generalize from a training set to unseen test data, transductive learning leverages information from both labeled and unlabeled data points, including the target cases of interest, during the training process (Joachims et al., 1999).

#### Inductive vs Transductive Learning

In inductive learning, models are trained on a labeled training set and are expected to generalize well to unseen test data drawn from the same underlying distribution. This approach assumes that the training and test data are independent and identically distributed (i.i.d.), which may not hold true in personalized modeling scenarios where we aim to make accurate predictions for specific individuals or cases with unique characteristics or data distributions (Vapnik, Vapnik, et al., 1998).

Transductive learning, on the other hand, focuses on making accurate predictions for a specific set of individuals or cases, rather than generalizing to an entire population. By exploiting the structure and patterns present in the unlabeled data, which often includes the target cases of interest, transductive models can learn representations and decision boundaries that are better adapted to the specific data distribution and characteristics of the target domain or individuals (Joachims, 2003).

#### Transductive Learning Methods

Several transductive learning approaches have been explored for personalized modeling in mental health, including:

1. **KNN-Based Methods**: K-nearest neighbors (KNN) and other instance-based learning techniques can be adapted to the transductive setting by considering both labeled and unlabeled data points when identifying the most relevant neighbors for a target case (Kasabov, 2007a). By leveraging the structure and patterns in the unlabeled data, these methods can make more informed and personalized predictions for the target individuals or cases.

2. **Transductive Support Vector Machines (TSVMs)**: TSVMs extend the popular support vector machine (SVM) algorithm to the transductive setting (Joachims et al., 1999). Unlike traditional SVMs that only consider labeled data during training, TSVMs simultaneously optimize the decision boundary and label assignments for both labeled and unlabeled data points. This allows the model to learn from the underlying data distribution of the target cases, resulting in more personalized and accurate predictions.

3. **Label Propagation and Graph-based Methods**: Graph-based semi-supervised learning techniques, such as label propagation and graph regularization, construct similarity graphs that capture the relationships between labeled and unlabeled data points (Zhu, 2005). By propagating label information along the edges of the graph, these methods can leverage the structure of the unlabeled data to make more informed and personalized predictions for the target cases.

#### Conclusion

Transductive learning is a promising avenue for personalized healthcare, particularly in mental health, where patient-specific predictions can significantly impact treatment outcomes and quality of life (Marquand et al., 2016). By adapting to individual-specific data distributions, and leveraging unlabeled data, which is often abundant in clinical settings, transductive models can provide nuanced insights tailored to individual patients. However, it is important to consider the cost-effectiveness of developing individualized models for every patient (Hatz, Schremser, and Rogowski, 2014). Despite this, the potential benefits of transductive models underscore their value in advancing personalized and precision medicine in mental health.

### 4.2.4 Key Transductive Methods: WWKNN and TWNFI

Two prominent transductive learning methods that have shown promise in personalized modeling for mental health are the Weighted Weighted k-Nearest Neighbors (WWKNN) and the Transductive Weighted Neuro-Fuzzy Inference (TWNFI). These techniques leverage the principles of transductive learning to make accurate and personalized predictions for target cases while effectively utilizing both labeled and unlabeled data.

#### Weighted Weighted k-Nearest Neighbors (WWKNN)

The Weighted Weighted k-Nearest Neighbor (WWKNN) algorithm, proposed in Kasabov, 2007a, is a transductive instance-based learning method that extends the traditional k-nearest neighbor (KNN) approach. WWKNN is particularly well-suited for personalized modeling in mental health due to its ability to handle mixed data types (e.g., continuous, categorical, text) and its interpretability.

The WWKNN algorithm operates as follows:

1. For each test case (target individual or case), identify the k nearest neighbors from the labeled and unlabeled data using a distance metric that considers all feature types.

2. Assign weights to the neighbors based on their similarity to the test case, where the similarity measure is also weighted i.e. accounts for feature importance.

3. Compute a weighted sum of the neighbor labels (for classification tasks) or outputs (for regression tasks) to generate the final prediction for the test case.

The key advantages of WWKNN for personalized modeling in mental health include:

- **Transductive Learning**: By considering both labeled and unlabeled data points, WWKNN can leverage the structure and patterns present in the target cases, leading to more accurate and personalized predictions.
- **Interpretability**: The instance-based nature of WWKNN allows for easy interpretation of the predictions, as the contributing neighbors and their respective weights can be examined.
- **Handling Mixed Data Types**: WWKNN can seamlessly handle mixed data types, including continuous, categorical, and textual features, which is crucial in mental health domains where data can be heterogeneous.

#### Transductive Weighted Neuro-Fuzzy Inference (TWNFI)

The Transductive Weighted Neuro-Fuzzy Inference (TWNFI) algorithm, developed by Song and Kasabov, 2006, is a transductive learning method that combines the principles of fuzzy inference systems and neural networks. TWNFI is particularly well-suited for personalized modeling in mental health due to its ability to handle uncertainty, non-linearity, and interpretability.

The TWNFI algorithm operates as follows:

1. Initialize a neuro-fuzzy inference system with a set of fuzzy rules and membership functions.
2. Iteratively optimize the fuzzy rules and membership functions using both labeled and unlabeled data, minimizing a transductive cost function that considers the prediction errors on labeled data and the consistency with unlabeled data.
3. Make predictions for target cases using the optimized neuro-fuzzy inference system, leveraging the personalized fuzzy rules and membership functions.

The key advantages of TWNFI include:

- **Personalized Modelling**: By leveraging both labeled and unlabeled data during the optimization process, TWNFI can learn personalized fuzzy rules and membership functions that are tailored to the target cases.
- **Handling Uncertainty and Non-linearity**: TWNFI's fuzzy inference system can effectively handle uncertainty and non-linear relationships, which are common in mental health domains.
- **Interpretability**: The learned fuzzy rules and membership functions provide interpretable representations of the underlying patterns and decision-making processes, facilitating trust and understanding in clinical settings.

Both WWKNN and TWNFI have demonstrated promising results in personalized modeling tasks across various domains, including mental health applications. However, their effectiveness may depend on factors such as data quality, feature engineering, and the specific characteristics of the target population. Careful evaluation and validation in real-world clinical settings are essential to ensure the responsible and trustworthy application of these transductive learning methods in mental healthcare.

### 4.2.5 Analysis of Transductive Methods for Personalized UHR Prediction

This analysis dives into the application of transductive learning methods, particularly Weighted Weighted k-Nearest Neighbors (WWKNN) and Transductive Weighted Neuro-Fuzzy Inference (TWNFI), for the personalized prediction of Ultra-High Risk (UHR) for psychosis using the LYRIKS microarray gene data. The study is structured to not only assess the predictive performance of these transductive methods but also to enhance the explainability of the models for clinical use by aggregating complex genetic data into more interpretable biological pathways.

#### Study Design

1. **Objective**: This analysis utilizes the LYRIKS microarray gene data to facilitate early diagnosis and prognosis of Ultra-High Risk (UHR) for psychosis in youth. The objective is to enhance the models' explainability for clinicians by aggregating over 30,000 genes into interpretable pathways. The primary task is binary classification, aiming to predict the 'healthy' or 'at-risk' status of individuals at future intervals using cross-sectional data.

2. **Data Aggregation**: The genes are aggregated to pathways using Kyoto Encyclopedia of Genes and Genomes (KEGG), and WikiPathwaysdatabases. This approach reduces the complexity of genetic data, presenting it within the context of biological pathways, which are more aligned with clinical interpretations and decisions.

3. **Models**: Transductive models, specifically Weighted k-Nearest Neighbors (WWKNN) and Transductive Weighted Neuro-Fuzzy Inference System (TWNFI), are utilized for their capacity to offer individualized predictions. Their performance is evaluated against traditional inductive models like Logistic Regression (LR), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Random Forests (RF), to ascertain the advantages of transductive methods in UHR prediction.

4. **Validation**: Leave-One-Out cross-validation (LOOCV) is the chosen method for validating the performance of the models. LOOCV is particularly suitable for datasets of this nature, ensuring a comprehensive assessment of each model's generalization capabilities across the dataset, which is crucial for the reliable prediction of UHR status.

5. **Hyperparameter Optimization**: Hyperparameters for all classifiers were tuned using a grid search on the training set within each fold of the cross-validation procedure. The selection was accuracy on 20% subset of the training data, which was held out as a validation set during tuning. This process was repeated independently within each fold to ensure that hyperparameter selection did not involve any information from the corresponding test fold.

TABLE 4.1: Prediction accuracies (%) of inductive and transductive models on LYRIKS microarray gene data

| Dataset | # Features | Classifier | | | | | |
|---|---|---|---|---|---|---|---|
| | | **WWKNN** | **TWNFI** | **LR** | **SVM** | **MLP** | **RF** |
| All Genes | 34694 | 91.7 | 88.1 | 89.3 | **92.9** | 90.5 | 83.3 |
| KEGG | 186 | **92.9** | 88.1 | 91.7 | 91.7 | 91.7 | 88.1 |
| WikiPathways | 586 | **96.4** | 92.9 | 91.7 | 94.0 | 91.7 | 86.9 |

6. **Computational Environment**: All experiments were implemented in `Python 3.8`, using libraries including `scikit-learn`, `pandas`, and `NumPy` for model development and evaluation. Visualisations were produced using `Matplotlib` and `Seaborn`. Experiments were executed on an institutional machine equipped with an `Intel Core i5 CPU` and `16 GB RAM`. GPU acceleration was not required for any of the experiments.

**Results**

The results demonstrate the performance of the transductive models, WWKNN and TWNFI, in comparison with traditional inductive models like Logistic Regression (LR), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Random Forests (RF), across different feature sets.

When considering all 34,694 genes as features, the SVM model achieved the highest accuracy of 92.9%, while WWKNN and MLP performed comparably with accuracies of 91.7% and 90.5%, respectively.

Notably, aggregating the genes into interpretable biological pathways not only improved the models' explainability but also led to higher prediction accuracies. Across various KEGG and WikiPathways databases, both WWKNN and TWNFI consistently demonstrated competitive or superior performance compared to the inductive models.

WWKNN achieved the highest accuracy of 96.4% when using the WikiPathways database, and 92.9% on KEGG pathway database, outperforming all other models. TWNFI also exhibited strong performance, achieving an accuracy of 92.9% on the WikiPathways database.

These results highlight the potential of transductive learning methods, WWKNN and TWNFI, for personalized prediction of UHR status using genetic data. Their ability to leverage both labeled and unlabeled data, combined with the interpretability gained through pathway aggregation, makes them valuable tools for early diagnosis and prognosis in mental healthcare.

## 4.3 Explainability of AI in Health

### 4.3.1 AI Safety, Trustworthiness, and Explainability

As artificial intelligence (AI) systems become increasingly prevalent in healthcare, ensuring their safety, trustworthiness, and transparency is of paramount importance. Explainable AI (XAI) techniques aim to shed light on the inner workings and decision-making processes of these complex models, fostering trust, accountability, and ethical AI practices in sensitive domains like mental healthcare.

#### AI Safety in Healthcare: Ensuring Reliability and Minimizing Risks

AI safety refers to the ability of AI systems to behave reliably and avoid unintended or harmful consequences. In mental healthcare settings, where decisions can profoundly impact patient well-being and outcomes, ensuring AI safety is crucial. This involves ensuring that AI-driven decisions or recommendations align with established clinical guidelines, respect patient preferences and values, and do not exacerbate existing biases or disparities in healthcare delivery.

#### Building Trustworthiness in Healthcare AI

Trustworthiness encompasses not only the technical aspects of AI safety but also ethical considerations such as fairness, accountability, and transparency. Building trust in AI systems is essential for their widespread adoption and acceptance in mental healthcare. Several factors contribute to the trustworthiness of AI in mental health, including interpretability and explainability, robustness and reliability, fairness and bias mitigation, privacy and data protection, human agency and oversight, and regulatory compliance and validation.

#### Need for Explainability of AI in Healthcare

Explainable AI (XAI) techniques play a crucial role in enhancing the transparency, interpretability, and trustworthiness of AI systems in mental healthcare. The ability to understand and explain the reasoning behind AI-generated predictions or recommendations is essential for building trust and ensuring appropriate use in clinical settings. XAI techniques can shed light on the decision-making processes of complex AI models, enabling clinicians, patients, and regulators to scrutinize and validate the AI system's behavior.

#### Conclusion

Addressing the challenges of AI safety, trustworthiness, and explainability in mental healthcare requires a multidisciplinary approach involving collaborations between mental health professionals, AI researchers, ethicists, policymakers, and patients. By prioritizing explainability and transparency, AI systems can be effectively integrated into mental healthcare delivery, augmenting clinical decision-making while ensuring patient safety, privacy, and ethical practices.

### 4.3.2 Approaches for Explainability

Explainability in AI uncovers the inner workings and decision-making processes of these complex models. Several methods have been proposed and explored in the field of XAI, each offering unique advantages and capabilities. Here, we discuss some of the key approaches for explainability of AI systems.

#### Model Interpretability

One approach to explainability is to develop inherently interpretable AI models, where the underlying structure and parameters of the model are transparent and can be directly interpreted by humans. Some examples of interpretable models include:

1. **Linear Models**: Linear statistical models such as linear regression are inherently interpretable, as their coefficients directly represent the contribution of each feature to the output.

2. **Decision Trees and Rule-based Models**: Decision trees and rule-based models make decisions based on a set of interpretable rules or decision paths, providing a clear understanding of the underlying logic.

3. **Generalized Additive Models (GAMs)**: GAMs are a type of regression model that can capture non-linear relationships while maintaining interpretability by expressing the output as a sum of interpretable component functions.

4. **Bayesian Networks and Causal Models**: Bayesian networks and causal models can represent and reason about complex dependencies and causal relationships between variables, offering insights into the underlying mechanisms driving outcomes.

While interpretable models may sacrifice some predictive performance compared to more complex black-box models, they can provide valuable insights and transparency, particularly in domains like mental healthcare where interpretability is critical for trust and accountability.

**Post-hoc Explanation Methods**

Post-hoc explanation methods aim to explain the predictions or decisions made by complex black-box models, such as deep neural networks or ensemble models, after they have been trained. These techniques can be applied to existing models without modifying their architectures or training processes. Some widely used post-hoc explanation methods include:

1. **Feature Importance and Saliency Maps**: These methods quantify the contribution of each input feature or data point to the model's output, enabling the identification of the most influential factors driving a particular prediction or decision.

2. **Local Interpretable Model-agnostic Explanations (LIME)**: LIME approximates the behavior of a complex model in the vicinity of a specific instance by fitting a simple, interpretable model (e.g., a linear model) to the local data and examining its parameters.

3. **Shapley Additive Explanations (SHAP)**: SHAP is a game-theoretic approach that assigns contribution values to each feature based on their marginal contributions to the model's output, providing insights into feature importance and interactions.

4. **Counterfactual Explanations**: Counterfactual explanations identify the minimal changes to the input features that would have led to a different model prediction, helping to understand the decision boundaries and sensitivity of the model.

5. **Concept Activation Vectors (CAVs)**: CAVs identify the high-level semantic concepts (e.g., diagnostic criteria or symptoms) that a model has learned and how they contribute to its predictions, providing human-interpretable explanations.

These post-hoc explanation methods can be applied to existing AI models in mental healthcare, enabling stakeholders to understand and scrutinize the models' decision-making processes without sacrificing predictive performance.

#### Interpretable Representations and Attention Mechanisms

Another approach to enhancing explainability is to develop AI models that learn interpretable representations or incorporate attention mechanisms that highlight the most relevant input features or components contributing to a particular prediction or decision. Some examples include:

1. **Attention-based Neural Networks**: These models use attention mechanisms to dynamically focus on the most relevant parts of the input data when making predictions, providing insights into the model's decision-making process.
2. **Disentangled Representations**: Disentangled representations aim to separate the underlying factors of variation in the input data into distinct, interpretable dimensions, enabling the understanding of how different factors contribute to the model's output.
3. **Concept-based Representations**: Concept-based representations learn to associate high-level semantic concepts (e.g., diagnostic criteria or symptoms) with specific patterns in the input data, providing human-interpretable explanations for the model's decisions.
4. **Hierarchical and Modular Models**: Hierarchical and modular models decompose complex tasks into simpler, interpretable subtasks or modules, each responsible for capturing specific aspects of the data or decision-making process.

By incorporating interpretable representations and attention mechanisms, AI models can provide more transparent and interpretable decision-making processes, facilitating trust and understanding in mental healthcare applications.

Approaches for enhancing explainability in healthcare AI are diverse, each offering unique insights into the workings of complex models. By carefully selecting and implementing these approaches, developers can create AI systems that are not only powerful and predictive but also transparent and understandable, paving the way for their responsible and effective use in clinical settings.

### 4.3.3 Intrinsic Explainability of Transductive Approaches

In section 4.2.5, we have observed the competitive performance of transductive learning approaches WWKNN and TWNFI, compared to traditional inductive models in predicting UHR status for psychosis using gene microarray data from the LYRIKS dataset. While accuracy is crucial, the advantages of transductive methods extend beyond predictive performance; they also offer intrinsic explainability and personalized interpretations that are essential in the context of mental healthcare.

#### Explainabiltiy of WWKNN

Figure 4.1 demonstrates the explainability of the WWKNN model for a specific patient (query) being assessed for UHR status. Interpreting the WWKNN Explainability Plot:

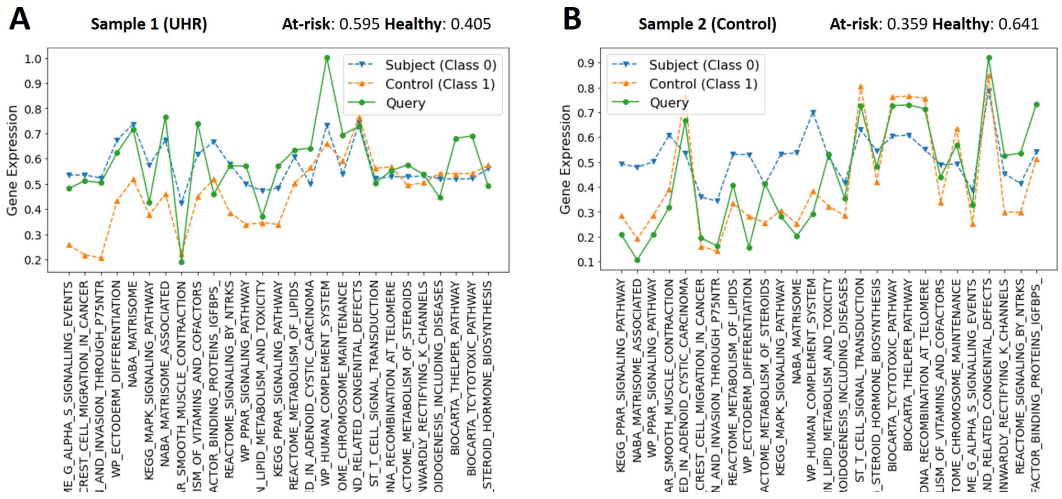


FIGURE 4.1: WWKNN explainability for (A) an at-risk individual and (B) a healthy individual. The plot shows the gene expression levels of the top features, sorted by their importance for the specific individual (green line) being assessed. The blue line represents the average values for the at-risk (UHR) class, while the orange line represents the average values for the healthy control class. The personalized feature importance ranking and comparison with class averages enables personalized decision making for individuals at risk of developing psychosis.

1. **Feature Importance Ranking**: The features on the x-axis are sorted based on their importance for the specific patient being assessed, determined by the Signal-to-Noise Ratio (SNR) algorithm. This ranking provides a clear understanding of which genetic markers are most influential in determining the patient's UHR status. By focusing on the top-ranked features, clinicians can prioritize their analysis and decision-making process.

2. **Comparison with Class Averages**: The plot allows for a direct comparison of the patient's gene expression levels (green line) with the average values of the at-risk (blue line) and healthy control (orange line) classes. By visually examining the deviation of the patient's values from the class averages, clinicians can identify patterns and abnormalities that may indicate an increased risk for psychosis.

3. **Personalized Interpretation**: The explainability plot is generated specifically for each individual patient, taking into account their unique genetic profile. This personalized approach enables clinicians to make informed decisions based on the patient's specific characteristics and risk factors, rather than relying on general population-based thresholds or guidelines.

4. **Identification of Relevant Biological Pathways**: The top-ranked features across various can be mapped to specific genes and biological pathways, providing insights into the underlying mechanisms that may contribute to the patient's UHR status. By examining the functions and interactions of the implicated genes, clinicians can gain a deeper understanding of the patient's condition and potential targets for intervention.

The WWKNN explainability plot serves as a powerful tool for clinicians to support their decision-making process. By visualizing the patient's genetic profile in

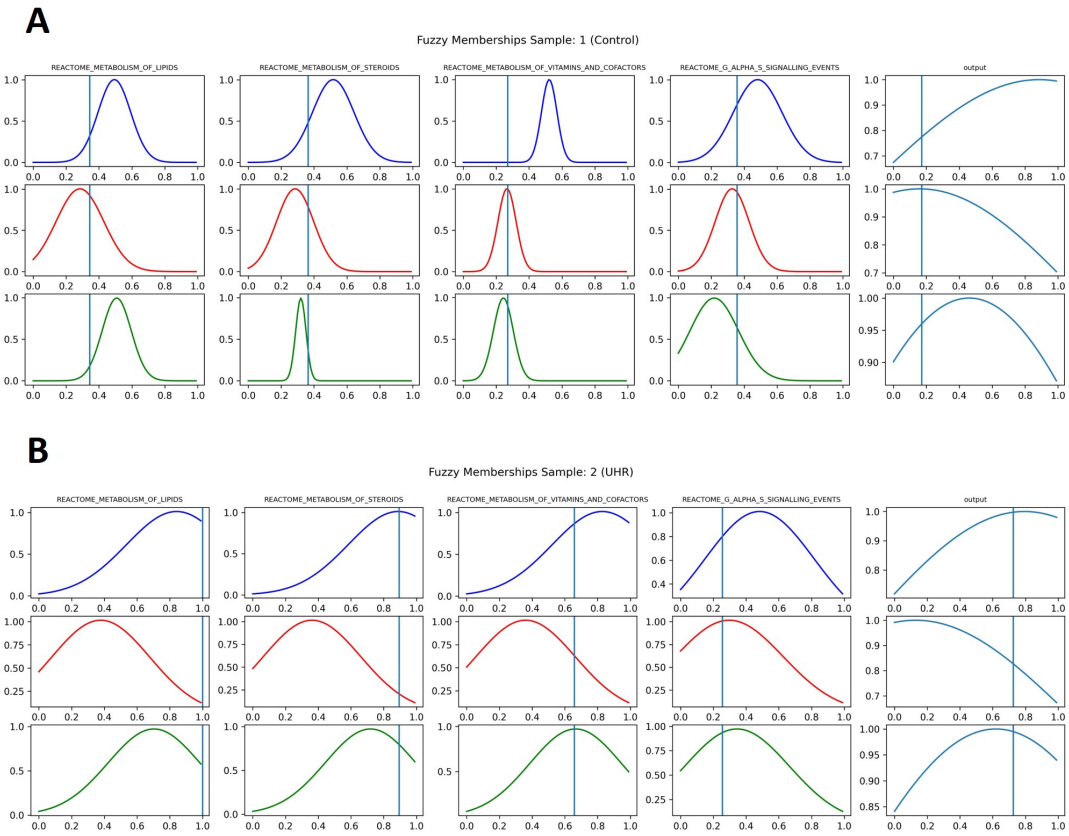


FIGURE 4.2: TWNFI explainability for (A) a healthy individual and (B) an at-risk individual. The plots display the learned personalized fuzzy membership functions for different input features. The intersection of the vertical line, representing the individual being assessed, with the membership functions shows the degree of membership to different fuzzy sets. Higher membership in the blue function indicates alignment with the at-risk group, while higher membership in the red function suggests alignment with the healthy control group. The memberships across features show the contribution of each feature to the final prediction, enabling personalized decision-making.

relation to the class averages and highlighting the most important features, clinicians can make informed assessments regarding the patient's UHR status, treatment options, and monitoring strategies.

### Explainability of TWNFI

Similarly, the TWNFI model offers explainability through its interpretable fuzzy rule-based structure. TWNFI learns fuzzy rules and membership functions that capture the relationships between input features and the predicted output (UHR status). Figure 4.2 illustrates the interpretability of the TWNFI model using fuzzy membership functions. Interpreting the TWNFI Model:

1. **Fuzzy Membership Functions**: The plots show the learned fuzzy membership functions for each input feature. The x-axis represents the feature values, while the y-axis indicates the degree of membership to different fuzzy sets. In the plots presented here, higher membership in the first membership function (blue) indicates a higher alignment with the at-risk group, whereas higher

membership in the second membership function (red) suggests a higher alignment with the healthy control group.

2. **Individual Assessment**: The vertical line in each plot represents the specific individual being assessed. The intersection of this line with the membership functions shows the degree of membership of the individual's feature values to different fuzzy sets. By examining these intersections, the alignment of each feature towards higher risk or lower risk can be determined.

3. **Feature Contribution**: By comparing the membership degrees across different features, the contribution of each feature towards the final prediction can be understood. Features with higher membership in the at-risk function (blue) are more indicative of UHR status, while features with higher membership in the healthy control function (red) suggest a lower risk.

4. **Personalized Explanations**: The TWNFI explainability plots are generated specifically for each individual using personalized membership functions that take into account the individual's unique genetic profile. This personalized approach enables informed decision making based on the individual's specific characteristics and risk factors.

The TWNFI explainability plots serve as a useful tool for clinicians to support their decision-making process. By visualizing the individual's membership degrees to the learned fuzzy rules across different features, clinicians can understand the contribution of each feature towards higher or lower risk and make informed assessments regarding the individual's UHR status, treatment options, and monitoring strategies.

Explainability is a critical aspect of transductive learning approaches in health, as it allows for scrutiny, validation, and continuous improvement of the models by domain experts. By providing interpretable personalized visualisations, these approaches can foster trust, accountability, and effective collaboration between AI systems and mental health professionals, ultimately leading to more informed and personalized clinical decision-making for UHR prediction and intervention.

### 4.3.4 Post-hoc Explainability using SHAP values

While transductive learning approaches like WWKNN and TWNFI offer intrinsic explainability, post-hoc explainability techniques can provide insights into the decision-making process of any machine learning model, including more complex inductive models. One powerful post-hoc explainability method is SHapley Additive exPlanations (SHAP), which uses game theory to compute the contribution of each feature to the model's prediction for a specific instance.

SHAP values quantify the impact of each feature on the model's output by considering all possible combinations of features and comparing the model's predictions with and without each feature. The SHAP value for a feature represents the average marginal contribution of that feature across all possible coalitions of features. Positive SHAP values indicate that the feature contributes positively to the model's output, while negative values suggest a negative contribution.

Figure 4.3 demonstrates the use of SHAP values for post-hoc explainability of the predictions of a Gradient Boosting Decision Tree (GBDT) model trained on the combined clinical, cognitive, and social assessment data from the LYRIKS dataset to identify UHR and healthy control individuals. Interpreting the SHAP Value Plots:

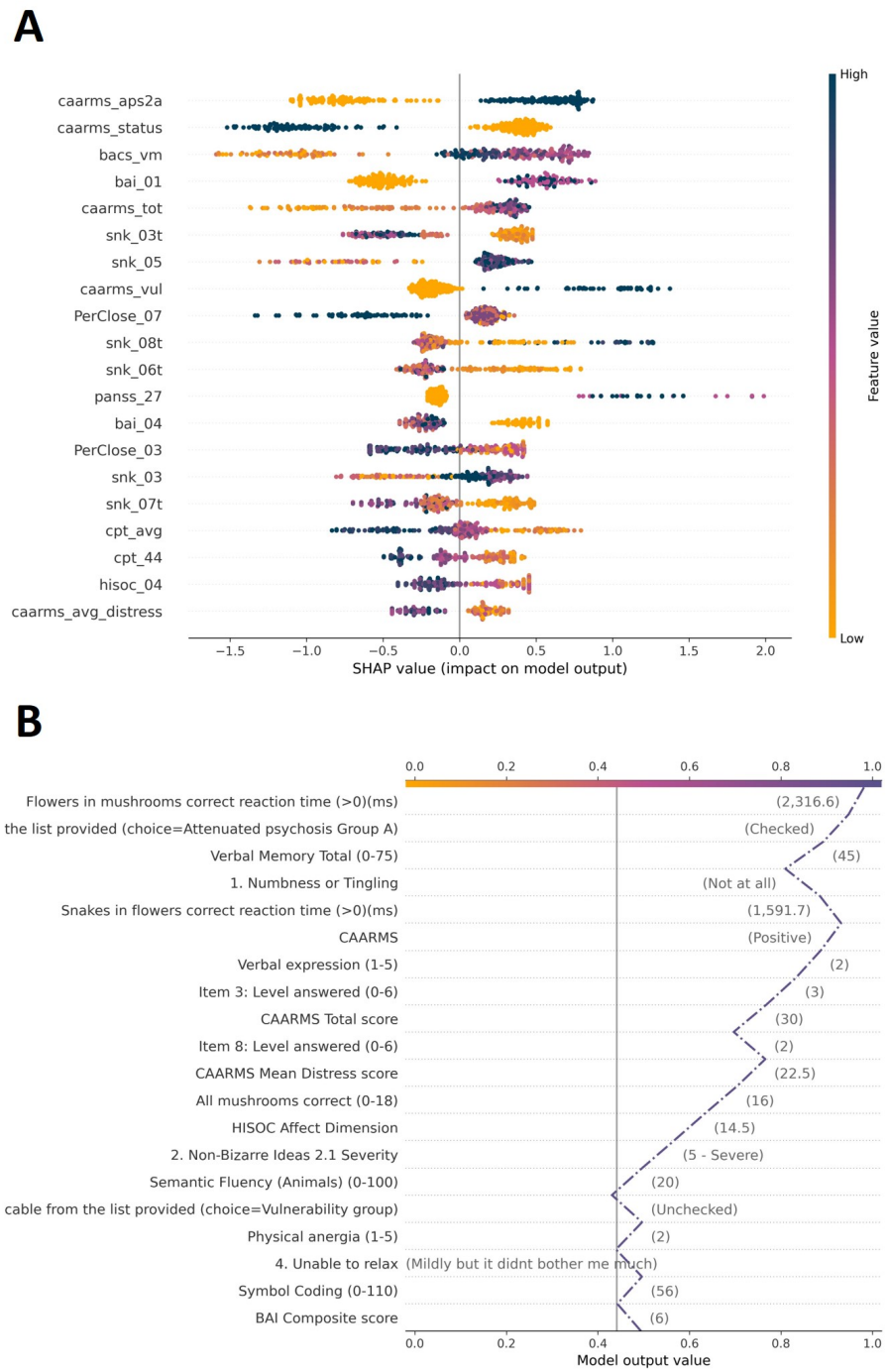


FIGURE 4.3: SHAP value plot for post-hoc explainability of UHR prediction models. (A) The summary plot ranks features by their overall importance based on the magnitude of SHAP values across the dataset. (B) The SHAP values for an individual patient show the contribution of each feature to the model's prediction for that specific instance, with positive values (indigo) indicating increased risk and negative values (yellow) suggesting reduced risk.

1. **Global Feature Importance**: The summary plot (Figure 4.3A) ranks features based on their overall importance across the entire dataset. Features with larger absolute SHAP values (both positive and negative) have a more significant impact on the model's predictions. This global view helps identify the most influential genetic markers or biological pathways in determining UHR status. These markers can be analyzed to identify links with early psychosis risk in prior clinical studies.

2. **Feature Value Interpretation**: The color coding of the data points represents the actual feature values, with indigo indicating higher values and yellow indicating lower values. This information helps to interpret the directionality of how the feature contributes across it's full range of values.

3. **Individual Feature Contributions**: The individual SHAP decision plot (Figure 4.3B) shows the contribution of each feature to the model's prediction for a specific patient. Positive SHAP values indicate that the feature contributes to an increased risk of UHR, while negative values suggest a reduced risk. By examining the individual feature contributions, clinicians can understand which specific factors are driving the model's prediction for that patient.

Post-hoc explainability using SHAP values complements the intrinsic explainability of transductive learning approaches. While transductive methods like WWKNN and TWNFI provide personalized explanations, SHAP values offer a more general framework for understanding feature contributions across any type of model. By providing global and local explanations of feature contributions, SHAP values enhance the transparency and trustworthiness of the models, facilitating effective understanding of AI decision making by health professionals, and ultimately leading to improved personalized care for individuals at risk of developing psychosis.

## 4.4 DAG: Explainable and Accurate Readout for Spiking Dynamics

### 4.4.1 Concept and Design

The Dynamic Attention Gateway (DAG) is an explainable and accurate readout layer for Liquid State Machines (LSMs), that enables interpretability of how LSMs process and extract information from the complex spiking dynamics of their liquid reservoir. The key concept behind DAG is to incorporate an attention mechanism that allows the readout layer to selectively focus on the most informative spiking patterns in the reservoir for a given task. This also leads to enhanced accuracy of the output through precise weighting of different temporal states of the reservoir.

#### Architecture

Figure 4.4 illustrates the architecture of the DAG readout layer and its integration with the liquid reservoir of an LSM. The main components of the DAG architecture are as follows:

1. **Liquid Reservoir**: The liquid reservoir is a network of randomly-connected recurrent spiking neurons that either fire or do not fire (binary activation) in response to input signals. The neurons can be either excitatory or inhibitory, and their firing at different time points leads to complex spiking dynamics in

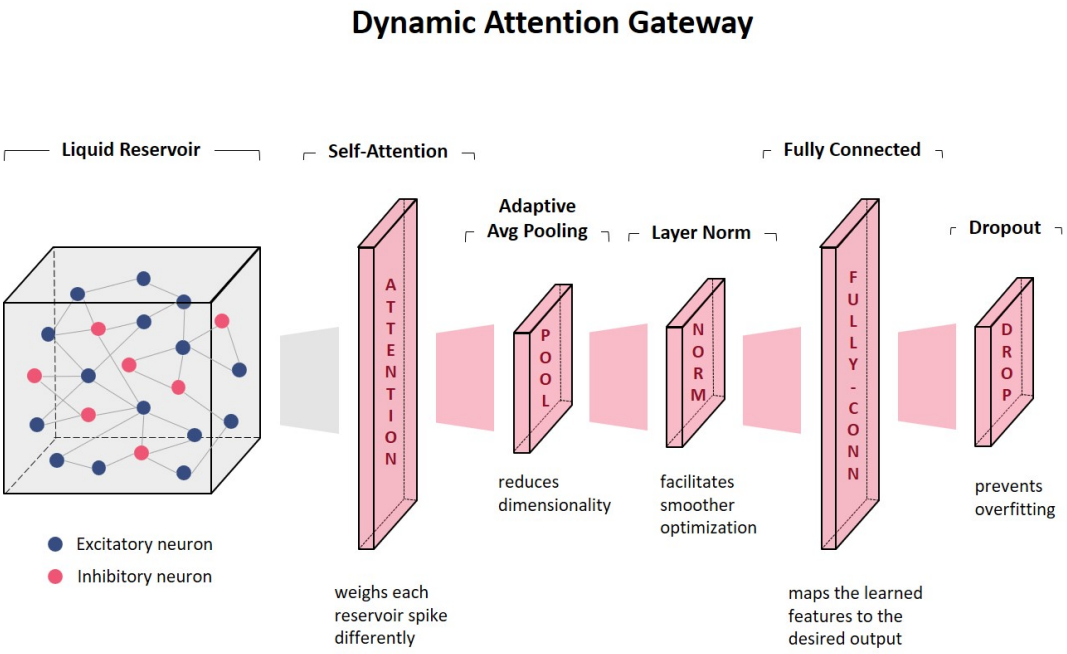


FIGURE 4.4: Dynamic Attention Gateway (DAG) architecture for explainable and accurate readout in Liquid State Machines.

the reservoir. This rich spiking dynamics provides a high-dimensional representation of the input signal over time, serving as the foundation for both explainability and accuracy.

2. **Self-Attention Mechanism**: The self-attention mechanism in DAG allows the readout layer to assign different weights to the spike generated at each time point in the liquid reservoir. By learning attention weights, the readout layer can focus on the most task-relevant spiking patterns, enhancing the model's predictive accuracy. The attention weights are learned through a trainable neural network that takes the reservoir spikes as input and outputs a set of attention scores. These scores are then used to weigh the reservoir spikes before passing them to the subsequent layers, ensuring that the model prioritizes features most correlated with the desired output.

3. **Adaptive Average Pooling**: The adaptive average pooling layer in DAG reduces the dimensionality of the attended spike representations while preserving the essential information, which is key for achieving high accuracy in predictions. This pooling operation helps to summarize the attended spikes and extract meaningful features for further processing. The adaptive nature of the pooling layer allows it to handle variable-length spike sequences and maintain a fixed output size, essential for consistent model performance across different inputs.

4. **Layer Normalization**: Following the adaptive average pooling layer, a layer normalization component is applied. Layer normalization normalizes the activations of the pooled features across the feature dimensions, helping to stabilize the learning process and improve convergence. By normalizing the activations, this layer ensures that the subsequent fully-connected layer receives input with consistent distribution, facilitating more efficient learning and quicker loss optimisation.

5. **Fully-Connected Layer**: The fully-connected layer in DAG takes the pooled features as input and learns a mapping to the desired output, such as class probabilities in a classification task. This layer consists of a set of trainable weights and biases that transform the pooled features into the final output representation. The fully-connected layer allows the readout to learn a task-specific mapping from the attended reservoir spikes to the desired output.

6. **Dropout Regularization**: Dropout is a regularization technique applied to the fully-connected layer in DAG to prevent overfitting and improve generalization. During training, dropout randomly sets a fraction of the input units to zero, encouraging the readout layer to learn more robust and generalizable representations. Dropout helps to mitigate the reliance on specific spiking patterns and promotes the learning of distributed representations that are more accurate.

#### Activation Functions in DAG

In the DAG architecture, two types of activation functions are employed for distinct purposes. First, within the self-attention mechanism, the softmax activation function is applied to the scaled dot-product of queries and keys. This softmax operation normalizes the raw attention scores into a probability distribution across timepoints, ensuring that the attention weights sum to one and allowing the model to meaningfully differentiate the importance of each temporal state. Second, in the feed-forward component of the readout layer, the ReLU activation function is used. ReLU was chosen due to its well-documented benefits in addressing the vanishing gradient problem. It introduces non-linearity, allowing DAG to model complex relationships between spiking patterns and outputs, while maintaining computational simplicity. Unlike sigmoid or tanh, ReLU does not saturate in the positive domain, making it suitable for learning sparse representations of temporally significant reservoir states. Alternatives such as GELU and sigmoid were also considered; however, ReLU offered a favorable balance between performance, ease of optimization, and interpretability when visualizing attention-weighted contributions.

#### Training

The DAG readout layer is trained using backpropagation to optimize the weights of both the attention mechanism and the fully connected layer. For the binary classification task, we use the `BCEWithLogitsLoss` loss function, which combines sigmoid activation followed by binary cross-entropy. This choice is motivated by numerical stability—`BCEWithLogitsLoss` combines both steps into a single optimized operation, reducing the risk of vanishing gradients when activations are close to 0 or 1. It also simplifies implementation and improves training robustness. Optimization is performed using Stochastic Gradient Descent (SGD) with a learning rate of 0.001. This setting reflects a trade-off between convergence speed and training stability, identified through preliminary tuning. Additional hyperparameters—including the dropout probability and weight initialization strategy—were selected based on standard best practices to promote generalization and efficient training.

By incorporating an attention mechanism and a fully-connected layer with dropout, the DAG readout layer provides an explainable and interpretable mapping from the liquid reservoir's spiking dynamics to the output, along with an emphasis on

enhancing accuracy through precise and adaptive feature weighting. The learned attention weights also offer insights into which time points and input features are most informative for a given task, enabling a deeper understanding of the LSM's decision-making process.

### 4.4.2 Temporal Weighting of Spikes in the Reservoir

The Dynamic Attention Gateway (DAG) employs a self-attention mechanism to learn and assign importance to different temporal states of the liquid reservoir, effectively treating each state as an embedding or representation of the input signal at a given timepoint. By leveraging the power of self-attention, DAG can dynamically focus on the most informative spiking patterns within the reservoir over time, enabling it to capture task-relevant temporal dependencies and improve the overall performance of the LSM.

#### Self-Attention Mechanism

Self-attention, a key component of the Transformer architecture (Vaswani et al., 2017), allows a model to attend to different parts of its input sequence to compute a representation of that sequence. In the context of DAG, self-attention is applied to the temporal states of the LSM reservoir to learn the importance of each state in relation to the others. The embedding for each temporal state of the LSM reservoir is represented using the spike or no spike at each neuron in the reservoir.

Let $\mathbf{X} \in \mathbb{R}^{T \times N}$ denote the matrix of reservoir states, where $T$ is the number of timepoints and $N$ is the number of neurons in the reservoir. Each row $\mathbf{x}_t \in \mathbb{R}^N$ represents the state of the reservoir at timepoint $t$.

The self-attention mechanism computes three matrices: the query matrix $\mathbf{Q}$, the key matrix $\mathbf{K}$, and the value matrix $\mathbf{V}$. These matrices are obtained by applying linear transformations to the input matrix $\mathbf{X}$:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q \tag{4.1}$$
$$\mathbf{K} = \mathbf{X}\mathbf{W}_K \tag{4.2}$$
$$\mathbf{V} = \mathbf{X}\mathbf{W}_V \tag{4.3}$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{N \times d}$ are learnable weight matrices, and $d$ is the dimension of the attention embedding. In case of DAG, $d$ is equal to the number of neurons in the reservoir.

The attention scores $\mathbf{A} \in \mathbb{R}^{T \times T}$ are computed as the softmax-normalized dot product between the query and key matrices:

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right) \tag{4.4}$$

The attention scores matrix $\mathbf{A}$ represents the importance of each reservoir state in relation to the others. The output of the self-attention mechanism is obtained by multiplying the attention scores with the value matrix:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{A}\mathbf{V} \tag{4.5}$$

### Temporal Weighting of Reservoir States

DAG applies the self-attention mechanism to the temporal states of the LSM reservoir to learn sample-specific attention weights. Given a matrix of reservoir states $\mathbf{X} \in \mathbb{R}^{T \times N}$, DAG computes the attention scores matrix $\mathbf{A} \in \mathbb{R}^{T \times T}$ using the formula described above.

The attention scores matrix $\mathbf{A}$ provides a rich representation of the temporal dependencies within the reservoir states. Each element $a_{ij}$ in $\mathbf{A}$ represents the attention weight assigned to the reservoir state at timepoint $i$ when computing the representation for timepoint $j$.

DAG then computes the attended reservoir states $\mathbf{X}_{\text{att}} \in \mathbb{R}^{T \times N}$ by multiplying the attention scores matrix with the original reservoir states:

$$\mathbf{X}_{\text{att}} = \mathbf{A}\mathbf{X} \tag{4.6}$$

The attended reservoir states $\mathbf{X}_{\text{att}}$ capture the most informative spiking patterns within the reservoir, weighted by their importance as determined by the self-attention mechanism. These attended states serve as a compact and informative representation of the temporal dynamics in the reservoir, which can be further processed by the subsequent layers of DAG for the specific task at hand.

By learning sample-specific attention weights, DAG can adapt to individual variability and capture subtle differences between different conditions or classes. The attention scores matrix $\mathbf{A}$ provides interpretable insights into the temporal weighting of reservoir states, allowing researchers to identify critical temporal windows and spiking patterns that contribute to the model's predictions.

### Example of Attention Across Samples in LYRIKS Dataset

To illustrate how attention in DAG works, let's consider the data from the cognitive modality of the LYRIKS dataset. Each sample is linearly interpolated from 5 to 52 timepoints to ensure sufficient spiking in the reservoir. Then these samples are encoded into spike trains using step forward algorithm and fed to the LSM and the DAG is trained. Figure 4.5 illustrates the attention alloted by DAG to a few unseen samples.

Each panel in the figure corresponds to a separate sample, labeled as A, B, and C. These panels demonstrate the distribution of attention weights over the neuronal spiking events captured at different timepoints within the LSM reservoir.

Each sample exhibits unique patterns of attention allocation, indicating that the DAG is effectively personalizing its focus based on the dynamic spiking characteristics of each individual's data. For example, Sample 2 in panel B shows concentrated attention in the central timepoints. By examining the differences in attention distribution among the samples, we can infer potential variations in cognitive responses or the progression stages of mental health conditions. Sample 3 in panel C, for example, displays a more dispersed attention pattern compared to Samples 1 and 2.

In summary, DAG leverages the self-attention mechanism to learn temporal weighting of the LSM reservoir states. By computing attention scores and attended reservoir states, DAG can focus on the most informative spiking patterns over time, leading to improved accuracy and interpretability in tasks.

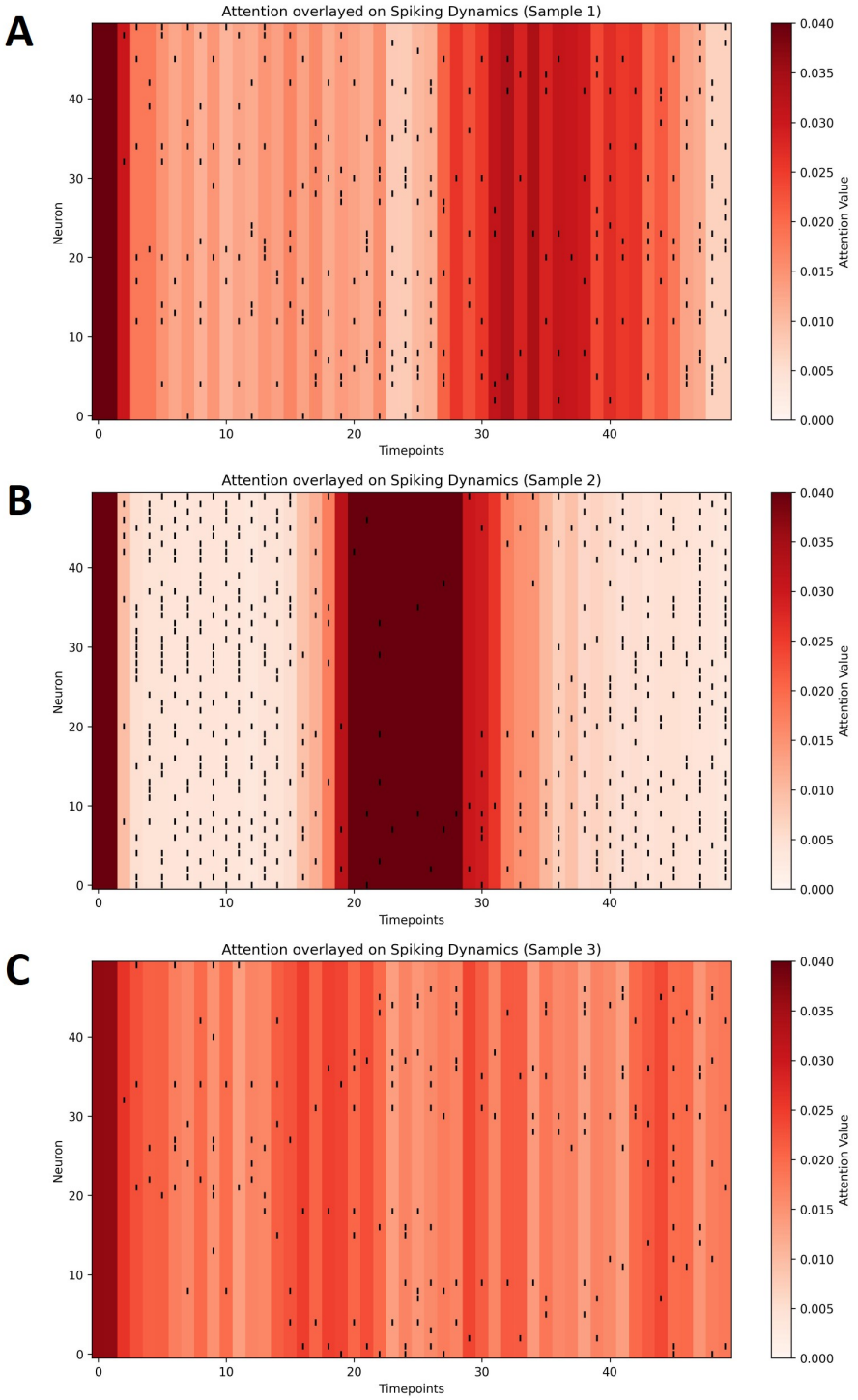


FIGURE 4.5: Visualization of the Dynamic Attention Gateway (DAG) applied to three different samples from the LYRIKS dataset, illustrating how attention weights vary across timepoints for each sample. (A) Relatively uniform attention distribution with a notable focus around the latter timepoints (B) Sharp concentration of attention around the centre (C) A consistent and distributed pattern of attention across most timepoints.

### 4.4.3 Comparative Analysis of DAG's performance against other Readout methods

To evaluate the effectiveness of the Dynamic Attention Gateway (DAG) readout layer, we compare its performance with other commonly used readout methods on two datasets.

#### Study Design

1. **Objective**: The focus of this study is to evaluate the effectiveness of the Dynamic Attention Gateway (DAG) readout layer in comparison to other commonly used readout methods. The comparison is conducted on two datasets: the cognitive modality from the LYRIKS dataset (Lee et al., 2013) and the TADPOLE (The Alzheimer's Disease Prediction Of Longitudinal Evolution) challenge dataset (Marinescu et al., 2018). The task is binary classification, to predict the cognitive status or mental health risk at 24 months using longitudinal data from baseline, 6 months, 12 months, and 18 months.

2. **Datasets**:

   - **LYRIKS Dataset (Cognitive Modality)**: The cognitive modality in LYRIKS includes neuropsychological test scores and cognitive assessments recorded from UHR and healthy participants. The task in this analysis is to predict whether participants are at-risk or healthy at 24 months using data from baseline, 6 months, 12 months, 18 months, and 24 months. The dataset has 94 participants who had cognitive assessments recorded at every timepoint, out of which 61 were healthy at 24 months and the remaining 33 are UHR.
   - **TADPOLE Dataset**: TADPOLE is a longitudinal dataset of Alzheimer's disease progression over 10 years, taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Mueller et al., 2005). Five timepoints were selected from the D2 set for this analysis: baseline, 6 months, 12 months, 18 months, and 24 months. These timepoints were chosen because they were the most commonly available across participants, resulting in a cohort of 608 individuals who had data recorded at all five timepoints. The objective is to predict the cognitive status at 24 months, classifying participants as either healthy or having cognitive impairment (mild cognitive impairment (MCI) or dementia). In the selected cohort, 245 participants were healthy at 24 months, while the remaining participants had MCI or dementia.

3. **Models**: The readout methods chosen for comparison include Logistic Regression (LogReg) (Cox, 1958), Gradient Boosting Decision Trees (GBDT) (Friedman, 2002), Dynamic Evolving Spiking Neural Networks (DeSNN) (Kasabov et al., 2013), and Echo State Networks (ESN) (Jaeger, 2001). These methods were selected based on their diverse characteristics and suitability for the size of the given datasets. LogReg is a simple and interpretable linear model that is most commonly used as readout in LSMs. GBDT is an ensemble method that combines multiple decision trees to capture complex relationships and has shown promise as a readout layer in applications like sleep stage classification (Budhraja et al., 2020). DeSNN, a readout layer in the established NeuCube architecture (Kasabov, 2014), is a spiking neural network that evolves its

connection weights dynamically based on the input data. ESN is a recurrent neural network with a fixed reservoir and a trainable readout layer, capable of capturing temporal dependencies. ESN has recently shown promising results as a readout layer in NeuCube for wrist movement classification (Koprinkova-Hristova et al., 2023).

4. **Data Preprocessing and Validation**: The five timepoints were interpolated to 52 timepoints for each participant. The interpolated data was then converted into spike trains using the step-forward algorithm. The resulting spike trains were introduced to the reservoir of the LSM. To ensure a fair comparison of performance across different readout layers, the same reservoir spiking activity is provided to all readout methods. 5-fold cross validation was employed to record better generalised performance. Hyperparameters for the readout classifiers were tuned using the training set in each fold based on Matthew's Correlation Coefficient (MCC).

5. **Performance Metrics**: Given the potential data imbalance between the healthy and at-risk classes, the selected performance metrics provide a balanced view of the models' effectiveness across both classes. These metrics include accuracy, which gives the overall correctness of the model, Matthew's Correlation Coefficient (MCC), offering a balanced measure even when class distributions are uneven, Area Under the Receiver Operating Characteristic Curve (AUC) assessing the model's ability to distinguish between the classes, and F1 score which shows the balance of precision and recall.

6. **Hyperparameter Optimization**: Hyperparameters for all readout classifiers were tuned using a grid search on the training set within each fold of the cross-validation procedure. The selection criterion was Matthews Correlation Coefficient (MCC), evaluated on a 20% subset of the training data, which was held out as a validation set during tuning. This process was repeated independently within each fold to ensure that hyperparameter selection did not involve any information from the corresponding test fold.

7. **Computational Environment**: All experiments were implemented in `Python 3.8`, using libraries including `scikit-learn`, `XGBoost`, `pandas`, `NumPy`, `PyTorch`, and `NeuCubePy` for model development and evaluation. Visualisations were produced using `Matplotlib` and `Seaborn`. Experiments were executed on an institutional machine equipped with an `Intel Core i5 CPU` and `16 GB RAM`. GPU acceleration was not required for any of the experiments.

#### Performance on the Cognitive modality of LYRIKS Dataset

Table 4.2 shows the performance comparison of readout methods on the cognitive modality of the LYRIKS dataset. Similar to the ADNI dataset, DAG achieves the highest accuracy of 69.15%, MCC of 0.2778, and AUC of 0.6232 among all the compared methods. The F1-score of DAG is also the highest at 0.67, demonstrating its effectiveness in handling the cognitive data.

#### Performance on the ADNI (TADPOLE) Dataset

Table 4.3 presents the performance metrics of different readout methods on the ADNI (TADPOLE) dataset. The DAG readout layer achieves the highest accuracy of 77.8%,

TABLE 4.2: Performance comparison of readout methods on the cognitive (LYRIKS) dataset.

| Method | Accuracy | MCC | AUC | F1-score |
|---|---|---|---|---|
| LogReg | 67.02% | 0.2251 | 0.5999 | 0.65 |
| GBDT | 62.77% | 0.1658 | 0.5810 | 0.62 |
| DeSNN | 60.64% | 0.2163 | 0.6133 | 0.62 |
| ESN | 61.70% | 0.1595 | 0.5797 | 0.62 |
| DAG | **69.15%** | **0.2778** | **0.6232** | **0.67** |

outperforming other methods. It also exhibits the highest MCC of 0.5394 and AUC of 0.7696, indicating its superior ability to distinguish between classes.

TABLE 4.3: Performance comparison of readout methods on the ADNI (TADPOLE) dataset.

| Method | Accuracy | MCC | AUC | F1-score |
|---|---|---|---|---|
| LogReg | 73.36% | 0.4713 | 0.7398 | 0.74 |
| GBDT | 74.34% | 0.4606 | 0.7251 | 0.74 |
| DeSNN | 72.86% | 0.4326 | 0.7139 | 0.73 |
| ESN | 73.36% | 0.4585 | 0.7322 | 0.74 |
| DAG | **77.8%** | **0.5394** | **0.7696** | **0.78** |

The comparative analysis demonstrates the superior performance of the DAG readout layer across both datasets. The comparative analysis also reveals the limitations of traditional machine learning methods, such as LogReg and GBDT, in handling the temporal dynamics and spiking nature of the LSM reservoir. While these methods can capture some relevant information, they may struggle to fully exploit the rich temporal patterns present in the spiking activity. In contrast, DAG's attention mechanism allows it to selectively focus on the most informative spiking patterns, leading to improved performance.

Furthermore, the comparison with DeSNN, a spiking neural network based readout layer in the NeuCube architecture, highlights the advantages of DAG's attention mechanism. Although DeSNN can evolve its weights dynamically based on the input data, it may not effectively capture the most relevant spiking patterns for the given task. DAG's attention mechanism enables it to adaptively weight the importance of different spiking patterns, resulting in better performance.

In summary, the attention mechanism in DAG enables it to effectively capture the most informative spiking patterns in the LSM reservoir, leading to improved accuracy and better class discrimination. The results highlight the potential of DAG as a powerful readout method for LSMs and other reservoir-based methods.

### 4.4.4 Personalized and Global Explainability of LSM's Decision-Making using DAG Attention

The Dynamic Attention Gateway (DAG) not only enhances the accuracy of Liquid State Machines (LSMs) but also provides a powerful tool for interpreting the model's decision-making process at both personalized and global levels. By taking the dot product of the average attention between timepoints and the difference in feature

values between timepoints, the importance of each feature in the dataset can be gauged. This approach provides a comprehensive view of how specific features influence the model's decision-making process, both on an individual level (personalized explainability) and across the entire dataset (global explainability).

Formally, let $A_{t,i}$ denote the attention weight assigned to timepoint $t$ for feature $i$, and let $X_{t,i}$ represent the value of feature $i$ at timepoint $t$. The importance of feature $i$ can be expressed as:

$$\text{Importance}(i) = \sum_{t=1}^{T-1} A_{t,i} \cdot |X_{t+1,i} - X_{t,i}| \tag{4.7}$$

This formula aggregates the product of attention weights and the changes in feature values over all timepoints, capturing the overall influence of each feature.

### Personalized Explainability

Figure 4.6 presents the personalized feature attention profiles for three different samples. Each panel shows the changes in input feature values over time (left) and the dot product of feature changes and attention received (right). This visualization highlights which features were most influential in the model's decision-making for each individual sample.

1. **Sample 1 (Panel A)**: The features that received the highest attention weights are predominantly related to cognitive performance measures, such as the CPT Average D-prime Score and BACS Verbal Memory Total. This indicates that these features played a crucial role in the model's assessment of this particular individual.

2. **Sample 2 (Panel B)**: The attention distribution reveals a different set of influential features, with a significant focus on the BACS Token Motor Task and PC Percentage of Correctly Identified Items. The model appears to weigh these cognitive tasks heavily when evaluating this sample, suggesting their importance in this individual's cognitive profile.

3. **Sample 3 (Panel C)**: The attention pattern is more evenly distributed across a range of cognitive tests, including BACS Digit Sequencing and SNK All Snakes Accuracy. This suggests that the model considers a broader array of cognitive factors for this individual, reflecting a more holistic assessment approach.

### Global Explainability

By examining the highest weighted features assigned by the Dynamic Attention Gateway (DAG) readout layer, we can gain insight into the key factors contributing to the LSM's decision-making process for predicting Ultra-High Risk (UHR) for psychosis and cognitive impairment.

For the LYRIKS dataset, which focuses on predicting UHR for psychosis, the DAG readout assigned the highest weights to BACS Token Motor Task, PC Percentage of Correctly Identified Items, and PC Filtered Perceptual closure average level. These features are highly relevant to the underlying cognitive and perceptual processes associated with UHR for psychosis. The BACS Token Motor Task assesses motor speed and coordination, which have been found to be impaired in individuals at high risk for psychosis (Sawada et al., 2017). The PC Percentage of Correctly

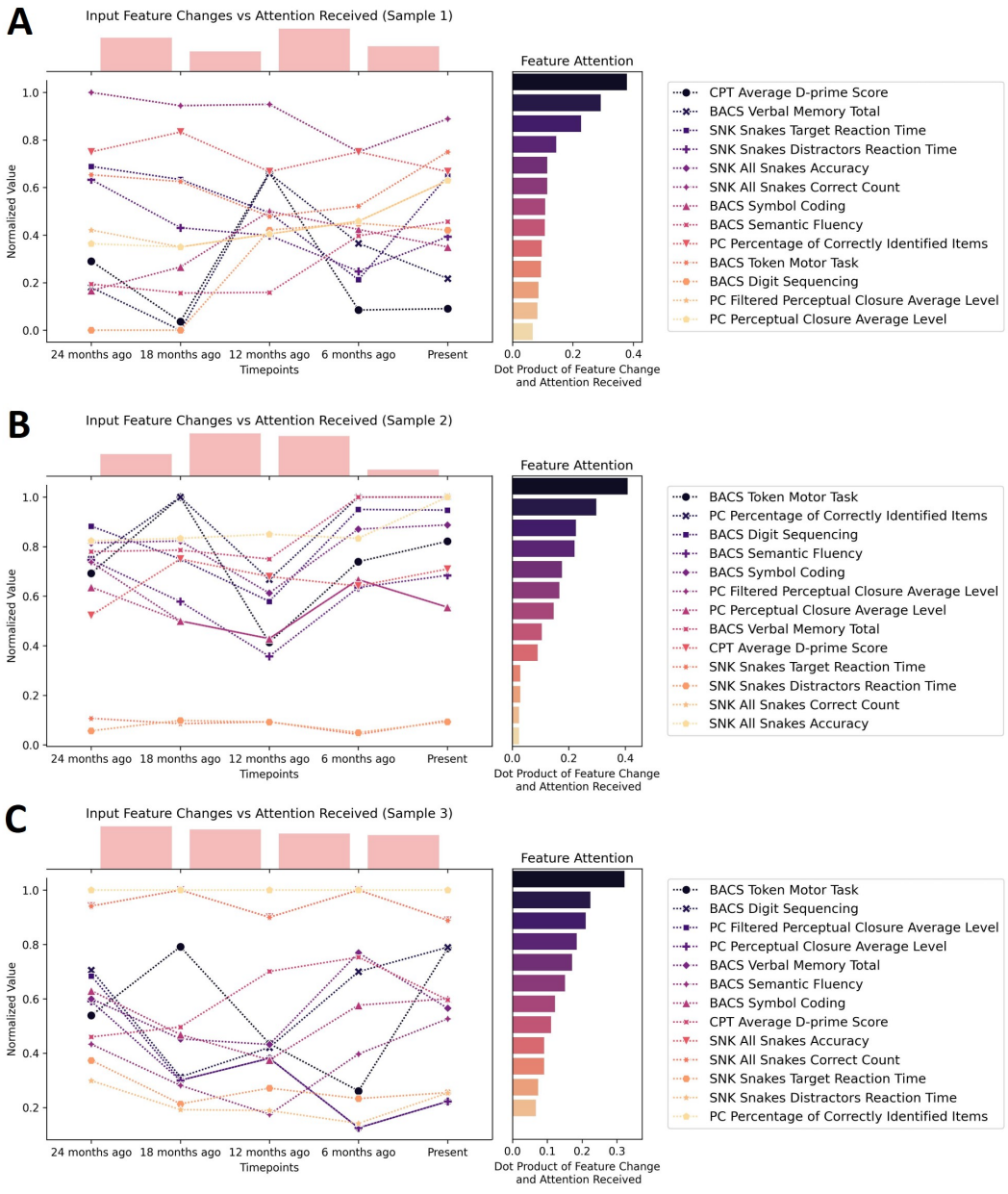


FIGURE 4.6: Personalized feature attention profiles for three different samples from the LYRIKS dataset. The left panels show the changes in input feature values over time, while the right panels display the dot product of feature changes and attention received, highlighting the most influential features for each sample.

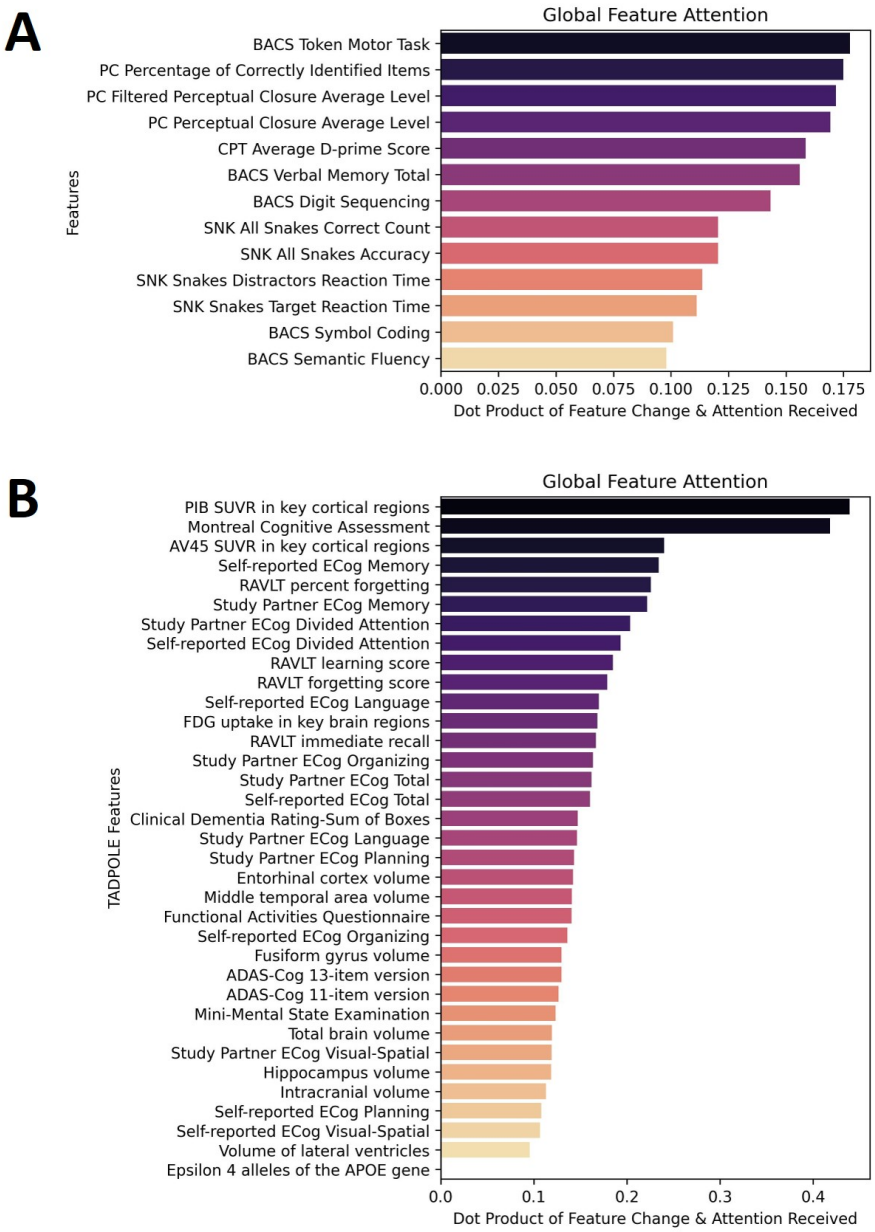


FIGURE 4.7: Global feature attention profiles for the LYRIKS dataset, showing the overall importance of each feature based on the dot product of feature changes and attention received.

Identified Items and PC Filtered Perceptual closure average level relate to visual perception and perceptual closure abilities, which are also known to be affected in UHR individuals (Doniger et al., 2001; Kimhy et al., 2007).

In the case of the TADPOLE dataset, where the aim is to predict cognitive impairment, the DAG model identifies the PIB SUVR in key cortical regions and Montreal Cognitive Assessment as most important. PIB SUVR (Pittsburgh Compound B Standardized Uptake Value Ratio) is a measure of amyloid-beta deposition in the brain, which is a hallmark pathological feature of Alzheimer's disease and is associated with cognitive decline (Engler et al., 2006). The Montreal Cognitive Assessment (MoCA) is a widely used screening tool for detecting cognitive impairment and has been shown to be sensitive to early stages of Alzheimer's disease and other dementias (Nasreddine et al., 2005).

The features identified by the DAG model align well with the current understanding of the most predictive and diagnostic markers for UHR for psychosis and cognitive impairment. This consistency with the existing literature supports the validity of the DAG model's feature weighting and highlights its potential for identifying clinically relevant factors.

However, it is important to acknowledge some limitations in the interpretation of these global explainability results. The identified features may be influenced by potential confounding factors, such as demographic variables or co-occurring conditions. Additionally, further validation of these findings in larger and more diverse samples would strengthen their generalizability.

In conclusion, the global explainability analysis using the DAG attention weights reveals key features that are highly relevant to the prediction of UHR for psychosis and cognitive impairment. These findings not only enhance our understanding of the LSM's decision-making process but also provide valuable insights for future research and clinical applications in mental health.

## 4.5 Conclusion

This chapter has dived into two pivotal components essential for advancing mental healthcare through artificial intelligence: personalized modelling and explainable AI. The focus was on developing methodologies that not only enhance predictive accuracy but also provide transparency and interpretability, thus fostering trust and practical applicability in clinical settings.

We began by establishing the necessity of personalized modelling in mental health, recognizing the complex and heterogeneous nature of mental health disorders. Conventional approaches often fail to account for individual variations, leading to less effective diagnostic and prognostic models. By leveraging transductive learning techniques, we demonstrated how personalized modelling could adapt to individual-specific data distributions and reveal underlying patterns unique to each patient's mental health trajectory. This approach addresses the critical need for tailored interventions that consider the unique aspects of each patient.

Explainability is the other cornerstone of this chapter. As AI systems grow more complex, their decision-making processes can become opaque, creating barriers to their acceptance and trust among clinicians and patients. Explainable AI techniques are designed to bridge this gap by offering transparent and interpretable explanations of model predictions. The chapter underscored the importance of explainability in mental health AI, discussing the challenges and strategies for developing models that not only perform well but also offer clear insights into their reasoning.

The central contribution of this chapter was the introduction of the Dynamic Attention Gateway (DAG) for Liquid State Machines (LSMs). This novel methodology integrates the strengths of personalized modelling with the need for explainability. The DAG-LSM approach was rigorously evaluated through empirical analyses on the LYRIKS and ADNI (TADPOLE) datasets. The results demonstrated the approach's effectiveness in predicting Ultra-High Risk (UHR) for psychosis and cognitive impairment, respectively. These findings highlight the dual benefits of the DAG-LSM method: high predictive performance and the ability to provide detailed, interpretable insights into the factors driving model decisions.

Our empirical analyses showcased how DAG-LSM could effectively identify key temporal patterns and early markers of mental health deterioration, addressing Research Question 2. Additionally, the interpretability provided by DAG-LSM directly responded to Research Question 3, which focused on designing AI models that offer clinically meaningful explanations. The ability to explain model predictions is crucial for gaining the trust of clinicians, who must understand the basis of these predictions to integrate AI effectively into their practice.

In conclusion, the DAG-LSM approach has made significant strides in achieving explainability in longitudinal data analysis, while still maintaining high accuracy for diagnosis and prognosis of mental health disorders. The insights gained from this research are poised to inform subsequent chapters, where this model will be further integrated with multimodal data. Ultimately, this work contributes to the overarching goal of improving mental healthcare through precise, reliable, and transparent AI systems that support clinical decision-making and improve patient outcomes.

# Chapter 5

# Multimodal Fusion: Integrating Diverse Data Sources

## 5.1 Introduction

The diagnosis and prognosis of mental health disorders are inherently complex processes that require comprehensive insights from a variety of data sources. Mental health conditions are influenced by a multitude of factors, including genetic, neurobiological, psychological, and environmental elements. To capture the full spectrum of these influences, integrating data from multiple modalities is essential. This chapter addresses the research objectives of integrating multimodal data sources (Objective 4 and Question 4).

Multimodal data integration involves combining information from different types of data, such as clinical assessments, neuroimaging, genetic data, and digital phenotyping from smartphones and wearables. Each of these data types provides unique and complementary insights into the mental health state of an individual. Clinical assessments offer detailed diagnostic information, neuroimaging provides structural and functional brain data, genetic data reveals underlying biological predispositions, and digital phenotyping captures real-time behavioral patterns. The challenge lies in effectively integrating these diverse data types to create a holistic and accurate representation of an individual's mental health.

Longitudinal data, which involves repeated observations of the same individuals over time, adds another layer of complexity and richness to this integration. Longitudinal studies are crucial for understanding the temporal dynamics of mental health disorders, identifying early markers of disease progression, and predicting future outcomes. The integration of multimodal longitudinal data can significantly enhance the predictive power of diagnostic and prognostic models by capturing the evolving nature of mental health conditions.

This chapter introduces the Mosaic Liquid State Machine (Mosaic LSM) method as a novel approach to integrating multimodal longitudinal data. Mosaic LSM leverages the strengths of Liquid State Machines (LSMs) in handling temporal data and combines it with advanced techniques for multimodal data fusion. The Mosaic LSM method is designed to effectively manage the high-dimensional and heterogeneous nature of multimodal data, ensuring that the integrated model can provide accurate and interpretable predictions.

The chapter is structured as follows. First, we review existing approaches to multimodal data integration, highlighting their strengths and limitations. Next, we introduce the Mosaic LSM method, detailing its architecture and the rationale behind its design. The methodology section describes the datasets used, the preprocessing steps, and the experimental setup for validating the Mosaic LSM approach. Results

from the experiments are then presented and analyzed, demonstrating the effectiveness of the Mosaic LSM method in integrating multimodal longitudinal data and improving predictive accuracy. Finally, the chapter discusses the implications of these findings for clinical practice, potential applications, and future research directions.

By developing and validating the Mosaic LSM method, this chapter makes significant contributions to the field of mental health diagnostics. The integration of diverse data sources not only enhances the predictive accuracy of diagnostic and prognostic models but also provides a more comprehensive understanding of the multifaceted nature of mental health disorders. The insights gained from this research will inform the subsequent chapters, where these integrated models will be further evaluated and applied to real-world clinical scenarios. Ultimately, this work aims to advance the field of mental health by leveraging the power of multimodal and longitudinal data integration to improve patient outcomes and support clinical decision-making.

## 5.2 Multimodal Fusion

### 5.2.1 Evolution of Multimodal Learning

The development of multimodal machine learning draws its inspiration from the complex nature of human perception and cognition. As humans, our experience of the world is never limited to a single sense. Our brains effortlessly merge visual (sight), auditory (sound), olfactory (smell), tactile (touch), and other sensory inputs, creating a dynamic, unified perception of our surroundings. This constant integration occurs as we interact with people and environments, enabling us to form a rich, comprehensive mental representation of every situation we encounter.

Neuroscience reveals that our brains contain specialized regions for processing single-sensory (unimodal) inputs — such as the auditory cortex for sound and the visual cortex for sight. Yet, it's the higher-order brain areas that are truly remarkable, as they amalgamate these unimodal signals into what cognitive scientists term 'amodal representations'. For example, our recognition of a dog is not just about seeing its shape; it also involves integrating the sensation of its fur and hearing its bark. This seamless weaving of multimodal stimuli underpins our sophisticated reasoning abilities and adaptable perception.

This concept of fluid multimodal integration, which is crucial for generalized human-like perception, has been a key source of inspiration for advancements in AI systems. Traditional machine learning models were typically limited to processing a single type of data, such as images or text. However, these singular modalities often miss the rich contextual associations that are intrinsic to human understanding. The realization that systems capable of concurrently processing multiple types of data — like images, texts, speech, and sensor information — could achieve a more comprehensive understanding of situations led to the development of more versatile AI models. By combining these diverse signals, these systems aim to capture a richer and more nuanced understanding of situations and objects, much like human cognition.

### 5.2.2 Why Multimodality Matters in Health

#### Holistic View

The integration of multimodal data in health provides a more holistic view of a patient's health status. Just as our brain processes information from various senses to form a complete picture, multimodal health data combines information from different sources — such as clinical observations, genetic profiles, lifestyle data, and environmental factors. This comprehensive approach allows healthcare professionals to see beyond the limitations of single-source data, offering a more rounded and complete understanding of a patient's health. This holistic view is essential for accurately diagnosing and treating complex health conditions where symptoms and causes span across various biological and environmental factors such as mental disorders.

#### Predictive Accuracy

Multimodal data significantly enhances the predictive accuracy of health assessments. By analyzing patterns and correlations across different types of data, healthcare providers can identify potential health risks and diseases earlier and with greater accuracy. This early detection is crucial for preventive healthcare, allowing for timely interventions that can alter the course of diseases. In conditions like heart disease or mental health disorders, where early intervention can make a significant difference, the predictive power of multimodal data becomes particularly valuable.

#### Personalized Healthcare

Personalized healthcare is one of the most significant benefits of multimodal data integration. Just as each person's biological composition is distinct and diverse, health conditions and responses vary significantly from one individual to another. These variations are influenced by a multitude of factors such as genetic makeup, lifestyle choices, environmental exposures, and more. Multimodal data allows for the customization of healthcare to the individual level, considering each patient's unique genetic makeup, lifestyle, environmental factors, and other personal health data. This tailored approach not only improves the effectiveness of treatments but also minimizes the risk of adverse reactions, leading to better overall patient care and outcomes.

In conclusion, the importance of multimodal data in healthcare cannot be overstated. It offers a more holistic and accurate understanding of health, enhances predictive capabilities, and paves the way for personalized healthcare. As we continue to advance in the field of health informatics, the integration of multimodal data will play a pivotal role in transforming healthcare into a more effective, efficient, and patient-centered service.

### 5.2.3 Challenges in Multimodal Data Integration

#### Heterogeneity of Data Types

One of the foremost challenges in multimodal data integration is the heterogeneity of data types, particularly evident in the field of health diagnostics. Data in this domain can vary widely, from structured formats like electronic health records and

clinical assessments to more complex types such as gene expression profiles, neuroimaging data, and neurocognitive test results. Each type of data presents its own set of characteristics and challenges:

- **Clinical Records**: Often in structured formats but may contain unstructured notes requiring natural language processing.
- **Gene Expression Data**: Complex and high-dimensional, necessitating sophisticated computational methods for integration.
- **Neuroimaging**: Includes various modalities like MRI, fMRI, and EEG each with unique processing and analysis requirements.
- **Neurocognitive Assessments**: Quantitative data that needs to be interpreted in the context of qualitative observations.

The diversity in data characteristics, quality, and formats poses a significant challenge in creating a unified framework that can process, analyze, and meaningfully integrate these diverse streams of data.

### Synchronization and Temporal Alignment

Synchronization and temporal alignment are pivotal in ensuring the coherence and efficacy of multimodal data analysis. Different data types often come with varying sampling rates and time stamps. For instance, clinical assessments, often requiring the presence of a clinician, are typically gathered at specific, less frequent intervals, whereas neurocognitive data, which can often be self-administered, might be collected more frequently and at varied times. Meanwhile, neuroimaging data is acquired during scheduled sessions, and wearable device data provides a constant stream, illustrating the challenge of aligning these diverse data types that are collected at different times and under varying conditions.

Aligning these disparate data streams to ensure temporal consistency and contextual relevance is crucial. This is particularly important in longitudinal studies focusing on the progression of mental health symptoms over time, where the timing of data points can significantly impact the interpretation and accuracy of the analysis.

### Dealing with Missing and Incomplete Data

The issue of missing or incomplete data is a common obstacle in multimodal datasets. This can arise from various factors such as sensor malfunctions, patient non-compliance, or gaps in data collection. The presence of missing data in one or more modalities can significantly impact the performance and reliability of AI models. Addressing this challenge involves:

- **Imputation Techniques**: Employing methods to estimate missing values based on available data, though this may introduce biases or inaccuracies.
- **Robust Model Architectures**: Designing AI models that can effectively handle incomplete data, either by being resilient to missing data or by using techniques to infer missing information.
- **Data Augmentation**: Generating synthetic data to compensate for missing modalities, which must be done carefully to avoid misleading the model.

Handling missing and incomplete data is essential for maintaining the accuracy, reliability, and overall integrity of multimodal data analysis, especially in health diagnostics.

### 5.2.4 Principal Approaches in Multimodal Fusion

The fusion of multimodal data is a critical step in constructing AI systems that can understand and interpret complex, real-world phenomena at a level comparable to human cognition. There are several principal approaches to multimodal fusion, each with its own methodologies, advantages, and challenges. These strategies can be broadly categorized into early fusion, late fusion, hybrid fusion, and intermediate fusion which consists of model-based and translation-based approaches.

#### Early Fusion (Feature-level)

Early fusion involves integrating data at the feature level before any model processing occurs. This method combines all modalities into a single feature vector, allowing the model to learn intermodal interactions from the very beginning of the process. Early fusion is particularly effective when the relationships between modalities are strong and the data is well synchronized. However, it requires all modalities to be present at the time of training and can be sensitive to misalignments or discrepancies in the data. Additionally, it may lead to complex and high-dimensional feature spaces that are challenging to manage, emphasizing the need for careful feature selection and dimensionality reduction.

Some examples of early fusion techniques include integration of visual, linguistic, and auditory features for sentiment analysis (Rosas, Mihalcea, and Morency, 2013), combination of physiological measures with facial videos for improved emotion recognition (Monkaresi, Hussain, and Calvo, 2012), and the integration of gene expression and DNA methylation data for prediction of Alzheimer's disease (Park, Ha, and Park, 2020).

#### Late Fusion (Decision-level)

Late fusion integrates decisions or predictions from separate models, each processing a different modality. This method is robust to misalignments and can handle missing data in one or more modalities by relying on the remaining modalities' outputs. Late fusion is particularly beneficial when the modalities are quite distinct or have different sampling rates. The approach to combining model outputs, whether through majority voting, weighted schemes, or other decision strategies, is a crucial aspect of late fusion.

Late fusion has been effectively applied in integrating heterogeneous modalities, such as imaging and medical records for pulmonary embolism diagnosis (Huang et al., 2020), and in combining separate classifiers for image and text in multimedia content categorization (Zadeh et al., 2016).

#### Intermediate Fusion (Representation-level)

Intermediate fusion, or representation-level fusion, occurs after some initial processing of the individual modalities but before the final decision-making stage. This approach allows for the integration of more abstract representations of data, typically after some transformation or feature extraction has been applied. It offers a middle ground between the deep integration of early fusion and the modality independence

| Approach | Advantages | Limitations |
|---|---|---|
| Early Fusion | • Integrates raw data features early to capture rich, low-level interactions between modalities<br>• Reduces model complexity by integrating all modalities at the beginning | • May lead to high-dimensional feature space, increasing complexity<br>• Requires precise synchronization of all modalities, limiting flexibility<br>• Tendency to miss modality-specific patterns |
| Late Fusion | • Can handle different modalities independently, offering robustness against missing or asynchronous data<br>• Simplifies the integration by focusing on decisions from individual models | • May overlook complex inter-modal interactions<br>• Overall performance heavily reliant on the accuracy of each modality's model |
| Intermediate Fusion | • Offers a balance between learning modality-specific patterns and cross-modal interactions.<br>• Provides flexibility in choosing how and which features to combine | • Increased design complexity.<br>• Dependent on the quality and relevance of intermediate representations |
| Hybrid Fusion | • Utilizes the advantages of other fusion methods, potentially leading to superior performance<br>• Flexible design to suit specific needs and characteristics of the task and modalities involved | • May lead to overfitting due to the large number of parameters and the complex structure<br>• Requires extensive knowledge and resources |

TABLE 5.1: Comparison of multimodal fusion approaches

of late fusion. It requires careful design to ensure that the representations are meaningful and compatible for subsequent fusion.

Within intermediate fusion, transformation-based and model-based strategies are prevalent:

1. **Transformation-based Approaches**: These involve translating all modalities into a common representational space before fusion. The aim is to convert disparate data types into a format that is more amenable to integration, often using dimensionality reduction techniques or learned embeddings (Kiros,

Salakhutdinov, and Zemel, 2014).

2. **Model-based Approaches**: Such approaches design specific architectures that can handle and integrate multiple data types. This might include adaptations of neural network architectures, such as CNNs or RNNs, to multimodal data, or employing advanced techniques like attention mechanisms and graph neural networks to manage and fuse representations (Ngiam et al., 2011).

Examples of intermediate fusion include deep canonical correlation analysis to align and combine textual and visual features for visual question answering (Andrew et al., 2013), and employing cross-modal attention mechanisms in neural networks to integrate language and vision for image captioning (Xu et al., 2015).

#### Hybrid Fusion

Hybrid fusion techniques combine different fusion strategies—early, intermediate, and late—to leverage the advantages of each. This approach aims to balance interaction modeling with flexibility in handling data variations and inconsistencies, adapting to the trade-offs between capturing complex inter-modal interactions and handling data variability.

Examples include employing a combination of feature-level and decision-level fusion for robust multimedia event detection (Lan et al., 2014), and integrating audio and video data using early and late fusion for speaker identification (Wu, Cai, and Meng, 2005).

Each approach to multimodal fusion has its unique set of challenges and considerations. The choice of strategy often depends on the specific characteristics of the data, the objectives of the task, and the available computational resources. An overview of the advantages and limitations of each type of fusion approach is presented in Table 5.1. As research progresses, the trend is towards more sophisticated intermediate model-based approaches that can adaptively manage the complexities of multimodal data, leading to more accurate, robust, and flexible AI systems.

### 5.2.5 Feasibility Analysis of Clinical, Genetic, Cognitive, and Social Modalities in UHR Prediction

In the section, we analyse effectiveness of different data modalities—each representing a distinct aspect of human health and behavior—in predicting the risk of developing severe mental health conditions. The focus is on identifying individuals at Ultra-High Risk (UHR) of psychosis, with the ultimate goal of facilitating early interventions. By leveraging various machine learning models, we assess the predictive power of clinical tests, gene expression, cognitive assessments, and social functioning, recorded in the LYRIKS dataset, across various future time points. This enables us to understand not just the immediate implications of these modalities but also their long-term relevance. From immediate predictions (0 months ahead) to forecasts extending up to 24 months, this analysis provides insights into which modalities are most indicative of UHR states.

For these experiments, given the context of early identification and monitoring in mental health, this work places emphasis on evaluation metrics that capture meaningful risk transitions. In particular, the accuracy on label-changing samples (acc-changed) metric is used to measure the model's ability to correctly predict a change in state between timepoints — the core goal in longitudinal prediction. This metric,

alongside MCC and F1-score, offers a balanced view of both overall performance and transition-specific sensitivity.

**Study Design**

1. **Objective**: The focus of this study is to leverage the LYRIKS dataset for early diagnosis and prognosis of Ultra-High Risk (UHR) for psychosis among youths. The dataset facilitates the analysis of clinical, genetic, cognitive, and social modalities individually and in combination, to predict the risk status of individuals at future intervals of 0, 6, 12, 18, and 24 months ahead. Hence the task will be binary classification, to predict the present or future label of 'healthy' or 'at-risk' for each individual using cross-sectional data (single time-point).

2. **Models**: Four machine learning models were selected for the analysis: Support Vector Machine (SVM), Gradient Boosting Decision Trees (GBDT), Logistic Regression (LogReg), and Multilayer Perceptron (MLP). These models were chosen for their diverse learning strategies—ranging from linear assumptions in LogReg to the non-linear capabilities of MLPs and the ensemble approach of GBDT—providing a broad spectrum of analytical capabilities to cater to the distinct characteristics of each modality.

3. **Validation**: To ensure the robustness and reliability of the predictive models, k-fold cross-validation (CV) strategy is employed with $k = 5$. This technique involves partitioning the dataset into $k$ subsets, sequentially using one subset for testing and the remaining for training. This is done to minimize the risk of overfitting on a particular train-test split and provide a more accurate estimate of the model's performance on unseen data.

4. **Performance Metrics**: Given the data imbalance between the healthy and at-risk classes, we selected performance metrics that provide a balanced view of the models' effectiveness across both classes. These metrics include accuracy, which gives the overall correctness of the model, Matthew's Correlation Coefficient (MCC), offering a balanced measure even when class distributions are uneven, Area Under the Receiver Operating Characteristic Curve (AUC), assessing the model's ability to distinguish between the classes, and accuracy of labels that changed status (Acc-changed), to specifically evaluate the model's performance in predicting transitions in risk status over time.

5. **Feature Selection**: For the gene expression data, since the dimensionality was high, the Filter and Wrapper Stacking Ensemble (FWSE) method (presented in Chapter 3) was employed, supplemented by ANOVA and Signal-to-Noise Ratio (SNR) as filter methods and Recursive Feature Elimination with Logistic Regression (RFELR) as the wrapper method, to select the top 500 features within each fold of the cross-validation. This number was chosen to ensure a broad yet manageable dataset that allows for comprehensive analysis without overfitting or undue computational burden. This design underscores our objective to evaluate the genetic modality's predictive capacity as a whole, rather than pinpointing a narrow set of biomarkers.

6. **Hyperparameter Optimization**: Hyperparameters for all classifiers were tuned using a grid search on the training set within each fold of the cross-validation procedure. The selection criterion was Matthews Correlation Coefficient (MCC),

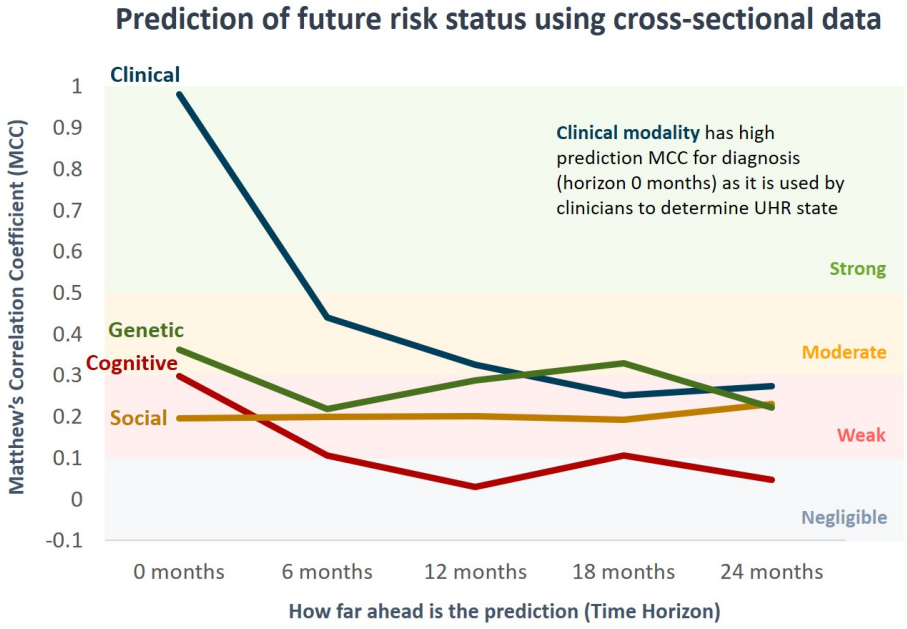


FIGURE 5.1: Top model performance of each modality at various time horizons for prediction of UHR status using cross-sectional data

evaluated on a 20% subset of the training data, which was held out as a validation set during tuning. This process was repeated independently within each fold to ensure that hyperparameter selection did not involve any information from the corresponding test fold.

7. **Computational Environment**: All experiments were implemented in `Python 3.8`, using libraries including `scikit-learn`, `XGBoost`, `pandas`, `NumPy`, `PyTorch`, and `NeuCubePy` for model development and evaluation. Visualisations were produced using `Matplotlib` and `Seaborn`. Experiments were executed on an institutional machine equipped with an `Intel Core i5 CPU` and `16 GB RAM`. GPU acceleration was not required for any of the experiments.

8. **Data Integrity**: The primary input for the models is the cross-sectional data collected at various points throughout the study. For the prediction of future risk status, we treat the data from each time point for a participant as a separate sample, while ensuring that the integrity of the temporal order is maintained to avoid lookahead bias. Specifically, when predicting the risk status at future intervals, we ensure that only data available up to the prediction point is used, respecting the temporal dynamics inherent in longitudinal studies.

### Clinical Modality

The exploration of the clinical modality's performance in predicting Ultra-High Risk (UHR) outcomes across various future time points, as summarized in Table 5.2, unveils significant insights into the role of clinical assessments in the early identification of psychosis.

Clinicians primarily rely on these clinical tests to determine UHR labels. Therefore, it is logical that the performance metrics at the immediate horizon (0 Months

TABLE 5.2: Performance of clinical modality for predicting UHR outcomes at different time horizons using cross-sectional data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Clinical | SVM | 98.88% | 0.9775 | 0.9893 | - |
| | | GBDT | 99.07% | 0.9814 | 0.9914 | - |
| | | LogReg | 99.07% | 0.9814 | 0.9914 | - |
| | | MLP | 98.88% | 0.9775 | 0.9893 | - |
| 6 Months Ahead | Clinical | SVM | 72.20% | 0.4411 | 0.7226 | 21.74% |
| | | GBDT | 66.59% | 0.3060 | 0.6480 | 37.39% |
| | | LogReg | 71.03% | 0.4033 | 0.6992 | 33.91% |
| | | MLP | 70.56% | 0.3920 | 0.6924 | 39.13% |
| 12 Months Ahead | Clinical | SVM | 69.78% | 0.3253 | 0.6547 | 61.48% |
| | | GBDT | 66.98% | 0.2476 | 0.6126 | 61.48% |
| | | LogReg | 69.16% | 0.3073 | 0.6444 | 62.22% |
| | | MLP | 68.58% | 0.3034 | 0.6467 | 58.52% |
| 18 Months Ahead | Clinical | SVM | 65.42% | 0.2357 | 0.6179 | 65.46% |
| | | GBDT | 67.76% | 0.2509 | 0.6162 | 75.63% |
| | | LogReg | 66.82% | 0.2244 | 0.6031 | 77.31% |
| | | MLP | 67.29% | 0.2461 | 0.6162 | 76.47% |
| 24 Months Ahead | Clinical | SVM | 65.42% | 0.1783 | 0.5820 | 69.01% |
| | | GBDT | 70.09% | 0.2740 | 0.6172 | 73.24% |
| | | LogReg | 65.42% | 0.1783 | 0.5820 | 69.01% |
| | | MLP | 65.42% | 0.1993 | 0.5957 | 66.20% |

Ahead) are notably high across all models. The near-perfect accuracy at this point—with GBDT and LogReg reaching an accuracy of 99.07%—reflects the direct correlation between the clinical assessments used in the study and the criteria for determining UHR status.

However, as the prediction horizon extends, a decline in performance metrics is observed. At 6 Months Ahead, MCC drops, with SVM showing the highest MCC of 0.4411 This trend of diminishing MCC continues further at 12 and 18 months ahead, underscoring the challenges in long-term prediction using solely clinical data. The 24 months ahead predictions exhibit a slight improvement in GBDT's performance, with a MCC of 0.2740, suggesting some resilience in predictive capability over an extended period.

The decline in predictive performance over time is mirrored by the corresponding accuracy and AUC metrics, which decrease as the prediction horizon extends. Interestingly, the accuracy for changed status (Acc-changed), which measures the model's ability to correctly predict changes in UHR status over time, also shows variability, with an increase in some models as the horizon extends. This could suggest that while overall prediction accuracy decreases, the models may still retain some capability to detect transitions in UHR status, potentially highlighting the evolving nature of clinical indicators over time.

This analysis reveals the critical importance of temporal dynamics in the predictive modeling of psychosis risk. While clinical data exhibit strong predictive power in the short term, their efficacy diminishes over longer horizons, highlighting the necessity for incorporating additional data modalities or longitudinal data to maintain predictive performance. Moreover, the variability in Acc-changed suggests that certain clinical indicators may become more or less salient as the risk of psychosis development evolves, underscoring the need for dynamic models that can adapt to

TABLE 5.3: Performance of cognitive modality for predicting UHR outcomes at different time horizons using cross-sectional data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Cognitive | SVM | 64.30% | 0.2842 | 0.6425 | - |
| | | GBDT | 65.05% | 0.2992 | 0.6500 | - |
| | | LogReg | 64.49% | 0.2838 | 0.6417 | - |
| | | MLP | 62.62% | 0.2479 | 0.6241 | - |
| 6 Months Ahead | Cognitive | SVM | 56.79% | 0.1068 | 0.5526 | 57.39% |
| | | GBDT | 54.21% | 0.0399 | 0.5191 | 52.17% |
| | | LogReg | 54.44% | 0.0443 | 0.5211 | 50.43% |
| | | MLP | 54.91% | 0.0441 | 0.5204 | 48.70% |
| 12 Months Ahead | Cognitive | SVM | 56.39% | 0.0305 | 0.5147 | 64.44% |
| | | GBDT | 55.76% | -0.0144 | 0.4934 | 68.15% |
| | | LogReg | 63.55% | 0.0223 | 0.5018 | 87.41% |
| | | MLP | 55.47% | 0.0039 | 0.5019 | 61.48% |
| 18 Months Ahead | Cognitive | SVM | 61.22% | 0.1068 | 0.5443 | 69.75% |
| | | GBDT | 59.35% | 0.0445 | 0.5205 | 70.59% |
| | | LogReg | 56.31% | 0.0403 | 0.5206 | 57.83% |
| | | MLP | 65.89% | 0.0942 | 0.5068 | 92.44% |
| 24 Months Ahead | Cognitive | SVM | 66.36% | 0.0000 | 0.5000 | 70.42% |
| | | GBDT | 61.68% | 0.0465 | 0.5196 | 81.69% |
| | | LogReg | 66.36% | 0.0000 | 0.5000 | 70.42% |
| | | MLP | 66.36% | 0.0000 | 0.5000 | 70.42% |

these changes over time.

### Cognitive Modality

The cognitive modality's performance in predicting Ultra-High Risk (UHR) outcomes is detailed in Table 5.3. At the outset, the performance metrics for the immediate horizon (0 Months Ahead) indicate a modest predictive capability with the highest Matthews Correlation Coefficient (MCC) achieved by GBDT at 0.2992. Unlike the clinical modality, the cognitive modality shows less predictive accuracy at this baseline, highlighting the complexity of using cognitive data alone for immediate UHR outcome prediction.

As the prediction horizon extends to 6, 12, 18, and 24 months ahead, a general trend of diminishing performance is observed across all models, reflected in both MCC and AUC metrics. This trend underscores the challenges inherent in long-term prediction using cognitive data, which may not capture the dynamic changes in risk status over time as effectively as clinical measures

This analysis highlights several key insights into the utility of cognitive data for predicting psychosis risk. First, while cognitive assessments provide valuable information, their predictive power, especially in the short term, is limited compared to clinical assessments. The modest MCC values at the baseline and their decline over time suggest that cognitive measures may be more reflective of general risk factors rather than immediate indicators of UHR status.

Second, the variability in Acc-changed across time points suggests that cognitive changes relevant to UHR status may manifest more clearly or become obscured over time. This could imply that cognitive indicators have a temporal aspect in their

TABLE 5.4: Performance of social modality for predicting UHR outcomes at different time horizons using cross-sectional data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Social | SVM | 58.28% | 0.1563 | 0.5772 | - |
| | | GBDT | 60.35% | 0.1957 | 0.5937 | - |
| | | LogReg | 58.28% | 0.1576 | 0.5781 | - |
| | | MLP | 57.24% | 0.1399 | 0.5699 | - |
| 6 Months Ahead | Social | SVM | 62.50% | 0.1999 | 0.5617 | 78.13% |
| | | GBDT | 59.05% | 0.1410 | 0.5683 | 64.06% |
| | | LogReg | 60.78% | 0.1444 | 0.5411 | 58.27% |
| | | MLP | 61.21% | 0.1554 | 0.5564 | 62.50% |
| 12 Months Ahead | Social | SVM | 66.67% | 0.2018 | 0.5877 | 80.00% |
| | | GBDT | 64.37% | 0.1621 | 0.5741 | 64.00% |
| | | LogReg | 62.64% | 0.0808 | 0.5333 | 78.67% |
| | | MLP | 66.09% | 0.1337 | 0.5439 | 65.33% |
| 18 Months Ahead | Social | SVM | 66.36% | 0.1909 | 0.5881 | 82.52% |
| | | GBDT | 64.67% | 0.1923 | 0.5955 | 73.13% |
| | | LogReg | 63.79% | 0.1094 | 0.5486 | 76.12% |
| | | MLP | 62.07% | 0.1390 | 0.5695 | 67.16% |
| 24 Months Ahead | Social | SVM | 67.24% | 0.1757 | 0.5792 | 82.50% |
| | | GBDT | 68.97% | 0.2312 | 0.6069 | 82.50% |
| | | LogReg | 62.07% | 0.0863 | 0.5417 | 75.00% |
| | | MLP | 63.79% | 0.1670 | 0.5847 | 72.50% |

predictive utility, which might be enhanced by incorporating longitudinal data or integrating cognitive data with other modalities.

Lastly, the persistence of some predictive capability, as seen in Acc-changed in certain models at longer horizons, underscores the potential of cognitive data to contribute to a multimodal predictive framework. This suggests that while cognitive data alone may face limitations in long-term UHR outcome prediction, its integration with other modalities like clinical, social, and genetic data could provide a more comprehensive and dynamic model for psychosis risk prediction.

### Social Modality

The exploration of the social modality's role in predicting Ultra-High Risk (UHR) outcomes across various future time points, as detailed in Table 5.4, offers a distinct perspective on the predictive utility of social functioning metrics in the context of psychosis risk.

A noteworthy observation across all prediction horizons is the relatively stable performance metrics. Unlike other modalities where a significant drop or improvement in predictive accuracy might be expected as the prediction horizon extends, the social modality exhibits a consistency in predictive capability. The highest Matthews Correlation Coefficient (MCC) observed is 0.2312 by GBDT at 24 Months Ahead, suggesting a modest but consistent relationship between social functioning and UHR status across time. The stability in performance is further reflected in the Acc-changed metric, which indicates the models' ability to accurately predict changes in UHR status over time.

This consistent performance pattern offers several insights. First, the stable predictive capability of social functioning data across all time horizons underscores the

inherent value of social metrics in UHR outcome prediction. It suggests that social functioning, while not the strongest predictor on its own, provides a steady signal that is relevant across different stages in the progression towards psychosis.

Second, when combined with clinical, cognitive, or genetic data, social functioning metrics can enhance the holistic assessment of psychosis risk, providing a more nuanced and comprehensive prediction model. Lastly, the predictive utility of social functioning across time points highlights the potential for developing targeted intervention strategies that address social factors as part of a comprehensive approach to psychosis risk management.

In summary, the analysis of the social modality in predicting UHR outcomes reveals a consistently modest yet significant predictive utility across various future time points. This stability in performance metrics suggests that social functioning plays a continuous and reliable role in the assessment of psychosis risk, emphasizing the importance of integrating social metrics into holistic, multimodal prediction models for early psychosis risk detection.

**Genetic Modality**

TABLE 5.5: Performance of genetic modality for predicting UHR outcomes at different time horizons using cross-sectional data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Genetic | SVM | 76.80% | 0.3628 | 0.6165 | - |
| | | GBDT | 75.47% | 0.3401 | 0.6430 | - |
| | | LogReg | 75.73% | 0.3443 | 0.6421 | - |
| | | MLP | 75.73% | 0.3376 | 0.6338 | - |
| 6 Months Ahead | Genetic | SVM | 77.60% | 0.1780 | 0.5413 | 85.71% |
| | | GBDT | 74.00% | 0.1264 | 0.5480 | 74.29% |
| | | LogReg | 77.20% | 0.1747 | 0.5610 | 77.14% |
| | | MLP | 61.21% | 0.2185 | 0.5748 | 77.14% |
| 12 Months Ahead | Genetic | SVM | 80.40% | 0.2879 | 0.5998 | 76.19% |
| | | GBDT | 79.20% | 0.2310 | 0.5784 | 78.57% |
| | | LogReg | 80.00% | 0.2751 | 0.5973 | 76.19% |
| | | MLP | 76.00% | 0.2196 | 0.5995 | 76.19% |
| 18 Months Ahead | Genetic | SVM | 81.60% | 0.3298 | 0.6300 | 90.00% |
| | | GBDT | 80.80% | 0.3267 | 0.6400 | 90.00% |
| | | LogReg | 80.80% | 0.3267 | 0.6400 | 90.00% |
| | | MLP | 80.00% | 0.3273 | 0.6500 | 83.33% |
| 24 Months Ahead | Genetic | SVM | 79.20% | 0.2228 | 0.5850 | 86.67% |
| | | GBDT | 77.60% | 0.1572 | 0.5600 | 83.33% |
| | | LogReg | 76.80% | 0.1937 | 0.5850 | 76.67% |
| | | MLP | 77.60% | 0.2100 | 0.5900 | 83.33% |

The genetic modality's contribution to predicting Ultra-High Risk (UHR) outcomes across different future time points, as captured in Table 5.5, provides insight into the predictive power of genetic information in the context of psychosis risk assessment.

One of the most striking findings from this analysis is the relatively robust and improving performance of genetic modality predictions over time. Unlike the clinical, cognitive, and social modalities, the genetic modality demonstrates a unique

trend: its predictive performance, particularly in terms of Matthews Correlation Coefficient (MCC) and Accuracy, tends to improve or remain stable as the prediction horizon extends. For instance, the MCC for SVM shows a notable increase from 0.3628 at the baseline (0 Months Ahead) to 0.3298 at 18 Months Ahead, reflecting a significant enhancement in the model's predictive capability over time.

This trend suggests that genetic factors may have a more pronounced or stable influence on the progression towards psychosis over longer periods, highlighting the potential of genetic data in contributing to long-term risk assessment strategies. The increase in Acc-changed across models and time points further underscores the utility of genetic information in detecting changes in UHR status, with notable improvements observed as the prediction horizon extends to 18 and 24 months.

The genetic modality's increasing or stable predictive performance over time emphasizes the enduring impact of genetic factors on psychosis risk. This contrasts with other modalities, where predictive power may diminish over longer horizons, suggesting that genetic information holds particular value in long-term risk assessment.

The relatively high and improving Acc-changed rates indicate that genetic factors are effective in signaling shifts in UHR status over time. This capability is crucial for developing interventions and monitoring strategies that adapt to an individual's evolving risk profile.

The distinct performance pattern of the genetic modality highlights its potential to complement other data types in multimodal predictive models. By integrating genetic information with clinical, cognitive, and social data, it may be possible to achieve a more comprehensive and accurate prediction of psychosis risk.

In conclusion, the analysis of the genetic modality in predicting UHR outcomes reveals its significant and possibly increasing contribution to understanding and forecasting psychosis risk over time. The stable to improving performance metrics across prediction horizons underscores the value of incorporating genetic information into predictive models, not just for immediate risk assessment but importantly for long-term monitoring and intervention planning. This finding advocates for a more pronounced role of genetic data in the multimodal prediction of psychosis, emphasizing the need for ongoing research and development in this area to harness the full potential of genetic insights in mental health diagnostics.

### 5.2.6 Analysing Effect of Using Longitudinal Data for UHR Prediction

In this section, we extend our investigation to analyze the effectiveness of multiple data modalities—encompassing clinical, genetic, cognitive, and social factors—in predicting the Ultra-High Risk (UHR) of developing severe mental health conditions over time. Unlike the cross-sectional analysis, this longitudinal study leverages data from three timepoints across one year to understand the temporal dynamics and predictive power of these modalities in forecasting psychosis risk. This approach allows us to capture the evolution of risk factors and their interactions over time, providing a more nuanced understanding of the progression towards psychosis.

#### Study Design

1. **Objective**: This study aims to harness the longitudinal data within the LYRIKS dataset to predict UHR for psychosis across multiple future timepoints (0, 6, and 12 months ahead) using three timepoint data collected over one year. The

task remains a binary classification challenge, focusing on identifying individuals as 'healthy' or 'at-risk' based on the evolution of their clinical, genetic, cognitive, and social profiles over time.

2. **Models**: To address the longitudinal nature of the data and capture the complex temporal dynamics, we incorporate a diverse set of machine learning and deep learning models:
    - **Support Vector Machine (SVM)** and **Gradient Boosting Decision Trees (GBDT)** for their robustness in handling high-dimensional data and their ability to model complex, non-linear relationships.
    - **Liquid State Machine (LSM)**, a type of spiking neural network, chosen for its capacity to process temporal patterns in dynamically changing environments.
    - **Long Short-Term Memory (LSTM)** networks, a form of recurrent neural network (RNN), ideal for modeling long and short-term dependencies in time-series data.
    - **Transformers**, which leverage self-attention mechanisms to capture temporal relationships at various scales, providing a powerful tool for understanding the sequential nature of the data.
3. **Validation**: The same 5-fold cross-validation strategy is employed. This is crucial for assessing the models' generalizability and minimizing overfitting. This strategy ensures each fold is used as a test set once while the remaining folds serve as the training set, providing a comprehensive evaluation across the dataset. In the longitudinal context, predictions for a given timepoint were always made using data from previous timepoints. This preserved the temporal order and avoiding look-ahead bias.
4. **Performance Metrics**: Accuracy, Matthew's Correlation Coefficient (MCC), Area Under the Receiver Operating Characteristic Curve (AUC), and accuracy for changed status (Acc-changed) remain our primary metrics. However, in the longitudinal context, special emphasis is placed on Acc-changed to evaluate the models' ability to capture the dynamics of risk status evolution over time.
5. **Feature Selection**: The longitudinal analysis requires careful consideration of feature selection to handle the increased complexity of temporal data. To address this, each timepoint of an individual was considered as a distinct sample (only for feature selection) with the label set to time horizon being predicted. This enabled the identification and selection of features that maintain their predictive relevance across different stages. The Filter and Wrapper Stacking Ensemble (FWSE) method was employed, integrating ANOVA and Signal-to-Noise Ratio (SNR) as filter methods, alongside Recursive Feature Elimination with Logistic Regression (RFE-LR) as the wrapper method with pruning factor set to 0.75. Top 500 features were selected within each fold, for each time horizon task. By adapting feature selection to the temporal dimension, this approach ensures that the most predictive features for each timepoint are utilized.
6. **Hyperparameter Optimization**: Hyperparameters for all classifiers were tuned using a grid search on the training set within each fold of the cross-validation

TABLE 5.6: Performance of clinical modality for predicting UHR outcomes at different time horizons using longitudinal data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Clinical | SVM | 99.38% | 0.9866 | 0.9915 | - |
| | | GBDT | 99.38% | 0.9866 | 0.9915 | - |
| | | LSM | 88.47% | 0.7539 | 0.8802 | - |
| | | LSTM | 97.19% | 0.9395 | 0.9652 | - |
| | | Transformer | 96.57% | 0.9259 | 0.9585 | - |
| 6 Months Ahead | Clinical | SVM | 78.97% | 0.5203 | 0.7469 | 27.03% |
| | | GBDT | 78.04% | 0.4954 | 0.7302 | 27.03% |
| | | LSM | 73.83% | 0.3919 | 0.6790 | 48.65% |
| | | LSTM | 73.83% | 0.4216 | 0.7108 | 27.03% |
| | | Transformer | 73.36% | 0.4171 | 0.7104 | 32.43% |
| 12 Months Ahead | Clinical | SVM | 74.77% | 0.4173 | 0.7003 | 36.00% |
| | | GBDT | 74.77% | 0.4060 | 0.6866 | 36.00% |
| | | LSM | 71.96% | 0.3229 | 0.6381 | 68.00% |
| | | LSTM | 71.03% | 0.3301 | 0.6585 | 44.00% |
| | | Transformer | 73.83% | 0.3989 | 0.6932 | 48.00% |

procedure. The selection criterion was Matthews Correlation Coefficient (MCC), evaluated on a 20% subset of the training data, which was held out as a validation set during tuning. This process was repeated independently within each fold to ensure that hyperparameter selection did not involve any information from the corresponding test fold.

7. **Computational Environment**: All experiments were implemented in `Python 3.8`, using libraries including `scikit-learn`, `XGBoost`, `pandas`, `NumPy`, `PyTorch`, and `NeuCubePy` for model development and evaluation. Visualisations were produced using `Matplotlib` and `Seaborn`. Experiments were executed on an institutional machine equipped with an `Intel Core i5 CPU` and `16 GB RAM`. GPU acceleration was not required for any of the experiments.

8. **Data Integrity**: Special attention is given to maintaining data integrity and preventing lookahead bias. This involves ensuring that predictions at any given timepoint are made using only information available up to that point, thereby preserving the chronological order of events and data authenticity.

### Clinical

The longitudinal analysis of clinical modality for predicting UHR outcomes enhances our understanding of the temporal dynamics in psychosis risk prediction. The immediate high performance metrics (0 Months Ahead) with SVM and GBDT achieving near-perfect accuracy, as observed in the cross-sectional study, are mirrored in the longitudinal approach, affirming the efficacy of clinical assessments in the short-term identification of UHR.

The longitudinal data analysis for predicting future UHR outcomes with clinical modality reveals significant improvements in predictive performance at 6 and 12 months ahead, as reflected in the 8% average increase in MCC and 4% average

TABLE 5.7: Performance of cognitive modality for predicting UHR outcomes at different time horizons using longitudinal data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Cognitive | SVM | 59.81% | 0.1961 | 0.6018 | - |
| | | GBDT | 60.44% | 0.2022 | 0.6049 | - |
| | | LSM | 63.96% | 0.2511 | 0.6278 | - |
| | | LSTM | 62.54% | 0.1327 | 0.5541 | - |
| | | Transformer | 59.01% | 0.1643 | 0.5844 | - |
| 6 Months Ahead | Cognitive | SVM | 61.68% | 0.1747 | 0.5893 | 62.16% |
| | | GBDT | 64.95% | 0.1698 | 0.5761 | 62.16% |
| | | LSM | 59.26% | 0.1594 | 0.5834 | 67.74% |
| | | LSTM | 62.96% | 0.1246 | 0.5567 | 54.84% |
| | | Transformer | 61.90% | 0.1682 | 0.5852 | 58.06% |
| 12 Months Ahead | Cognitive | SVM | 64.21% | 0.1588 | 0.5764 | 61.11% |
| | | GBDT | 65.26% | 0.1637 | 0.5759 | 55.56% |
| | | LSM | 65.42% | 0.2416 | 0.6230 | 64.00% |
| | | LSTM | 68.42% | 0.1912 | 0.5743 | 66.67% |
| | | Transformer | 66.32% | 0.1703 | 0.5754 | 66.67% |

increase in AUC, when compared to the cross-sectional analysis. The enhanced performance indicates a more accurate model ability to predict future risk states, highlighting the advantage of using longitudinal data for understanding the progression of psychosis risk over time. This improvement underscores the importance of temporal analysis in clinical assessments for UHR, suggesting that longitudinal approaches offer a more nuanced and effective method for early psychosis intervention strategies.

The LSM model, compared to the others, demonstrates better ability to predict changes in UHR status (Acc-changed) at later time points, suggesting its potential in capturing the progressive changes in clinical indicators over time.

**Cognitive**

Analyzing the trends from the longitudinal analysis of the cognitive modality for UHR prediction, we observe that the performance across models varies. LSM stands out slightly better in the 0-month and 12-month ahead predictions, indicating its potential in capturing cognitive changes over time. Both MCC and AUC metrics generally improve in the longitudinal analysis compared to cross-sectional data over time, especially at 12 months ahead, where MCC improves by 21% and AUC by 11%. Unlike the clinical modality, the Acc-changed metric for the cognitive modality aligns closely with the accuracy values, suggesting that the models when using cognitive modality do not exhibit a bias toward predicting the current state of the individual as their future state.

The analysis also reveals that tracking how cognitive function changes, is a better indicator of UHR status, as seen by the improvement of predictive performance over future time horizons, rather than static cognitive states. This aligns with existing research indicating that cognitive decline or variability may be a critical marker or mediator in the development of psychosis, emphasizing the importance of monitoring cognitive changes longitudinally for early intervention and tailored treatment strategies in at-risk populations.

TABLE 5.8: Performance of social modality for predicting UHR outcomes at different time horizons using longitudinal data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Social | SVM | 68.97% | 0.2646 | 0.6171 | - |
| | | GBDT | 63.79% | 0.1892 | 0.5934 | - |
| | | LSM | 64.37% | 0.1621 | 0.5741 | - |
| | | LSTM | 66.67% | 0.1810 | 0.5719 | - |
| | | Transformer | 67.24% | 0.1649 | 0.5487 | - |
| 6 Months Ahead | Social | SVM | 67.24% | 0.1677 | 0.5675 | 55.00% |
| | | GBDT | 69.83% | 0.2301 | 0.5867 | 55.00% |
| | | LSM | 67.24% | 0.2366 | 0.6147 | 59.46% |
| | | LSTM | 64.66% | 0.1596 | 0.5752 | 55.00% |
| | | Transformer | 71.55% | 0.2756 | 0.5860 | 50.00% |
| 12 Months Ahead | Social | SVM | 70.69% | 0.2847 | 0.6347 | 75.00% |
| | | GBDT | 72.41% | 0.3556 | 0.6778 | 66.67% |
| | | LSM | 72.41% | 0.3364 | 0.6625 | 60.00% |
| | | LSTM | 72.41% | 0.2721 | 0.6014 | 50.00% |
| | | Transformer | 72.41% | 0.2721 | 0.6014 | 58.33% |

**Social**

In the longitudinal study of the social modality for predicting UHR outcomes, we can observe an enhancement in the performance metrics compared to the cross-sectional analysis. This improvement underscores the significance of utilizing longitudinal data to capture the evolution of social factors over time in predicting psychosis risk.

Comparing the longitudinal and cross-sectional results, there is a noticeable increase in performance across all metrics in the longitudinal analysis. For example, the MCC at 0 months ahead increased from 19.6% (GBDT in cross-sectional) to 26.46% (SVM in longitudinal), The MCC increased by an average of 11% across all modalities. This suggests that the temporal aspect of social data provides a more robust framework for predicting UHR outcomes.

The Acc-changed metric, which tracks the models' ability to predict transitions in UHR status, shows variability across models and timepoints. Notably, the Acc-changed at 12 months ahead for the SVM model in the longitudinal study (75%) is lower than in the cross-sectional study (80%), suggesting a nuanced interaction between temporal data and prediction of status changes that warrants further investigation. Among different models, the Transformer model shows a consistent performance at longer prediction horizons (6 and 12 months ahead). This might indicate the model's ability to effectively capture and integrate temporal social dynamics for risk prediction.

The longitudinal analysis of the social modality offers valuable insights into how social interactions and behaviors evolve over time in individuals at UHR. The enhanced predictive performance in the longitudinal setting suggests that dynamic social patterns are crucial indicators of progression towards psychosis, highlighting the importance of continuous monitoring and intervention in the social domain for at-risk individuals.

TABLE 5.9: Performance of genetic modality for predicting UHR outcomes at different time horizons using longitudinal data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Genetic | SVM | 84.80% | 0.4916 | 0.7036 | - |
| | | GBDT | 84.80% | 0.4795 | 0.6760 | - |
| | | LSM | 78.80% | 0.3435 | 0.6655 | - |
| | | LSTM | 82.40% | 0.3659 | 0.5918 | - |
| | | Transformer | 83.20% | 0.4075 | 0.6107 | - |
| 6 Months Ahead | Genetic | SVM | 84.00% | 0.4023 | 0.6300 | 61.54% |
| | | GBDT | 83.20% | 0.3686 | 0.6250 | 61.54% |
| | | LSM | 81.60% | 0.3646 | 0.6600 | 76.92% |
| | | LSTM | 83.20% | 0.3686 | 0.6250 | 84.62% |
| | | Transformer | 84.00% | 0.4481 | 0.6650 | 84.62% |
| 12 Months Ahead | Genetic | SVM | 87.20% | 0.5507 | 0.6950 | 66.67% |
| | | GBDT | 87.20% | 0.5507 | 0.6950 | 66.67% |
| | | LSM | 84.80% | 0.4669 | 0.6950 | 80.00% |
| | | LSTM | 86.40% | 0.5160 | 0.6750 | 86.67% |
| | | Transformer | 84.00% | 0.4001 | 0.6150 | 73.33% |

**Genetic**

The longitudinal analysis of the genetic modality for predicting Ultra-High Risk (UHR) outcomes demonstrates a marked improvement over the cross-sectional approach, the highest out of all the modalities.

The longitudinal study's notable improvement is largely due to selection of better features. The data at each time point from a participant in the training set, is treated as separate samples just for the feature selection process, to leverage the temporal nature of the data. This process ensures that the selection of most predictive genetic features over time, enhancing the models' ability to identify the progression towards psychosis.

There is a significant increase in accuracy across all models in the longitudinal analysis compared to the cross-sectional approach. The MCC increases by an average of 22%, highlighting the effectiveness of incorporating temporal dynamics in genetic data analysis.

The Acc-changed metric, particularly important in the longitudinal context, shows that models like the LSTM and Transformer are particularly adept at capturing the evolution of genetic risk factors, with the LSTM model showing an Acc-changed of 84.62% at 6 months ahead and 86.67% at 12 months ahead.

These results suggest that genetic factors, when analyzed over time, offer valuable insights into the trajectory of psychosis risk. The ability to predict changes in UHR status based on genetic data can inform proactive interventions and contribute to the development of personalized treatment plans.

**Discussion**

The longitudinal analysis of multiple data modalities—clinical, genetic, cognitive, and social factors—provides insights into the predictive power of these factors for Ultra-High Risk (UHR) of developing severe mental health conditions. The utilization of longitudinal data significantly enhances the understanding and prediction

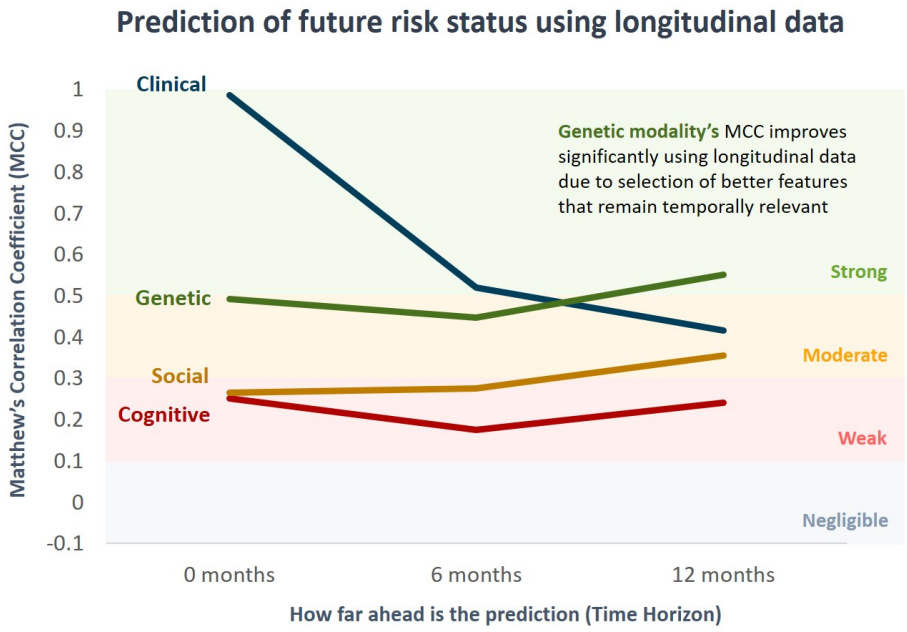


FIGURE 5.2: Top model performance of each modality at various time horizons for prediction of UHR status using longitudinal data

capabilities compared to cross-sectional analyses, illustrating the dynamic nature of risk factors associated with psychosis.

The clinical modality's performance in the longitudinal analysis, especially with near-perfect accuracy at the baseline (0 Months Ahead), reinforces the critical role of clinical assessments in the early identification of UHR individuals. The predictive performance improvement at 6 and 12 months ahead suggests that longitudinal clinical data provide a deeper understanding of how clinical risk factors evolve, offering a more accurate forecast of an individual's trajectory towards psychosis. This finding supports the notion that clinical factors are not static and that their changes over time are crucial in predicting the progression of mental health conditions.

For the cognitive modality, the longitudinal approach demonstrates the importance of temporal changes in cognitive functions as indicators of UHR status. The improvement in predictive performance over time indicates that it's not just the cognitive state at a single time point, but its trajectory that is vital for risk assessment. This aligns with and extends existing literature, suggesting that cognitive decline or variability could be a significant marker in the development of psychosis. The findings highlight the potential of cognitive monitoring over time as a critical component of early intervention strategies.

The enhanced predictive accuracy in the social modality underscores the value of examining social behaviors and interactions over time. The increase in performance metrics suggests that the progression of social factors is a significant indicator of psychosis risk, echoing the sentiment that social changes are as crucial as static assessments. This enhancement in prediction through longitudinal data illustrates the dynamic nature of social risk factors and their impact on mental health, offering new insights into how social dynamics play a role in the development of psychosis.

The genetic modality showed the most significant improvement in the longitudinal analysis, highlighting the importance of considering the temporal aspect of

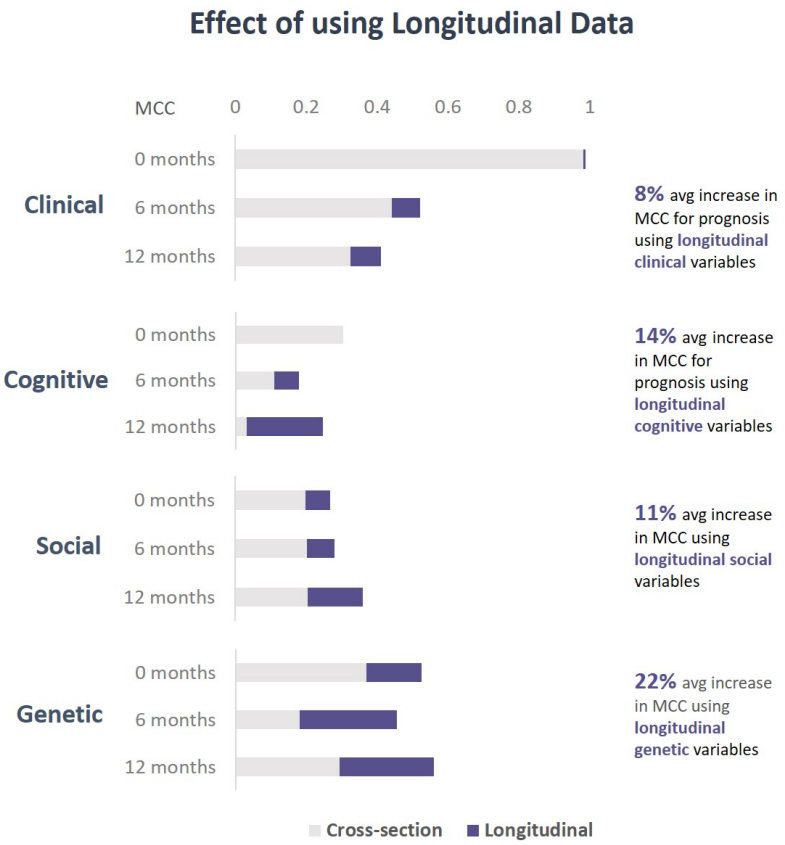


FIGURE 5.3: Improvement in performance of each modality at various time horizons for prediction of UHR status when using longitudinal data over cross-sectional data

genetic data. The feature selection process, tailored to capture temporal changes, significantly enhanced the predictive capabilities. This suggests that the genetic predisposition towards psychosis is not only a matter of static genetic markers but also how these markers interact with or are expressed over time. The ability to predict changes in UHR status based on genetic data opens avenues for personalized medicine, where interventions can be tailored based on an individual's genetic risk trajectory.

The improvement in prognostic performance when using longitudinal data for predicting UHR outcomes can be scientifically attributed to the dynamic nature of psychosis risk factors over time. Longitudinal data capture the temporal progression and fluctuations in these risk factors, providing a more comprehensive picture of an individual's mental health trajectory. This allows for the identification of patterns, trends, and changes in risk status that cross-sectional data, which offers only a snapshot in time, cannot. Consequently, models trained on longitudinal data can better account for the evolving context of risk indicators, enhancing their predictive accuracy for future outcomes. This approach aligns with the understanding that mental health conditions, especially psychosis, develop through complex interactions over time, necessitating a temporal perspective for effective prediction and intervention.

## 5.3 Mosaic LSM: Pioneering Multimodal Integration with LSMs

### 5.3.1 Rationale

The development of Mosaic LSM represents a significant step forward in enhancing the capabilities of Liquid State Machines (LSMs) for predictive modeling, particularly in the realm of multimodal data integration. This advancement is aimed at improving both the predictive performance and the explainability of LSM-based models.

#### Elevating Performance with Self-Attention

Mosaic LSM introduces a novel component, the Dynamic Attention Gateway (DAG), which refines the predictive accuracy and explainability of LSM. DAG incorporates a self-attention mechanism, allowing the model to focus on the most relevant features in the data dynamically. It consists of a sequence of layers: a self-attention layer, adaptive average pooling, layer normalization, and a fully-connected layer with dropout. This architecture enables the model to adaptively weigh the importance of different features, enhancing both performance and interpretability. The specifics of the self-attention mechanism and the functionality of each DAG layer, which contribute to the model's explainability, are elaborated in Appendix B and will be discussed further in a subsequent chapter dedicated to AI explainability in mental health.

#### Enhanced Representation of Data

Traditional LSMs transform input series into spike trains, focusing on learning temporal patterns while often losing absolute value information. However, in certain contexts, like gene expression analysis, absolute values hold critical importance. Mosaic LSM addresses this by introducing direct connections from the input data to the readout classifier, enabling the model to leverage both the temporal dynamics captured by the reservoir and the absolute values. This integration ensures a more comprehensive understanding of the data, as the output of the DAG is concatenated with the flattened input data, allowing the model to combine temporal inferences with absolute values effectively.

#### Strategic Fusion for Multimodal Data

To harness the full potential of multimodal data, Mosaic LSM incorporates various fusion strategies, each designed to capitalize on the strengths of different data integration approaches:

- **Early Fusion**: This strategy involves the concatenation of all modalities into a single dataset before transformation into spike-encoded data for LSM processing. It allows the model to capture interactions between modalities from the outset.
- **Intermediate Fusion**: With intermediate fusion, each modality is processed through its dedicated reservoir to capture unique temporal dependencies. The outputs from these reservoirs are then concatenated and fed into a common readout layer, enabling the model to integrate modality-specific insights.

- **Late Fusion**: This approach uses a separate reservoir for each modality, similar to intermediate fusion, but combines the outputs at a later stage. Each reservoir's output is processed by a distinct readout layer, and the final predictions are merged using weighted voting, allowing for a nuanced integration of modality-specific predictions.

By addressing these key areas, Mosaic LSM sets a new standard for multimodal data processing with LSMs, offering enhanced performance and deeper insights into the underlying patterns and relationships in complex datasets.

### 5.3.2 Architecture Design

The Mosaic LSM is designed to handle multimodal data, designated as $\mathcal{D} = \{(X_1^1 \dots X_1^m, y_1), \dots, (X_n^1 \dots X_n^m, y_n)\}$, where $X_n^m$ represents the $n^{th}$ sample for the $m^{th}$ data modality, and $y_n$ denotes the corresponding class label. Each $X_n^m \in \mathbb{R}^{t \times d}$ signifies a dataset containing $n$ samples, $d$ features, and $t$ time points $X_n^m = \{x_1 \dots x_t\}$. The primary aim is to learn the mapping $f : \mathcal{X} \to \mathcal{Y}$ utilizing the proposed Mosaic LSM, which comprises several components outlined below.

#### Spike Encoder

To utilize LSMs for modeling, the continuous data needs to be transformed into discrete spikes, as LSMs operate on event-based processing, similar to the action potentials of neurons in the brain. The step forward algorithm [30] is employed, which produces a positive spike when the input feature value at time $t$ exceeds the baseline $B$ (input value at $t = 0$) plus a threshold $th$. In this case, $B$ is updated as $B + th$. If the input feature value at time $t$ is less than $B + th$, a negative spike is produced, and $B$ is updated as $B - th$. This process is performed on all features in each modality of every sample.

#### Input Layer

The input layer establishes sparse connections to the reservoir neurons using weights from a uniform distribution. Formally, the connections are defined by the matrix $W_i$ of size $d \times N$, drawn from $\mathcal{U}(a, b)$. Each modality $X_n^1 ... X_n^m$ has a separate input layer defined by this procedure.

#### Liquid Layer (Reservoir)

The liquid layer is a reservoir composed of sparsely interconnected spiking neurons. The connections within the reservoir are established based on topological constraints informed by the small-world assumption [31], meaning neurons closer together have a higher probability of being connected than those farther apart. The connection weight between neurons $a$ and $b$ is defined by:

$$P_{a,b} = \begin{cases} C \cdot e^{-(d_{a,b}/\lambda)^2} & \text{if } d_{a,b} \leq d_{thresh} \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Where $d_{a,b}$ is the Euclidean distance between neurons $a$ and $b$, $d_{thresh}$ is the distance threshold, $C$ defines the maximum connection probability, and $\lambda$ defines the small-world connectivity radius.

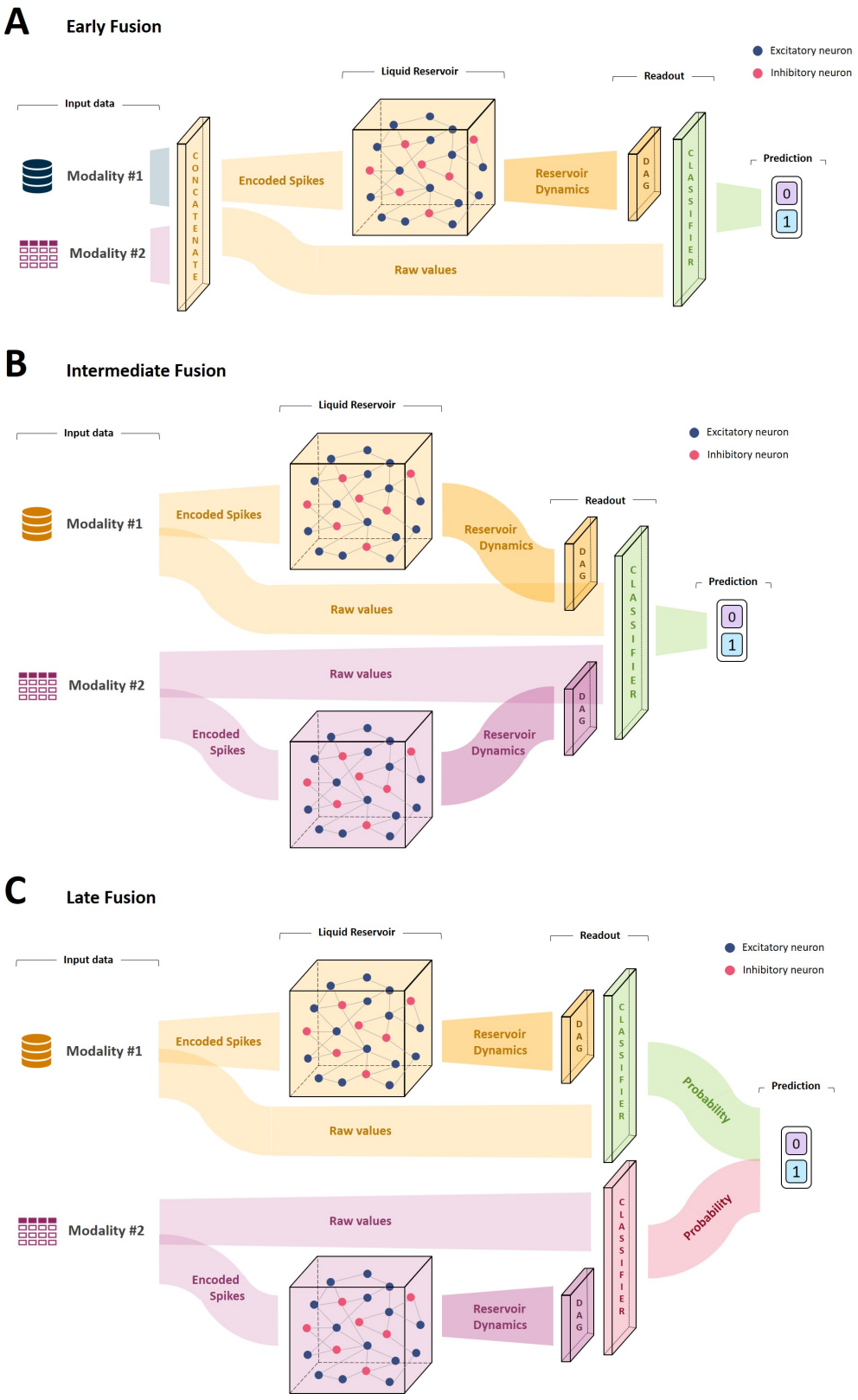


FIGURE 5.4: Various fusion architectures of the Mosaic LSM

The Leaky Integrate-and-Fire (LIF) neuron model is used for the reservoir dynamics, described by the following equations:

$$\mu[t] = \mu[t-1]e^{-\frac{1}{\tau}}(1-\theta_j[t]) + I[t] \tag{5.2}$$

$$I[t] = \sum_{i=1}^{d} W_i\theta_i[t] + \sum_{j=1}^{N} W_j\theta_j[t] \tag{5.3}$$

$$\theta_j[t] = \begin{cases} 1 & \text{if } \mu_j[t-1] \geq \mu_{th} \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

Here, $\mu[t]$ is the membrane potential of a neuron at time $t$, $\tau$ is the membrane's time constant, $\theta_j[t]$ represents the occurrence of a spike in the $j^{th}$ neuron at time $t$, $I[t]$ is the current injected into the neuron, and $\mu_{th}$ is the firing threshold.

**Concatenation Layer (for intermediate fusion)**

For intermediate fusion, to learn both joint and marginal representations of the modalities, the features and representations from each modality are stacked into a single high-dimensional vector, fed as input to the subsequent readout layer.

**Readout Layer**

The readout layer interfaces the highly nonlinear reservoir dynamics with the final outputs. For intermediate fusion, the number of times each neuron spiked for each modality and the original data is passed to a classifier for combined readout and classification. For late fusion, separate readouts are used for each modality to obtain separate classification probabilities.

**Combined Probability (for late fusion)**

In late fusion, the probabilities obtained from separate readouts for each modality are combined by taking a weighted average, resulting in marginal representations learned independently for each modality without considering inter-modality relationships.

The choice of fusion strategy (early, intermediate, or late) for a classification or regression task depends on the nature of the multimodal data and the desired level of integration between modalities. Figure 1 illustrates the Mosaic LSM architecture for the three fusion strategies.

### 5.3.3 Comparative Analysis of Mosaic LSM Architectures for Multimodal UHR Prediction

This study aims to explore the effectiveness of Mosaic Liquid State Machine (LSM) architectures for multimodal fusion in predicting Ultra-High Risk (UHR) for psychosis. By integrating clinical, genetic, cognitive, and social data modalities, we seek to develop a comprehensive predictive model that captures the complex, interconnected nature of risk factors associated with UHR.

**Study Design**

This study focuses on evaluating the efficacy of combining clinical, genetic, cognitive, and social data modalities through various machine learning models, including innovative Mosaic LSM architectures, for predicting Ultra-High Risk (UHR) of psychosis. The design incorporates a nuanced approach to data modality integration, especially for the diagnosis task at the horizon of 0 months ahead.

1. **Objective**: The objective of this study is threefold: firstly, to evaluate the combined effectiveness of genetic, cognitive, and social modalities in predicting Ultra-High Risk (UHR) for psychosis at the immediate horizon (0 months ahead), providing an objective alternative to subjective clinical assessments. Secondly, to assess the predictive capabilities of all four modalities—clinical, genetic, cognitive, and social—integrated for predicting UHR at extended horizons of 6 and 12 months ahead, thereby exploring the full potential of multimodal data fusion in UHR prognosis. Lastly, to compare the three different architectures of MosaicLSM (EF, IF and LF) in their ability to fuse and process the multiple modalities effectively, for UHR prediction.

2. **Models**: The five models from the longitudinal analysis-SVM, GBDT, LSM, LSTM, and Transformer-are employed for the multimodal . The data is fused and passed to the models. In addition, the three architectures of MosaicLSM are also compared. MosaicLSM-EF (Early Fusion) fuses data at the input level before processing through LSM. MosaicLSM-IF (Intermediate Fusion) integrates modality-specific LSM outputs using a common classifier. MosaicLSM-LF (Late Fusion) employs modality-specific LSM and readouts and combines output probabilities at the decision level.

3. **Validation**: The same 5-fold cross-validation strategy from the longitudinal study is used to assess the models' generalizability and minimize overfitting. This method ensures each data fold is utilized as a test set once while the remaining serve as the training set.

4. **Performance Metrics**: The study employs the same performance metrics as the cross-section and longitudinal analyses: Accuracy, Matthew's Correlation Coefficient (MCC), Area Under the Receiver Operating Characteristic Curve (AUC), and Accuracy for Changed Status (Acc-changed). These metrics help evaluate each model's effectiveness in UHR prediction using multimodal data.

5. **Feature Selection**: Feature selection for the genetic modality is consistent with the longitudinal study's methodology. Each timepoint is treated as a distinct sample for feature selection purposes, ensuring that the most predictive features across different stages are utilized, enhancing the models' predictive validity.

6. **Hyperparameter Optimization**: Hyperparameters for all classifiers were tuned using a grid search on the training set within each fold of the cross-validation procedure. The selection criterion was Matthews Correlation Coefficient (MCC), evaluated on a 20% subset of the training data, which was held out as a validation set during tuning. This process was repeated independently within each fold to ensure that hyperparameter selection did not involve any information from the corresponding test fold.

7. **Computational Environment**: All experiments were implemented in `Python 3.8`, using libraries including `scikit-learn`, `XGBoost`, `pandas`, `NumPy`, `PyTorch`, and `NeuCubePy` for model development and evaluation. Visualisations were produced using `Matplotlib` and `Seaborn`. Experiments were executed on an institutional machine equipped with an `Intel Core i5 CPU` and `16 GB RAM`. GPU acceleration was not required for any of the experiments.

8. **Data Integrity**: Data integrity measures mirror those in the longitudinal study, with a strict protocol ensuring that predictions at any given timepoint are based solely on data available up to that point. This approach maintains the chronological order of events and the authenticity of the data.

TABLE 5.10: Comparison of methods for predicting UHR outcomes at different time horizons using multimodal longitudinal data

| Prediction Horizon | Modality | Model | Accuracy | MCC | AUC | Acc-changed |
|---|---|---|---|---|---|---|
| 0 Months Ahead | Multimodal | SVM | 76.92% | 0.5314 | 0.7628 | - |
| | (Cognitive | GBDT | 73.08% | 0.4528 | 0.7239 | - |
| | + | LSM | 69.23% | 0.4074 | 0.7023 | - |
| | Social | LSTM | 83.97% | 0.6766 | 0.8387 | - |
| | + | Transformer | 82.05% | 0.6400 | 0.8213 | - |
| | Genetic) | MosaicLSM-EF | 85.90% | 0.7194 | 0.8495 | - |
| | | MosaicLSM-IF | 85.90% | 0.7160 | 0.8588 | - |
| | | MosaicLSM-LF | 87.18% | 0.7432 | 0.8731 | - |
| 6 Months Ahead | Multimodal | SVM | 80.77% | 0.6064 | 0.7879 | 70.00% |
| | (Clinical | GBDT | 77.88% | 0.5457 | 0.7720 | 75.00% |
| | + | LSM | 75.96% | 0.5040 | 0.7492 | 65.00% |
| | Cognitive | LSTM | 75.00% | 0.5822 | 0.7803 | 45.00% |
| | + | Transformer | 80.77% | 0.6451 | 0.8242 | 70.00% |
| | Social | MosaicLSM-EF | 85.58% | 0.7052 | 0.8417 | 90.00% |
| | + | MosaicLSM-IF | 87.50% | 0.7429 | 0.8674 | 65.00% |
| | Genetic) | MosaicLSM-LF | 89.42% | 0.7827 | 0.8871 | 75.00% |
| 12 Months Ahead | Multimodal | SVM | 82.69% | 0.6438 | 0.8197 | 85.71% |
| | (Clinical | GBDT | 73.08% | 0.4485 | 0.7242 | 64.29% |
| | + | LSM | 71.15% | 0.4058 | 0.7015 | 78.57% |
| | Cognitive | LSTM | 84.62% | 0.6860 | 0.8303 | 92.86% |
| | + | Transformer | 80.77% | 0.6210 | 0.7788 | 92.86% |
| | Social | MosaicLSM-EF | 86.54% | 0.7265 | 0.8652 | 92.86% |
| | + | MosaicLSM-IF | 86.54% | 0.7231 | 0.8591 | 64.29% |
| | Genetic) | MosaicLSM-LF | 88.46% | 0.7636 | 0.8818 | 78.57% |

**Diagnosis (0 months)**

During the initial diagnosis phase at the 0-month horizon, the primary focus was on the cognitive, social, and genetic modalities, excluding clinical data due to its role in determining UHR labels. The MosaicLSM-LF architecture demonstrated superior performance with the highest MCC value of 0.7432, indicating its effectiveness in integrating these modalities for immediate UHR prediction. The LSTM model also showed strong performance with an MCC of 0.6766, suggesting its capability in capturing temporal relationships within the data even without clinical input.

**Prognosis (6 months)**

For the 6-month prognosis, incorporating all four modalities, the MosaicLSM-LF architecture again stood out, showing the highest MCC of 0.7827. This indicates a robust capacity to fuse and interpret the multimodal data for medium-term UHR predictions. Interestingly, the MosaicLSM-EF architecture, while having a slightly lower MCC, demonstrated the highest Acc-changed at 90.00%, suggesting its particular strength in identifying dynamic shifts in UHR status over time.

**Prognosis (12 months)**

At the 12-month horizon, the LSTM model exhibited a notably high MCC of 0.6860, surpassing other traditional models and indicating its effectiveness for long-term prognosis. Among the Mosaic LSM architectures, MosaicLSM-LF again showed the best performance with an MCC of 0.7636. The high Acc-changed values for LSTM and MosaicLSM-EF highlight their capability in predicting long-term transitions in UHR status, emphasizing the importance of longitudinal data analysis in mental health assessments.

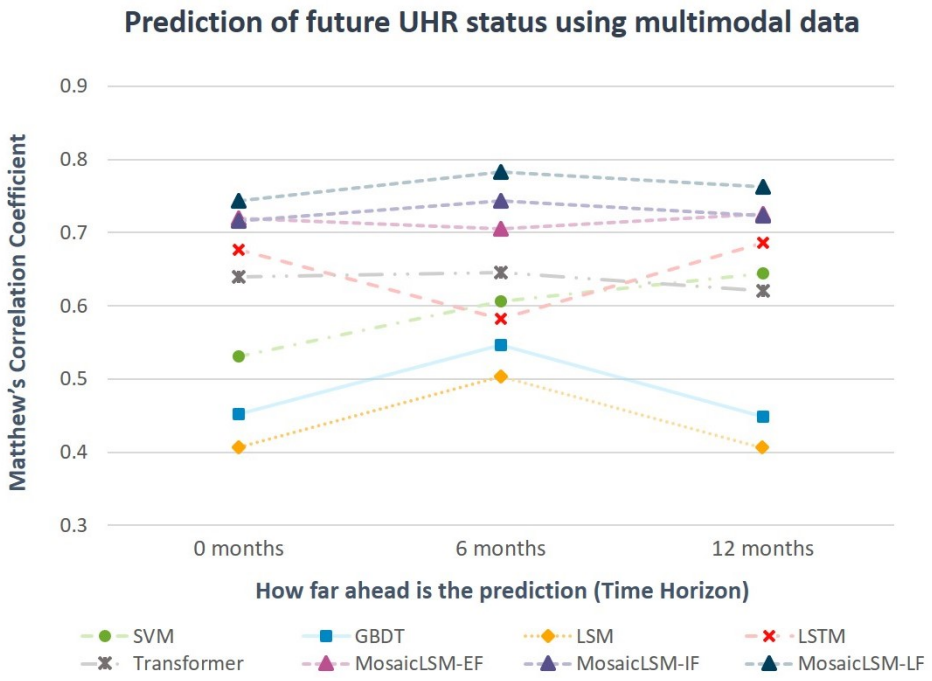


FIGURE 5.5: Comparison of performance of different methods for prediction of UHR status at various time horizons using multimodal longitudinal data

**Discussion**

The analysis of the MosaicLSM architectures for predicting UHR of psychosis provides important insights into the integration of multimodal data in mental health predictive models.

At the initial diagnosis stage, MosaicLSM-LF emerged as particularly effective, with its MCC surpassing 0.75. This indicates that late fusion, which combines insights post individual modal analysis, can synthesize a comprehensive view from

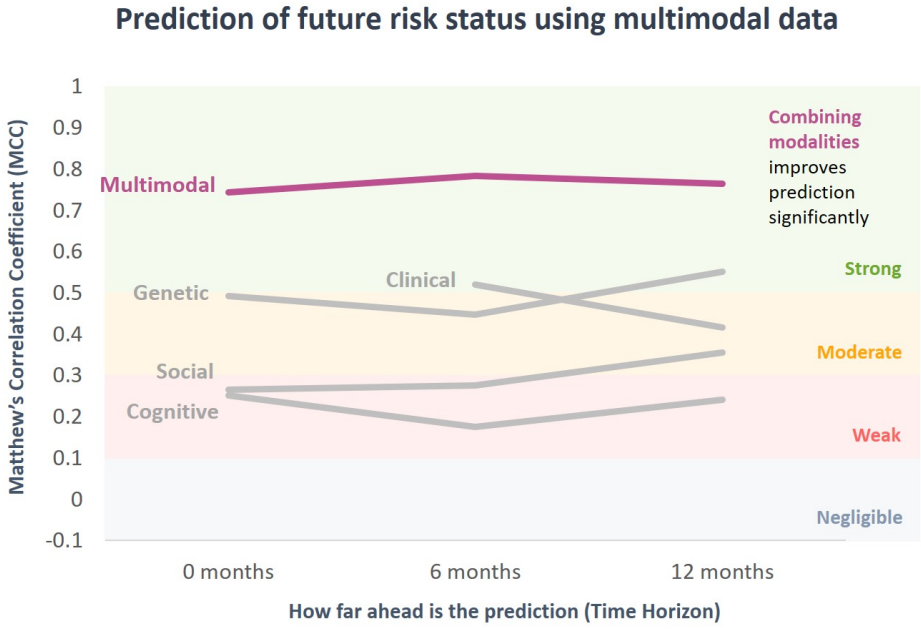


FIGURE 5.6: Top model performance at various time horizons for prediction of UHR status using multimodal longitudinal data

cognitive, social, and genetic data. The strength of MosaicLSM-LF suggests that understanding the nuanced interplay of these modalities can enhance immediate risk assessments, offering a data-driven alternative to traditional clinical evaluations.

Over longer horizons, MosaicLSM-LF consistently showed superior performance, highlighting its capability to integrate and interpret evolving multimodal data. The advantage of late fusion likely stems from its capacity to preserve and leverage the distinct, yet interconnected, contributions of each modality, facilitating a more accurate long-term risk prediction.

The differential performance among early, intermediate, and late fusion strategies within MosaicLSM architectures sheds light on the complexity of multimodal data integration. Late fusion's effectiveness for UHR prediction suggests it can accommodate the distinct temporal evolutions and interactions of risk factors better than early or intermediate fusion, which may blend modalities too soon, potentially obscuring crucial intermodal relationships.

These findings endorse the MosaicLSM, particularly its late fusion variant, as a robust tool for mental health diagnostics and prognosis. By offering a sophisticated method to integrate diverse data types, MosaicLSM-LF could significantly enhance the accuracy of UHR predictions. This advancement not only has implications for the early detection and intervention in psychosis but also exemplifies the potential for similar multimodal approaches in broader mental health and medical fields, paving the way for more personalized and effective healthcare solutions.

## 5.4 Conclusion

This chapter has presented the Mosaic Liquid State Machine (Mosaic LSM) method as a novel approach for integrating multimodal longitudinal data to improve the

diagnosis and prognosis of mental health disorders. By addressing the research objectives of integrating multimodal data sources (Objective 1) and enhancing predictive accuracy using longitudinal data (Objective 2), the Mosaic LSM method offers a comprehensive and effective solution to the complexities inherent in mental health research.

The integration of diverse data sources, such as clinical assessments, cognitive assessments, genetic data, and social functioning assessments, is crucial for capturing the multifaceted nature of mental health disorders. Each data type contributes unique insights that, when combined, provide a holistic view of an individual's mental health. The Mosaic LSM method effectively manages the high-dimensional and heterogeneous nature of these multimodal datasets, ensuring that the integrated model can leverage the strengths of each data type.

Longitudinal data, with its repeated observations over time, adds significant value by capturing the temporal dynamics and progression of mental health conditions. The Mosaic LSM method's ability to handle temporal data ensures that these dynamics are accurately represented, facilitating the identification of early markers and patterns that are critical for effective diagnosis and prognosis.

The experimental results demonstrated the superiority of the Mosaic LSM method in terms of predictive accuracy and robustness compared to traditional approaches. The method's ability to integrate and analyze multimodal longitudinal data resulted in more accurate and reliable predictions, which are essential for clinical decision-making and patient care. These findings address Research Question 1, showing how multimodal data sources can be effectively integrated to enhance mental health diagnostics and prognostics.

Despite its strengths, the Mosaic LSM method is not without limitations. The computational complexity associated with processing high-dimensional and longitudinal data can be substantial. Future research should focus on optimizing the computational efficiency of the Mosaic LSM method and exploring its scalability for larger datasets and real-world applications. Additionally, further validation using diverse clinical populations and settings will be necessary to ensure the generalizability and robustness of the model.

In conclusion, this chapter has made contributions to the field of mental health diagnostics by developing and validating the Mosaic LSM method for multimodal longitudinal data integration. The insights gained from this research underscore the importance of integrating diverse data sources to capture the full complexity of mental health disorders. The Mosaic LSM method provides a powerful tool for improving predictive accuracy and enhancing the understanding of mental health conditions. The integration of multimodal and longitudinal data represents an advancement in mental health diagnostics, paving the way for more precise, reliable, and comprehensive AI-driven models. The research presented in this chapter has been published at IJCNN (Budhraja et al., 2023b).

# Chapter 6

# NeuroGeMS: Translating Research to Real-World Applications

## 6.1 Introduction

The advancements in artificial intelligence (AI) and machine learning (ML) have revolutionized the field of mental health diagnostics and prognosis. However, the practical application of these sophisticated models in clinical settings remains a significant challenge. Interdisciplinary researchers often face difficulties in utilizing advanced AI models due to their complexity and the lack of user-friendly tools that facilitate seamless integration into existing workflows. This chapter addresses the research objective of facilitating the adoption of multimodal AI methods and advancing research in the biomedical domain (Objective 5 and Question 5).

To bridge the gap between AI innovation and clinical practice, it is necessary to develop user-friendly software that can handle the complexity of multimodal data while providing intuitive interfaces for non-technical users. This chapter introduces NeuroGeMS (Neuro Genetic Multimodal System), a graphical user interface (GUI) software designed to simplify the process of multimodal modelling and analysis. NeuroGeMS aims to make advanced AI tools accessible to researchers, thereby enhancing the practical applicability of AI models in mental health care.

NeuroGeMS is built to support the integration and analysis of various data types, including assessments and genetic data. By providing a unified platform for managing diverse datasets, NeuroGeMS enables users to perform comprehensive analyses without the need for extensive technical expertise. The software incorporates personalized AI methods, such as WWKNN and TWNFI discussed in previous chapters, ensuring that users can leverage the latest advancements in AI for mental health diagnostics.

The chapter is structured as follows. First, we review the existing tools and software used in mental health research, highlighting their strengths and limitations. Next, we introduce the NeuroGeMS software, detailing its design principles, architecture, and key features. The methodology section describes the implementation process, including the integration of multimodal data handling capabilities and the development of user-friendly interfaces. We then present case studies and examples demonstrating the practical applications of NeuroGeMS in real-world clinical and research settings. Finally, the chapter discusses the implications of NeuroGeMS for advancing mental health research, potential areas for improvement, and future research directions.

By developing and validating NeuroGeMS, this chapter makes contributions to the field of mental health diagnostics. The software not only facilitates the adoption

of advanced AI methods by the broader research community, and exploring new insights and innovations. NeuroGeMS represents a meaningful step towards bridging the gap between AI research and medical practice.

## 6.2 NeuroGeMS: A GUI for Accessible and Interpretable Multimodal Modelling

### 6.2.1 Vision and Motivation

The rapid advancement of multimodal AI has been driven by the need to effectively process and analyze multiple modalities of data related to the same outcome or individual. This is especially critical in the field of biomedical engineering, where diverse multimodal data, such as genetic, proteomic, neuroimaging, clinical, cognitive, and behavioral, are increasingly available (Lahat, Adali, and Jutten, 2015). The efficient integration and interpretation of this multimodal data is pivotal for advancing diagnosis, prognosis, and treatment strategies in the rapidly evolving field of medical data analysis (Miotto et al., 2018). However, the potential of such integration is often bottlenecked by the complexity of existing analytical tools, which necessitates a user-friendly solution to make multimodal data analysis accessible to a wider array of medical and research professionals.

#### Examples of Multimodal AI in Medicine

Several studies have explored the application of multimodal learning in various biomedical domains. Sui et al., 2012 provided a comprehensive review of multimodal neuroimaging methods, discussing the advantages and challenges of integrating multiple imaging modalities for brain disorder analysis. They highlighted the potential of multimodal approaches in capturing the complex nature of brain disorders and improving diagnostic accuracy.

Zhang et al., 2019 proposed a deep learning framework for multimodal biomedical image analysis, demonstrating its effectiveness in tasks such as brain tumor segmentation and Alzheimer's disease diagnosis. Their framework leveraged the complementary information from multiple imaging modalities to enhance the performance of the analysis tasks.

In the context of mental health, Cummins et al., 2015 explored the use of multimodal data, including speech, text, and visual cues, for the assessment and monitoring of depression. Their review highlighted the potential of multimodal approaches in capturing the complex nature of mental disorders and providing more comprehensive insights into the patient's condition.

#### The Need for Accessible Multimodal Learning Tools

However, the complexity of integrating and analyzing this data often necessitates sophisticated computational techniques, typically accessible only to specialists with extensive programming expertise (Baltrusaitis, Ahuja, and Morency, 2019). This gap limits the widespread adoption and innovation in multimodal medical analysis. Moreover, the interpretability and explainability of multimodal learning models are consequential in the medical domain, where decisions can have significant implications for patient care (Tjoa and Guan, 2020). Black-box models, while potentially accurate, may not provide the necessary transparency and trust required in clinical settings (Holzinger et al., 2017).

Despite the promising results of multimodal learning in biomedical research, the accessibility and usability of these techniques remain a challenge. O'Halloran et al., 2012 discussed the need for standardized tools and frameworks to facilitate the development and deployment of multimodal learning systems. They emphasized the importance of user-friendly interfaces and modular architectures to enable researchers and practitioners to leverage these techniques effectively.

However, there is still a lack of comprehensive GUI-based software that specifically targets multimodal learning in the biomedical domain. Existing tools often focus on specific data types or analysis tasks, limiting their applicability to a broader range of multimodal problems. Moreover, the integration of advanced techniques such as deep learning and the incorporation of interpretability and explainability measures remain limited in current software solutions.

### 6.2.2 Related Works

#### General Machine Learning Softwares

Several general-purpose machine learning software and frameworks have been developed to facilitate the development and deployment of ML models. Weka (Hall et al., 2009) is a popular open-source tool that provides a collection of machine learning algorithms for data preprocessing, classification, regression, clustering, and association rules mining. It offers a user-friendly GUI and supports various data formats, making it accessible to users with limited programming experience. However, Weka's focus is primarily on traditional machine learning techniques and does not provide specific functionalities for multimodal learning or biomedical data analysis.

Weights and Biases (W&B) (Biewald et al., 2020) is a cloud-based platform that enables experiment tracking, visualization, and collaboration for machine learning projects. It provides tools for monitoring model performance, comparing different runs, and sharing results with team members. While W&B supports a wide range of ML frameworks and can be used for tracking multimodal learning experiments, it still requires expertise in building models and writing code.

PyTorch (Paszke et al., 2019) is a popular open-source deep learning framework that provides a flexible and dynamic programming interface for building and training neural networks. It offers a wide range of tools and libraries for tasks such as computer vision, natural language processing, and reinforcement learning. Although PyTorch can be used for multimodal learning tasks, it requires significant programming expertise and does not provide a GUI-based environment for biomedical data analysis.

#### Biomedical Domain Tools

In the biomedical domain, several tools have been developed to facilitate the analysis of specific types of biomedical data. Nilearn (Abraham et al., 2014) is a Python package that provides a set of tools for statistical learning on neuroimaging data, supporting tasks such as data preprocessing, feature extraction, and model evaluation. However, it focuses on a specific data type and lacks the capabilities for multimodal learning.

Similarly, PyMVPA (Hanke et al., 2009) is a Python toolbox for multivariate pattern analysis of neuroimaging data, offering a wide range of algorithms for classification, regression, and feature selection. While it provides a comprehensive set of tools for analyzing neuroimaging data, its applicability to a broader range of multimodal problems is limited.

In the field of bioinformatics, tools like Bioconductor (Gentleman et al., 2004) and Galaxy (Afgan et al., 2018) have been developed to facilitate the analysis and integration of multi-omics data. These platforms provide user-friendly interfaces and workflows for tasks such as data preprocessing, normalization, and statistical analysis. However, they lack the integration of advanced techniques such as deep learning and the incorporation of interpretability and explainability measures.

#### NeuroGeMS: Bridging the Gap

NeuroGeMS (Neuro-genetic Multimodal System) aims to bridge the gap between advanced multimodal learning techniques and their accessibility to researchers in the biomedical domain. It provides a unified GUI framework for multimodal modeling, specifically designed to address the challenges and requirements of multimodal learning in this field. Unlike conventional machine learning GUIs that primarily focus on single data type analysis for classification and regression tasks, NeuroGeMS additionally supports multimodal learning approaches such as early and late fusion (Atrey et al., 2010).

With its friendly interface, extensive suite of machine learning techniques, and support for various multimodal fusion strategies and interpretability measures, NeuroGeMS is poised to accelerate the adoption and application of multimodal learning in biomedical research. Its specialized functionalities and tailored design set it apart from general-purpose machine learning software and existing tools in the medical domain. By empowering researchers with accessible and interpretable tools for multimodal data analysis, NeuroGeMS enables the full potential of multimodal learning to be harnessed for improving patient outcomes and advancing biomedical discoveries.

### 6.2.3 Architecture of NeuroGeMS

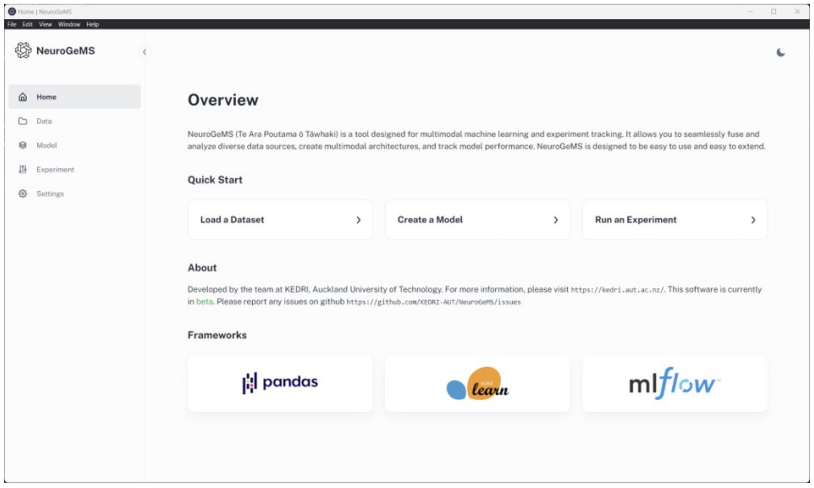


FIGURE 6.1: Home page of NeuroGeMS. The home page introduces the key features and capabilities of NeuroGeMS, and provides quick start links to the data, model, and experiment pages. The page also acknowledges the frameworks and libraries used in the development of NeuroGeMS.

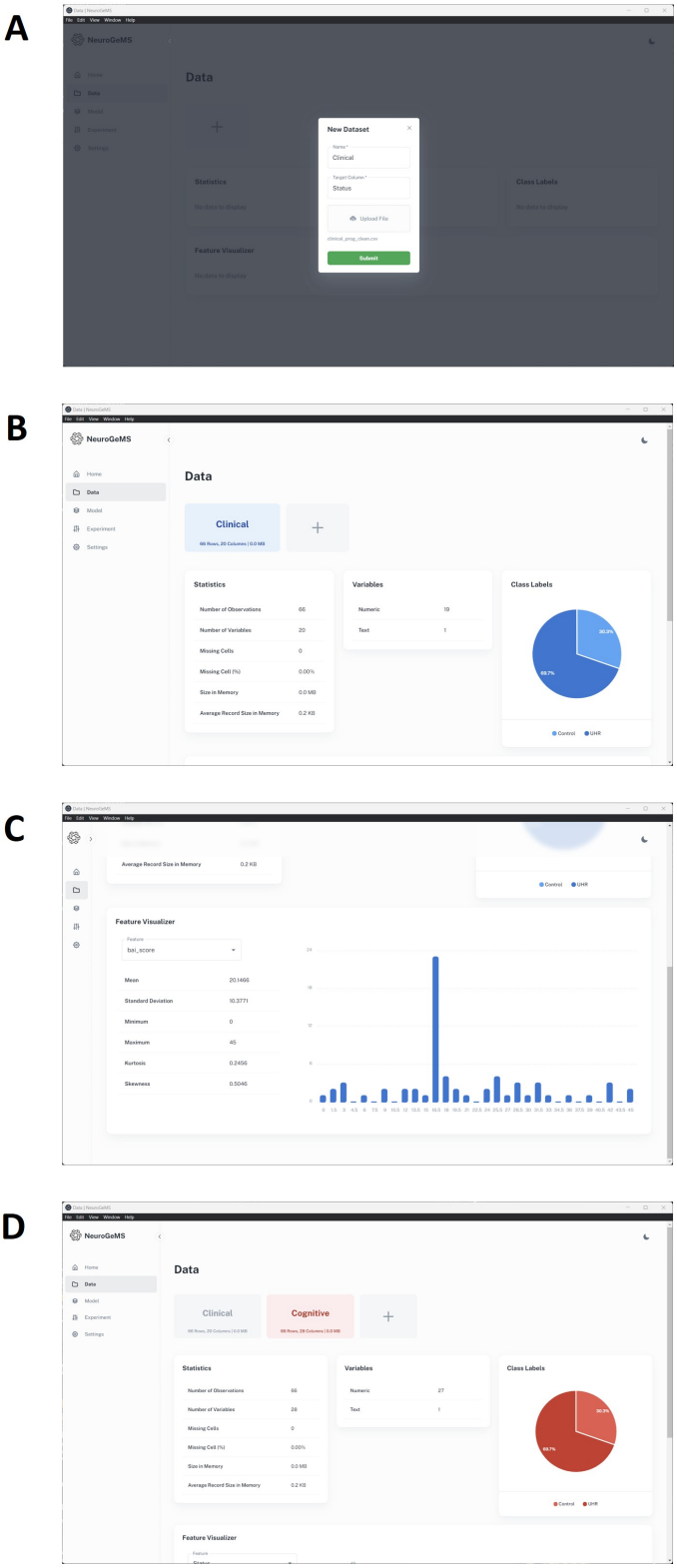


FIGURE 6.2: Data page of NeuroGeMS. The software provides a user-friendly interface for creating and managing datasets. (A) Users can add new data modalites by specifying a name, target column and loading the file. (B) Each data modality is displayed along with their summary statistics, and (C) visualisation of feature distributions. (D) This facilitates the organization and exploration of multimodal datasets.

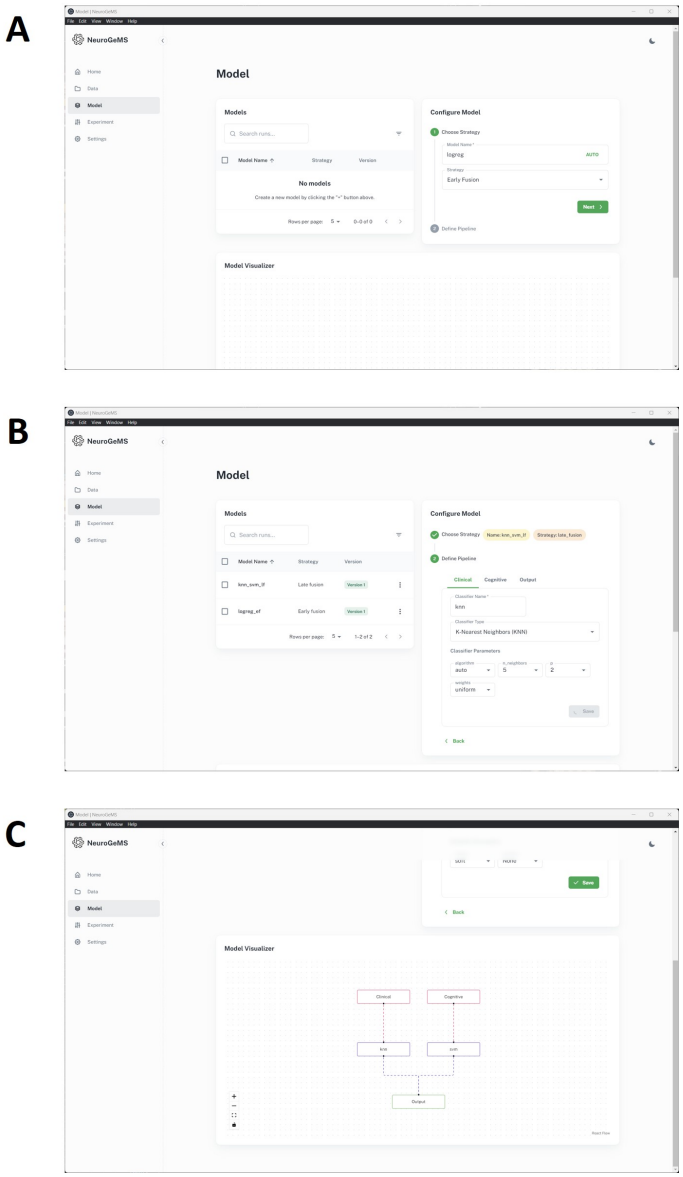


FIGURE 6.3: Model page of NeuroGeMS. NeuroGeMS provides a dedicated interface for creating and configuring multimodal models. (A) Users can create new models by specifying the model name and selecting the desired model type, such as early fusion or late fusion. (B) Existing models are displayed along with their details. (C) The model visualizer provides a graphical representation of the model architecture. This interface enables researchers to design and experiment with different multimodal architectures.

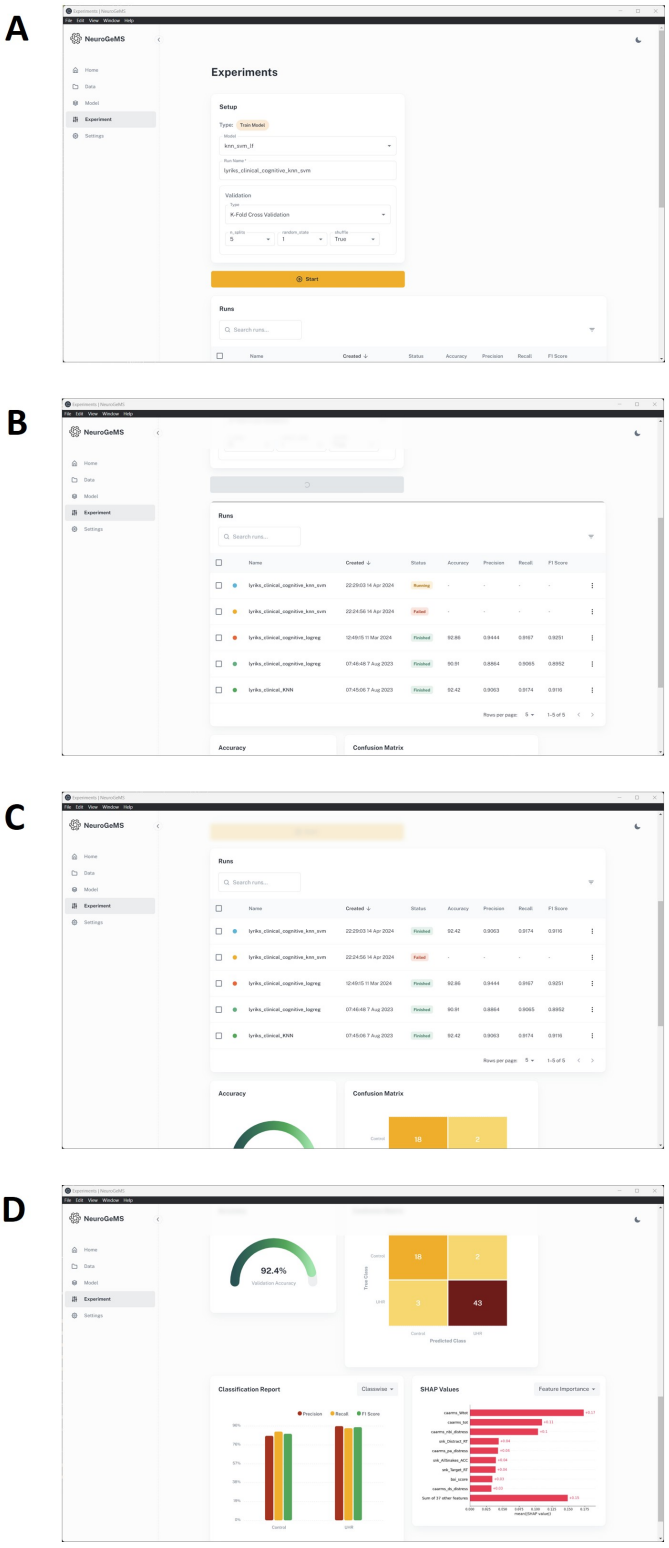


FIGURE 6.4: Experiment page of NeuroGeMS. NeuroGeMS provides a centralized interface for managing and tracking experiments. (A) New experiments can be initiated by specifying the necessary setup parameters. (B)-(D) Users can view a list of existing experiments along with their setup details, validation metrics, and plots. This interface enables researchers to keep track of their experimental results, compare different models and configurations, and ensure reproducibility.

NeuroGeMS (Neuro-genetic Multimodal System) is designed to simplify and enhance the process of multimodal data analysis in biomedical research. Its architecture integrates various advanced technologies to provide a robust, flexible, and user-friendly platform. NeuroGeMS consists of two primary layers: the User Interface (UI) layer, and the Application Logic layer. Each layer is designed to handle specific tasks, ensuring a clear separation of concerns and promoting modularity and scalability.

#### User Interface (UI) Layer

The UI layer provides an intuitive interface for users to interact with NeuroGeMS. Built using React and Material-UI, this layer ensures a responsive and visually appealing experience. Key components of the UI layer include:

- **Data Interface:** Allows users to import and explore multimodal datasets. It supports various data formats and provides visualization tools for initial data inspection.
- **Model Interface:** Facilitates the construction and configuration of machine learning models. Users can select from predefined models or define custom models tailored to their specific needs.
- **Experiment Interface:** Integrates with MLflow to provide comprehensive experiment tracking. Users can monitor, compare, and reproduce experiments, ensuring transparency and reproducibility in their research.

#### Application Logic Layer

The Application Logic layer, built with Python Flask, handles the core functionalities of NeuroGeMS. This layer is responsible for processing user requests, managing workflows, and executing machine learning algorithms. Key components of the Application Logic layer include:

- **Data Handling Module:** Manages the import, and storage of multimodal datasets. It leverages libraries like Pandas, NumPy, and YData Profiling to ensure efficient data manipulation.
- **Model Handling Module:** Provides tools for defining, training, and evaluating machine learning models. It supports many models integrating with libraries like scikit-learn and PyTorch.
- **Experiment Management Module:** Integrates with MLflow to track and manage experiments. This module logs experiment details, stores model artifacts, and provides tools for result analysis and comparison.

## 6.3 Software Capabilities

### 6.3.1 Overview

NeuroGeMS offers a range of capabilities to empower users to analyze, integrate, and interpret multimodal data, develop personalized models, ensure reproducibility, and extend the software's functionality to meet their specific needs.

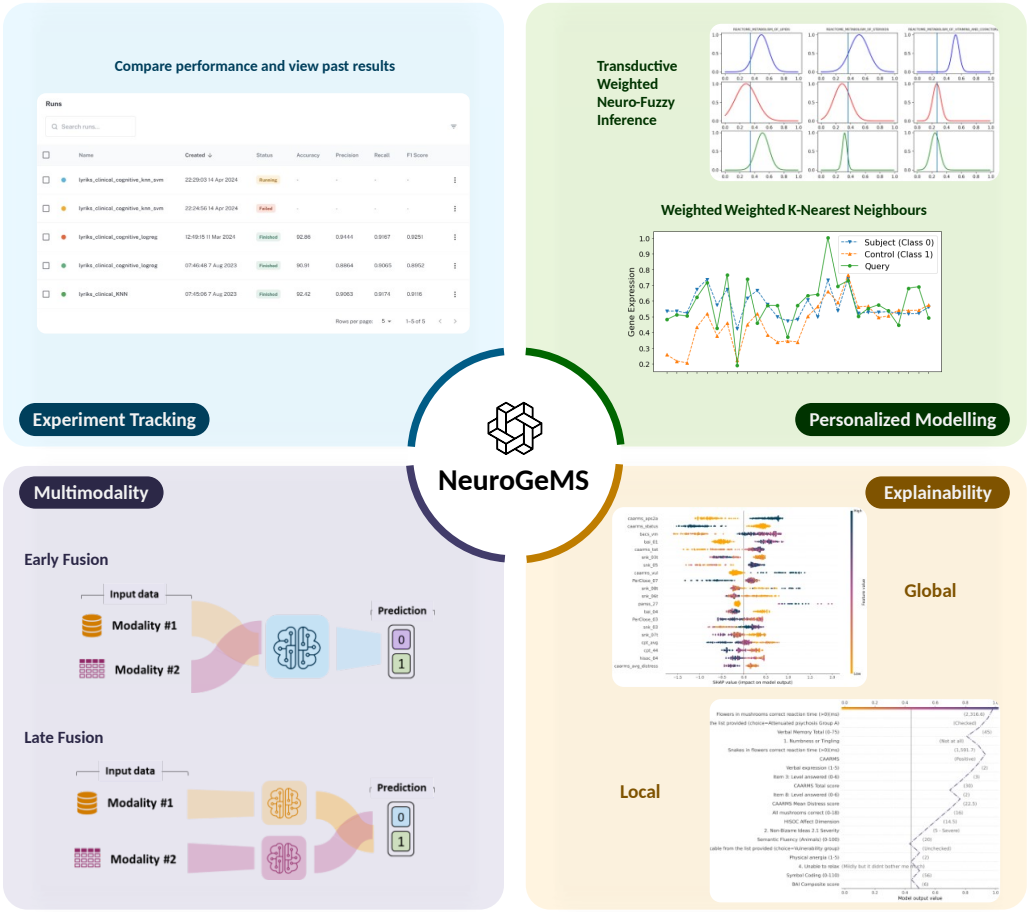


FIGURE 6.5: Overview of the NeuroGeMS software capabilities. The software is centered around the core functionalities of multimodality, experiment tracking, personalized modeling, and explainability. The multimodality component supports both early fusion and late fusion strategies for integrating data from multiple modalities. The experiment tracking feature, powered by MLflow, enables users to compare the performance of different models and view past results. Personalized modeling is achieved through advanced techniques such as Transductive Weighted Neuro-Fuzzy Inference (TWNFI) and Weighted Weighted K-Nearest Neighbors (WWKNN). The explainability component provides both global and local explanations of model predictions using techniques like SHAP (SHapley Additive exPlanations). These features collectively enable researchers to build robust and interpretable multimodal models for biomedical applications.

The software provides state-of-the-art multimodal learning capabilities, supporting various strategies for integrating and analyzing data from multiple modalities (Section 6.3.2). These strategies allow researchers to leverage the complementary information present in different data types to gain comprehensive insights into complex biomedical phenomena.

NeuroGeMS also focuses on explainability, incorporating frameworks like SHAP (SHapley Additive exPlanations) to help researchers understand and interpret the predictions made by their models (Section 6.3.3). This explainability is crucial for building trust, validating results, and making informed decisions in biomedical research and medical applications.

Personalized modeling is another key aspect of NeuroGeMS, offering advanced techniques that allow researchers to develop models that capture individual-level heterogeneity and adapt to the specific characteristics of each patient (Section 6.3.4). These methods enable the creation of personalized models that can provide more accurate and relevant predictions for individual patients.

NeuroGeMS fully supports reproducibility and experiment tracking, seamlessly integrating with platforms like MLflow (Section 6.3.5). This integration provides a centralized and standardized framework for experiment management, comparison, and collaboration, ensuring the reproducibility of research results.

Extensibility is a core feature of NeuroGeMS, with its modular and flexible architecture allowing researchers to easily integrate new preprocessing techniques, machine learning models, and multimodal strategies (Section 6.3.6). This extensibility ensures that NeuroGeMS can adapt to the evolving needs of biomedical research and incorporate the latest advancements in the field.

In the following subsections, we will explore each of these software capabilities in detail, discussing their technical foundations, practical applications, and potential impact on advancing biomedical knowledge and personalized medicine.

### 6.3.2 Multimodal Learning

NeuroGeMS is designed to support various multimodal learning strategies, enabling users to analyze and integrate data from multiple modalities effectively. The software provides a flexible framework that accommodates different fusion techniques, allowing researchers to choose the most suitable approach for their specific problem. In this section, we discuss the multimodal learning capabilities of NeuroGeMS, including unimodal analysis, early fusion, late fusion, and future directions such as co-learning and translation.

#### Unimodal Strategy

Unimodal analysis is a fundamental capability of NeuroGeMS, allowing users to process and analyze data from individual modalities independently. This strategy is useful when researchers want to explore the predictive power of each modality separately or when the relationship between different modalities is not well understood. Users can easily select a specific modality, such as genetic data or clinical features, and apply appropriate techniques to analyze the data.

The unimodal analysis capabilities of NeuroGeMS enable researchers to gain insights into the individual contributions of each modality to the prediction task. By analyzing modalities separately, users can identify the most informative features and assess the predictive performance of each modality. This information can guide

further analysis and help in selecting the most relevant modalities for multimodal integration.

#### Early Fusion Strategy

Early fusion is a multimodal learning strategy that combines different modalities at the feature level before feeding them into a single predictive model (discussed in Chapter 5 Section 5.2.4).

NeuroGeMS supports early fusion by providing tools for feature-level integration of multiple modalities. In the early fusion approach, features from different modalities are concatenated or transformed into a common representation space. This allows the model to learn the joint representation of the multimodal data and capture the interactions between modalities.

Users can easily configure the early fusion process in NeuroGeMS by selecting the desired modalities and specifying the fusion technique. The software handles the alignment and synchronization of the multimodal data, ensuring that the features from different modalities are properly integrated. Once the fused features are obtained, users can apply machine learning algorithms to train the predictive model.

#### Late Fusion Strategy

Late fusion is another multimodal learning strategy supported by NeuroGeMS (discussed in Chapter 5 Section 5.2.4). In contrast to early fusion, late fusion combines the predictions or decisions from multiple unimodal models to obtain the final prediction.

NeuroGeMS provides a framework for late fusion that enables users to train separate models for each modality and then combine their predictions using voting techniques. Users can define the method and hyperparameters for each unimodal model, allowing them to optimize the performance of individual modality-specific models. The software handles the synchronization and alignment of the predictions, ensuring that the fusion process is performed correctly.

Late fusion in NeuroGeMS enables researchers to leverage the strengths of different modalities. This approach allows for more flexibility in modeling each modality separately and can be beneficial when the optimal model architecture differs for each modality.

#### Future Directions: Co-Learning and Translation

NeuroGeMS is designed with extensibility in mind, allowing for the integration of future multimodal learning strategies. Two promising directions for future development are co-learning and translation.

Co-learning is a multimodal learning approach that enables the joint training of multiple modality-specific models, where each model learns from the others. In this approach, the models are trained simultaneously, and the knowledge gained from one modality is used to guide the learning process of the other modalities. Co-learning can help in capturing the complex interactions between modalities and improving the overall predictive performance.

Translation, on the other hand, focuses on mapping data from one modality to another. This approach can be useful when there is a lack of data in one modality or when the goal is to generate synthetic data in a target modality based on the information from a source modality.

NeuroGeMS aims to incorporate these advanced multimodal learning strategies in future updates. The modular architecture of the software allows for the seamless integration of new algorithms and techniques. Researchers and developers can contribute to the development of co-learning and translation modules, expanding the capabilities of NeuroGeMS and enabling users to explore novel approaches in multimodal learning.

By providing a comprehensive set of multimodal learning strategies, including unimodal analysis, early fusion, and late fusion, NeuroGeMS empowers researchers to effectively analyze and integrate multimodal data. The software's flexible and extensible framework facilitates the exploration of different approaches and the development of innovative solutions for multimodal learning in the biomedical domain.

### 6.3.3 Explainability

Explainability is a crucial aspect of machine learning models, particularly in the biomedical domain, where understanding the factors contributing to a prediction is essential for trust, transparency, and informed decision-making. NeuroGeMS recognizes the importance of explainability and incorporates the SHAP (SHapley Additive exPlanations) framework to provide insights into the model's predictions.

SHAP is a game-theoretic approach that assigns each feature an importance value, known as the Shapley value, which represents the feature's contribution to the model's output. By computing Shapley values for each feature, SHAP provides a unified and interpretable measure of feature importance, enabling users to understand how each feature influences the model's predictions (Lundberg and Lee, 2017).

NeuroGeMS integrates the SHAP library, allowing users to generate explanations for their trained models with ease. The software supports various types of SHAP explanations, including:

- **Global Explanations:** Global explanations provide an overall understanding of the model's behavior by quantifying the importance of each feature across the entire dataset. NeuroGeMS enables users to visualize global explanations using summary plots, which display the average absolute Shapley values for each feature. This helps identify the most influential features driving the model's predictions.
- **Local Explanations:** Local explanations focus on explaining individual predictions by quantifying the contribution of each feature to a specific instance. NeuroGeMS allows users to generate local explanations for any desired instance, providing a detailed breakdown of how each feature contributes to the model's output. This is particularly useful for analyzing specific cases and understanding the factors leading to a particular prediction.

The integration of SHAP in NeuroGeMS empowers users to interpret and explain their models effectively. By providing both global and local explanations, the software enables researchers to gain insights into the overall behavior of the model and the factors influencing individual predictions. This transparency enhances trust in the model's outputs and facilitates the validation of the model's decision-making process.

Furthermore, the multimodal explanations offered by NeuroGeMS are particularly valuable in understanding the complex interactions between different data modalities. By quantifying the contributions of features from each modality, users

can identify the most informative modalities and assess their relative importance in the prediction task. This knowledge can guide further analysis, feature selection, and model refinement.

NeuroGeMS also provides interactive visualizations and reports for SHAP explanations, making it easy for users to explore and communicate the model's behavior to stakeholders. The software generates intuitive plots, such as force plots and dependence plots, which illustrate the impact of features on the model's predictions. These visualizations aid in the interpretation and dissemination of the model's explainability, facilitating collaboration and knowledge sharing among researchers.

By incorporating SHAP-based explainability, NeuroGeMS takes a significant step towards responsible and transparent AI in the biomedical domain. The ability to understand and explain the model's predictions is essential for building trust, validating the model's decisions, and enabling informed decision-making. With NeuroGeMS, users can harness the power of explainable AI to gain deeper insights into their multimodal models and enhance the interpretability of their results.

### 6.3.4 Personalized Modeling

Personalized modeling is an essential aspect of healthcare analytics, as it aims to tailor predictions and recommendations to individual patients based on their unique characteristics and data. NeuroGeMS recognizes the importance of personalized modeling and provides a suite of methods specifically designed to capture individual-level heterogeneity and improve the accuracy and relevance of predictions.

The software incorporates two state-of-the-art personalized modeling techniques: Weighted Weighted K-Nearest Neighbors (WWKNN) (Kasabov, 2007a) and Transductive Weighted Neuro-Fuzzy Inference (TWNFI) (Song and Kasabov, 2006). These methods leverage the principles of local learning and instance-based reasoning to build personalized models that adapt to the specific characteristics of each patient.

#### Weighted Weighted K-Nearest Neighbors (WWKNN)

WWKNN is an extension of the traditional K-Nearest Neighbors (KNN) algorithm that incorporates both instance weights and feature weights to capture the importance of individual instances and features in the local neighborhood of a target instance (Kasabov, 2007a). The method assigns higher weights to instances that are more similar to the target instance and features that are more relevant for the prediction task (Doborjeh et al., 2019).

NeuroGeMS implements the WWKNN algorithm, allowing users to build personalized models based on the local neighborhood of each patient. The software provides an intuitive interface for configuring the WWKNN parameters, such as the number of nearest neighbors (K) and the weighting schemes for instances and features. Users can choose from various distance metrics and weighting functions to customize the model to their specific data and problem domain.

The WWKNN approach in NeuroGeMS enables the creation of patient-specific models that capture the unique characteristics and relationships within each patient's local neighborhood. By assigning higher weights to more similar instances and relevant features, WWKNN can adapt to the individual patterns and dependencies present in the data, leading to more accurate and personalized predictions.

**Transductive Weighted Neuro-Fuzzy Inference (TWNFI)**

TWNFI is a personalized modeling technique that combines the principles of transductive learning, neuro-fuzzy inference, and weighted instance-based reasoning (Song and Kasabov, 2006). The method builds a local neuro-fuzzy model for each target instance, leveraging the information from the most similar instances in the training set.

NeuroGeMS incorporates the TWNFI algorithm, providing users with a powerful tool for personalized modeling. The software allows users to specify the parameters of the TWNFI model, such as the number of fuzzy rules, the membership function type, and the optimization criteria. Users can also choose from various similarity measures and weighting schemes to customize the model to their specific data and problem domain.

The TWNFI approach in NeuroGeMS enables the creation of interpretable and personalized neuro-fuzzy models for each patient. By learning from the most similar instances and adapting the fuzzy rules and membership functions to the local characteristics of each patient, TWNFI can capture complex non-linear relationships and provide intuitive explanations for the model's predictions.

The personalized modeling capabilities of NeuroGeMS, through WWKNN and TWNFI, empower researchers to build models that are tailored to the unique characteristics of each patient. These methods can uncover individual-level patterns and relationships that may be overlooked by global models, leading to more accurate and relevant predictions.

NeuroGeMS provides a user-friendly interface for applying these personalized modeling techniques to various biomedical datasets, including multimodal data. The software handles the data preprocessing, feature selection, and model training steps, allowing users to focus on interpreting the results and making informed decisions.

By incorporating personalized modeling techniques like WWKNN and TWNFI, NeuroGeMS enables the development of precision medicine applications that can adapt to the specific needs and characteristics of individual patients. This personalized approach has the potential to improve patient outcomes, optimize treatment strategies, and support individualized decision-making in healthcare.

### 6.3.5 Experiment Tracking

Experiment tracking and reproducibility are critical aspects of scientific research, particularly in the field of machine learning and healthcare analytics. NeuroGeMS recognizes the importance of these factors and provides a comprehensive page for tracking experiments, managing results, and ensuring reproducibility by integrating MLflow, a state-of-the-art platform for the complete machine learning lifecycle (Zaharia et al., 2018).

MLflow is an open-source platform that provides a unified interface for tracking experiments, packaging code into reproducible runs, and sharing and deploying models. NeuroGeMS seamlessly integrates MLflow, allowing users to leverage its powerful features for experiment management, version control, and reproducibility.

NeuroGeMS utilizes MLflow's tracking component to manage and organize experiments. Users can create new experiments, specify the dataset, select the desired preprocessing steps, and configure the model architecture and hyperparameters. MLflow automatically logs all the relevant information associated with each experiment, including the parameters, metrics, and artifacts.

The MLflow integration in NeuroGeMS enables users to track and compare multiple runs of an experiment, each with different parameter configurations. The software provides an intuitive web interface for visualizing and comparing the results of different runs, making it easy to identify the best-performing models and optimize the hyperparameters.

With NeuroGeMS and MLflow, researchers can have confidence in the reliability and reproducibility of their experiments, promoting trust and credibility in the scientific community. The software's experiment tracking and reproducibility capabilities set a strong foundation for conducting high-quality research in the biomedical domain, ultimately contributing to improved patient outcomes and data-driven decision-making in healthcare.

### 6.3.6 Extensibility

Extensibility is a key design principle in NeuroGeMS, ensuring that the software can adapt to the evolving needs of biomedical research and accommodate new methodologies and technologies. The software is built with a modular and flexible architecture that allows for easy integration of new preprocessing techniques, machine learning models, and multimodal strategies without requiring changes to the core functionality or handlers.

NeuroGeMS follows a modular architecture, where different components of the software are designed as independent modules. Each module encapsulates a specific functionality, such as data preprocessing, feature selection, model training, or evaluation. The modular design allows for seamless addition, removal, or modification of modules without affecting the overall system.

This modular architecture is particularly beneficial when it comes to extending NeuroGeMS with new preprocessing techniques or machine learning models. Researchers and developers can create new modules that implement novel algorithms or techniques and easily integrate them into the existing pipeline. The software provides well-defined interfaces and guidelines for developing these modules, ensuring compatibility and smooth integration.

Similarly, the modular design of NeuroGeMS facilitates the incorporation of new multimodal strategies. As discussed in Section 6.3.2, the software currently supports unimodal analysis, early fusion, and late fusion strategies. However, the extensible architecture allows for the straightforward addition of new multimodal strategies, such as co-learning or translation, without requiring modifications to the core functionality or handlers.

The extensibility of NeuroGeMS is further enhanced by its adherence to standard data formats and conventions. The software relies on widely adopted data structures and formats, such as NumPy arrays and Pandas DataFrames, for data representation and processing. This standardization makes it easier to integrate new preprocessing techniques, machine learning models, or multimodal strategies that are compatible with these formats.

Moreover, NeuroGeMS provides comprehensive documentation and guidelines for extending its functionality. The documentation includes detailed instructions

on how to create new modules, implement custom preprocessing techniques, integrate machine learning models, and add multimodal strategies. This documentation serves as a valuable resource for researchers and developers who wish to extend NeuroGeMS to meet their specific requirements.

The extensibility of NeuroGeMS is a crucial aspect of its design, enabling researchers to adapt the software to their evolving needs and incorporate new advancements in biomedical research. By providing a modular architecture that allows for easy integration of preprocessing techniques, machine learning models, and multimodal strategies, NeuroGeMS establishes itself as a flexible and future-proof platform for biomedical research.

This extensibility empowers researchers to leverage the latest methodologies and technologies in their studies without being constrained by the limitations of the software. As new preprocessing techniques emerge, researchers can easily integrate them into NeuroGeMS to enhance data quality and feature representation. Similarly, as novel machine learning models are developed, they can be seamlessly incorporated into the software to improve prediction accuracy and performance.

Furthermore, the extensibility of NeuroGeMS fosters collaboration and knowledge sharing among the research community. Researchers can develop and share their own extensions, preprocessing techniques, or machine learning models, contributing to the collective advancement of biomedical research. The modular design of NeuroGeMS ensures that these contributions can be easily integrated and utilized by other researchers, promoting reproducibility and accelerating scientific progress.

In summary, the extensibility of NeuroGeMS is a fundamental aspect of its design, allowing for the seamless integration of new preprocessing techniques, machine learning models, and multimodal strategies without requiring changes to the core functionality or handlers. This extensibility empowers researchers to adapt the software to their specific needs, leverage the latest advancements in biomedical research, and contribute to the collective knowledge of the scientific community.

## 6.4 Code Architecture and Implementation

### 6.4.1 Overview

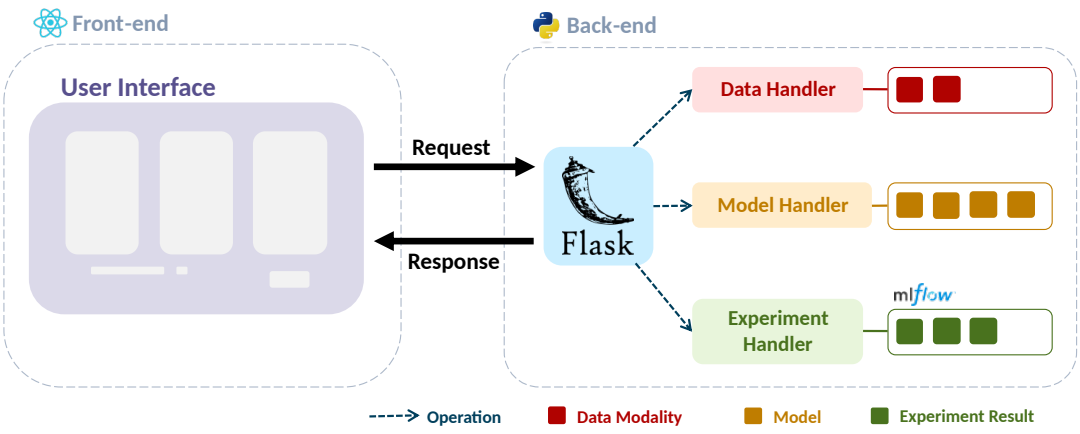


FIGURE 6.6: Overview of the NeuroGeMS software architecture. The front-end, based in React.js, interacts with the back-end, based in Flask.

NeuroGeMS blends web technologies for UI with Python's robust data processing capabilities to create a versatile and user-friendly platform for multimodal data analysis. NeuroGeMS is based on the Electron-React-Flask template[1], which bridges the gap between desktop and web applications, offering the responsiveness and accessibility of web interfaces with the native capabilities of desktop software.

The use of Electron in NeuroGeMS provides a consistent user experience across different operating systems while leveraging the rich ecosystem of web development. Electron facilitates the integration of React for the frontend, enabling a dynamic and component-driven user interface, while Flask serves as a lightweight yet powerful backend, capable of handling complex data operations with ease.

The UI is based on the Minimal Material UI kit[2] ensuring that the application not only functions effectively but also appeals visually to the end-users, promoting ease of use and enhancing user engagement.

Below is a summary of the primary technologies and packages used in NeuroGeMS, and the roles they play in the software:

| Package | Type | Purpose |
|---|---|---|
| React | npm | Powers the interactive user interface with component-based structures. |
| Electron | npm | Facilitates cross-platform desktop application development. |
| Flask | Python | Acts as the backend server, offering a lightweight API layer. |
| Axios | npm | Manages asynchronous HTTP requests for dynamic data interaction. |
| Material-UI | npm | Provides a suite of UI components for a cohesive user interface design. |
| MLflow | Python | Tracks experiments, manages models, and streamlines workflow. |
| Pandas | Python | Offers extensive data manipulation and analysis functionalities. |
| scikit-learn | Python | Implements a broad range of efficient machine learning algorithms. |

The subsequent sections dive deeper into the frontend, backend, and their integration, discussing how each contributes to the overarching goal of facilitating advanced data analysis in a user-centric application.

### 6.4.2 Backend: Python Flask

The backend of NeuroGeMS, built with the Python Flask framework, is the engine behind the application's data processing, analysis, and interaction capabilities. Flask's lightweight and extensible nature makes it an ideal choice for developing a responsive and scalable backend, capable of handling the complex workflows inherent in multimodal data analysis in biomedical research.

The Flask backend serves as a conduit between the frontend user interface and the underlying data processing and analysis logic. It is responsible for receiving HTTP requests from the frontend, processing these requests through various internal components, and sending back the appropriate responses. This architecture ensures a clear separation of concerns, with the backend focusing on data handling, computational tasks, and business logic, while the frontend concentrates on user interaction and presentation.

The organization of the backend into distinct components aligns with the principles of modularity and reusability. Each component, encapsulated in its respective Python module, is dedicated to a specific aspect of the backend's functionality. In

[1] Based on iPzard's electron-react-python-template: https://github.com/iPzard/electron-react-python-template

[2] Minimal Material UI kit: https://github.com/minimal-ui-kit/material-kit-react

the following subsections, we will explore the inner workings of the NeuroGeMS backend and highlight the application's capability to support advanced data analysis and modeling in the context of biomedical research.

#### Data Handling (`data.py`)

The `Data` module is core to NeuroGeMS, providing sophisticated data management capabilities essential for handling multimodal biomedical data. It encapsulates the complexities of data manipulation, ensuring that data across different modalities are accessible and analyzable within the system.

- `DataHandler` **Class:** Central to the data handling module is the DataHandler class, a versatile manager for datasets within NeuroGeMS. It serves as a repository, storing data in Pandas DataFrames, a choice reflecting the need for efficient and flexible data manipulation. This class facilitates the addition, retrieval, and basic management of datasets, ensuring a streamlined flow of data throughout the application.
- **Dataset Registration:** The `add_dataset` method allows the registration of new datasets into the system, where each dataset is indexed and stored as a `Pandas DataFrame`. This method is designed to handle various dataset specifications, accommodating the diverse data types encountered in multimodal biomedical analysis.
- **Exploratory Data Analysis (EDA):** Leveraging `ydata_profiling`, the module offers automated exploratory data analysis, generating comprehensive reports that provide insights into the data's structure, distribution, and potential quality issues. This functionality enriches the data handling process, offering a nuanced understanding of datasets prior to further processing or analysis.

The `Data` module stands as a cornerstone within NeuroGeMS's backend architecture, exemplifying the system's commitment to robust and nuanced data management. Through its sophisticated handling of diverse data types and its integration of advanced EDA tools, the module significantly contributes to the system's analytical rigor and operational efficiency.

#### Multimodal Strategies (`strategy.py`)

The `Strategy` module in NeuroGeMS, along with its auxiliary modules, orchestrates the application of unimodal and multimodal learning strategies, accommodating the intricate nature of multimodal data analysis in biomedical research.

- **Base Strategy Class:** At the foundation is the `Strategy` abstract base class defined in `base.py`, which outlines the essential structure and methods that all specific strategy classes must implement. This abstraction ensures a consistent approach to strategy execution, allowing for flexibility and extendibility in strategy definition.
- **Unimodal Strategy:** The `UnimodalStrategy` class (`unimodal.py`) addresses scenarios where a single data modality is analyzed. This strategy encapsulates the process of applying a model to a single dataset, providing a foundational approach to learning that can be extended or used directly for unimodal data analysis.

- **Early Fusion Strategy:** The `EarlyFusionStrategy` (`early_fusion.py`) merges data from different modalities at the feature level before model training, enabling the model to learn from a unified representation of the multimodal data. This approach is beneficial when the interplay of features across modalities is crucial for the learning task.
- **Late Fusion Strategy:** Conversely, the `LateFusionStrategy` (`late_fusion.py`) maintains modality-specific models that are later combined, typically at the decision level. This strategy allows individual modalities to be processed and learned separately, with their predictions or features integrated post hoc, offering flexibility in handling diverse data characteristics.
- **Integration and Execution:** The `StrategyHandler` provides a unified interface for their execution. It supports the dynamic selection and application of strategies based on the analysis requirements, facilitating a versatile and adaptable analysis framework within NeuroGeMS.

Through the `Strategy` module and its related components, NeuroGeMS offers a robust and flexible framework for applying diverse learning strategies. Whether dealing with single or multiple data modalities, the system provides tailored approaches to harness the unique characteristics and insights that each modality offers, underpinning the advanced analytical capabilities of NeuroGeMS.

#### Model Definition and Training (`model.py`)

The `Model` module within NeuroGeMS provides a robust framework for defining, training, and managing a diverse array of machine learning models. This module underscores the system's flexibility, supporting standard models from established libraries and custom models for specialized tasks.

- **Support for Standard and Custom Models:** NeuroGeMS accommodates a wide variety of standard machine learning models, such as classifiers and regressors from the scikit-learn library. Beyond these, the system is distinctively equipped to integrate custom models, enabling users to address unique analytical challenges with tailored solutions.
- **Custom Personalized Modeling:** Emphasizing personalized healthcare, NeuroGeMS includes specialized models like Transductive Weighted Neuro Fuzzy Inference (TWNFI) and Weighted Weighted K-Nearest Neighbors (WWKNN), which are particularly adept at handling individual patient data and incorporating personalization into the analysis:
  - **Transductive Neuro Fuzzy Inference (TNFI):** TNFI combines fuzzy logic with learning algorithms to model complex, nonlinear relationships in data. It is particularly effective in scenarios where individualized predictions are crucial, making it a valuable asset for personalized medicine within NeuroGeMS.
  - **Weighted Weighted K-Nearest Neighbors (WWKNN):** The WWKNN model enhances the conventional KNN approach by introducing weights for both instances and features, providing a more nuanced and adaptive mechanism for classification or regression tasks. This model exemplifies the system's capability to deliver personalized insights, tailoring predictions to individual data points.

- **Model Training:** Within NeuroGeMS, model training is a streamlined process, with the `model.py` module facilitating the efficient configuration, training, and optimization of both standard and custom models. This ensures that the system can adapt to various data characteristics and research objectives, providing robust and reliable analytical outcomes.

The `Model` component is a cornerstone of NeuroGeMS, reflecting the system's comprehensive and adaptable approach to data analysis in the biomedical domain. By supporting a blend of standard and personalized models, NeuroGeMS empowers researchers to harness advanced computational techniques for insightful, patient-centered analysis.

#### Validation and Testing (`validation.py`)

The `Validation` module in NeuroGeMS plays a crucial role in the system's analytical pipeline, focusing on the robust validation and testing of the models developed within the framework. This module ensures that the models are not only trained effectively but also evaluated rigorously to guarantee their reliability and validity in real-world applications.

- **Integration with Model Training:** While the core functionalities of validation are encapsulated within `validation.py`, the process is closely integrated with model training routines described in `model.py`. This integration ensures a cohesive workflow where models are not only trained with optimal parameters but also validated using best practices in machine learning.
- **Cross-validation Techniques:** NeuroGeMS employs a variety of cross-validation techniques to assess model performance. These techniques, which include K-Fold and Stratified K-Fold cross-validation, are essential for understanding how models perform on unseen data, thereby mitigating overfitting and underfitting issues. Such rigorous validation approaches contribute to the development of models that are robust and generalizable across different datasets.
- **Testing and Robustness Checks:** Beyond validation, `validation.py` also oversees the final testing phase, where models are evaluated on a separate test set to simulate real-world application. This phase is crucial for assessing the model's robustness and its ability to generalize beyond the training data.

The `Validation` component of NeuroGeMS underscores the system's commitment to quality and reliability in model development. Through comprehensive validation and testing processes, the system ensures that its predictive models are not only accurate but also robust and dependable for critical applications in biomedical research and healthcare.

#### Experiment Tracking (`experiment.py`)

The `Experiment` module in NeuroGeMS is dedicated to the systematic tracking of experiments, ensuring that every detail from the model training and validation processes is meticulously recorded and accessible for future reference and analysis.

- **Integration with MLflow:** Central to the experiment tracking process is the integration with MLflow, a platform that specializes in managing the end-to-end machine learning lifecycle. This integration allows NeuroGeMS to leverage

MLflow's robust capabilities for tracking experiments, recording parameters, metrics, and results, thereby enhancing reproducibility and transparency in the research process.

- **Model Evaluation Metrics:** The module maintains a comprehensive list of supported metrics, including accuracy, precision, recall, F1 score, and ROC AUC score, among others. These metrics are crucial for evaluating the performance of models across different experimental setups, providing a basis for comparative analysis and informed decision-making.
- **Experiment Logging:** Experiment tracking involves logging various details, such as model parameters, training data characteristics, and evaluation metrics. This logging ensures that researchers can revisit experiments, understand the context and settings under which models were trained and evaluated, and replicate results or further refine the models.
- **Result Analysis and Comparison:** By systematically tracking experiments, NeuroGeMS facilitates the analysis and comparison of different models and strategies. Researchers can review past experiments, draw insights from accumulated data, and make evidence-based decisions to enhance model performance and address research questions effectively.

The `Experiment` component embodies NeuroGeMS's commitment to rigorous and structured research practices. By providing a detailed record of experiments and their outcomes, the module supports the continuous improvement of models, fosters transparency in the research process, and contributes to the overall reliability and credibility of the system's analytical capabilities.

### 6.4.3 Frontend: React and Material-UI

The frontend of NeuroGeMS is built using React, a declarative and efficient JavaScript library for constructing user interfaces (UI), and Material-UI, a popular React UI framework that follows the Material Design guidelines. This combination provides a robust foundation for developing a user-friendly and aesthetically pleasing interface, facilitating seamless interaction with the system's comprehensive backend functionalities.

React, known for its efficiency and flexibility, enables NeuroGeMS to manage its dynamic content and stateful interactions with ease. The library's declarative nature simplifies the process of building interactive UIs, where the state changes dictate the rendering logic, ensuring that the UI is always in sync with the underlying data. React's component-based architecture facilitates the development of interactive user interfaces, allowing for modular and reusable code that enhances the application's maintainability and scalability.

#### Design and Usability

A key factor that shaped the interface design was the adoption of Material UI (MUI), a modern React component library grounded in Google's Material Design system. MUI offers a consistent and professional aesthetic along with strong support for accessibility and responsiveness. Its component abstraction significantly accelerated development while ensuring that key interface elements — such as tabbed modality views, explainability visualisations, and model configuration panels — remained

clear and easy to navigate. The framework's responsiveness and theme customization capabilities allow NeuroGeMS to adapt to various devices and user preferences, ensuring a broad accessibility range.

The design of the interface was shaped through ongoing collaboration with researchers at the Institute of Mental Health (IMH) Singapore, Nanyang Technological University (NTU) Singapore, Nottingham Trent University (NTU) UK, and the Auckland Bioengineering Institute (ABI). Regular demonstrations and development sessions provided informal but targeted feedback, leading to refinements in layout, interaction flow, and output presentation. This co-design process helped ensure that the tool met the practical needs of researchers engaging with multimodal data and explainable AI outputs.

#### UI Components (`components/`)

The `components/` directory in NeuroGeMS's frontend architecture encapsulates a variety of UI components, each serving a distinct purpose in the application's user interface. These components, developed with React and styled using Material-UI, contribute to a modular and maintainable codebase, facilitating ease of development and consistency across the platform.

- **Component Structure:** Each subdirectory within `components/` represents a specific UI element or a group of related elements, showcasing the application's modular design approach. This structure not only enhances code reusability but also streamlines the development process, allowing for isolated testing and maintenance of individual components.
  - **Chart Components:** The `chart` directory likely contains components for data visualization, enabling users to interact with and interpret complex data sets visually.
  - **Utility Components:** Directories like `color-utils` and `iconify` suggest the presence of utility components that provide specific functionalities, such as color manipulation or icon integration, enhancing the application's UI flexibility and aesthetic appeal.
  - **Interactive Elements:** Components like `counter` and `scroll-to-top` indicate interactive UI elements that enhance user engagement and navigation efficiency within the application.
  - **Layout Components:** Elements such as `nav-section` and `titlebar` are crucial for structuring the application's layout, ensuring a coherent and user-friendly navigation experience.
  - **Branding Elements:** The `logo` component underscores the incorporation of branding elements, vital for maintaining a consistent visual identity across the platform.
- `App.js` **and** `routes.js`**:** The `App.js` and `routes.js` files play eseential roles in the application's main structure and routing logic. `App.js` serves as the root component, where the overall layout and component hierarchy are defined, while `routes.js` manages the application's routing, ensuring that users can navigate through different views and functionalities efficiently.

The `components/` directory exemplifies the frontend's commitment to modularity and reusability, with each component designed to serve a specific function within

the user interface. This approach not only streamlines the development process but also enhances the application's maintainability and scalability, ensuring that NeuroGeMS remains adaptable to evolving user needs and functionalities.

#### Application Pages (`pages/`, `sections/`)

In NeuroGeMS, the `pages/` directory serves as a container for the main views of the application, each represented by a JavaScript file that corresponds to a distinct page in the UI. These pages are crucial in defining the user journey, presenting data, and allowing interactions with the system's functionalities.

- **Core Pages:** The individual page files, such as `DataPage.js`, `ExperimentPage.js`, and `ModelPage.js`, define the layout and content of each respective page. These files orchestrate the assembly of various components and sections to present information and interactive elements aligned with the page's purpose.
  - `HomePage.js`: Serves as the landing page of the application, providing users with an overview and quick access to various functionalities.
  - `DataPage.js`: Defines the layout and functionality of the page where users interact with and manage their data, essential for the initial stages of the data analysis workflow.
  - `ModelPage.js`: Focuses on model management, allowing users to define and visualise different unimodal and multimodal models, central to the application's analytical capabilities.
  - `ExperimentPage.js`: Represents the interface where users can configure, execute, and review experiments, playing a pivotal role in the model training and evaluation process.
  - `SettingsPage.js`: Offers users the ability to customize and configure application settings, enhancing the adaptability and personalization of the user experience.
  - `Page404.js`: A standard error page, ensuring users are guided back to the application's functional areas when an invalid URL is accessed.
- **Modular Sections:** NeuroGeMS employs a modular approach to organizing content. By segmenting pages into sections, the application promotes a clear separation of concerns and facilitates the maintenance and enhancement of individual UI elements. This underscores the modularity and reusability within NeuroGeMS's frontend.

The `pages/` and `sections/` directories are foundational to the NeuroGeMS user interface, dictating the layout, structure, and interaction patterns of the application. By meticulously organizing these core components, NeuroGeMS ensures a user-friendly environment that supports comprehensive data analysis workflows, model management, and system configuration, all while maintaining a clear and intuitive navigation schema.

#### State Management

State management is pivotal in React applications, dictating how data is stored, modified, and passed throughout the application. In NeuroGeMS, `useState` and `useEffect` are essential hooks that facilitate dynamic and responsive user interfaces by efficiently managing local component states and side effects.

- `useState` **Hook:** Imagine `useState` as a personal notebook for each component in the application. This notebook keeps track of important details that can change over time, like a user's input or a selection on the screen. When something is written or modified in the notebook (`useState`), the component knows it needs to update what's shown on the screen, making sure that everything the user sees is current and relevant.

- `useEffect` **Hook:** `useEffect` acts as a vigilant observer in the application. It watches for specific changes and acts when it's necessary. Think of it as a helper who runs to fetch data or perform tasks whenever something important happens or changes. It ensures that the application reacts appropriately to user interactions or other events, keeping the user interface synchronized with the underlying data.

- **Integration of** `useState` **and** `useEffect`**:** In practice, NeuroGeMS combines `useState` and `useEffect` to create interactive and dynamic pages that respond to user actions and system changes. For example, `useState` might track user inputs on a form, while `useEffect` could handle data fetching based on those inputs, updating the UI in response to new data or user interactions.

These React hooks, `useState` and `useEffect`, are foundational to creating an environment where every click, every input, and every action leads to a natural and immediate response on the screen. They ensure that NeuroGeMS remains not just a tool, but an interactive companion for its users, capable of adapting and responding to their every need in real-time.

#### Utility Functions (`utils/`)

The `utils/` directory is an integral part of the NeuroGeMS frontend, offering a suite of utility functions that bolster various application functionalities. These utilities enhance the user experience, ensure data consistency, and facilitate communication with the backend.

- **Styling Utilities (**`cssStyles.js`**):** Contains CSS styles and themes for consistent and appealing visual elements across the application, including background effects, text styles, and scrollbar customizations.

- **Data Formatting:**

  - `formatNumber.js` provides functions to format numerical data, ensuring consistency and readability across user interfaces.
  - `formatTime.js` includes methods to standardize time representation, crucial for displaying temporal data accurately.

- **Network Requests (**`requests.js`**):** Handles HTTP GET and POST requests, enabling efficient data exchanges between the frontend and backend, essential for dynamic data interactions and server communications.

- **Application Services (**`services.js`**):** Offers functions for window management, such as minimizing, maximizing, and closing the application window, integrating desktop-like functionalities within the web application.

The `utils/` directory's utility functions are foundational to the operational efficiency and user-centric design of NeuroGeMS. By centralizing these key functionalities, the application maintains a high standard of performance and user interaction quality, supporting the sophisticated requirements of biomedical data analysis.

**Theming and Styling (`theme/`, `index.scss`)**

NeuroGeMS's frontend architecture includes a comprehensive theming and styling system, facilitating a cohesive visual experience across the application. The `theme/` directory contains the core theming functionalities, while `index.scss` provides global style definitions applied throughout the application.

- **Theme Customization (`theme/`):** The theme directory contains JavaScript and SCSS files that define the visual aspects of the application, including colors, typography, shadows, and more.
  - `palette.js` and `shadows.js` specify the color schemes and shadow styles, contributing to a consistent look and feel.
  - `typography.js` sets the typography standards, ensuring text is readable and aesthetically pleasing across different UI components.
  - `globalStyles.js` and `customShadows.js` provide additional styling details that can be applied globally or customized per component.
  - `themeContext.js` handles the context for theming, allowing for dynamic theme changes or user-preference-based theming.
  - The `overrides/` directory contains specific style overrides for Material-UI components, ensuring that even third-party components align with the overall design language.
- **Global Styles (`index.scss`):** This SCSS file defines global CSS styles that are applied across the application. For instance, the body and code block styles set the foundational font families and smoothing properties, which are crucial for maintaining a consistent and professional look.

Through its dedicated theming and styling structure, NeuroGeMS achieves a flexible and uniform design that enhances user interaction and accessibility. The careful separation of theme definitions and global styles allows for scalable and maintainable frontend development, supporting the application's adaptability to evolving design standards and user needs.

### 6.4.4 Integration and Communication

This section highlights the seamless orchestration between NeuroGeMS's frontend and backend, facilitated by Electron's main process (main.js). This script acts as the nerve center, initializing the application, creating browser windows, and managing inter-process communication.

- **Application Initialization:** `main.js` initiates the Electron app, setting up the main browser window and loading the React frontend. In development mode, it also handles the loading of developer tools and the Flask server, ensuring that all necessary components are in place for the application to function.

- **Browser Window Management:** This script is responsible for creating and managing Electron's browser windows, where the React frontend is rendered. It manages different application states, like maximizing, minimizing, and closing, ensuring a native app-like user experience.
- **Inter-Process Communication (IPC):** Using Electron's `ipcMain` module, `main.js` establishes communication channels between the Electron main process and the renderer process (React frontend). This setup allows for the execution of functions like app quitting, minimizing, and maximizing from within the React application, ensuring responsive user interactions.
- **Flask Server Integration:** `main.js` dynamically assigns a port and starts the Flask backend server, enabling the React frontend to communicate with the backend via HTTP requests. This integration is crucial for data processing, retrieval, and other server-side operations essential to NeuroGeMS's functionality.
- **API Endpoints and Data Exchange:** NeuroGeMS utilizes a set of API endpoints defined in the Flask backend to facilitate data exchange between the frontend and backend. The frontend, built with React, communicates with these endpoints using HTTP requests to retrieve or send data. This interaction is crucial for executing data processing tasks, fetching analysis results, and managing user inputs and outputs. The React components interact with these endpoints to perform CRUD operations, ensuring dynamic and real-time data handling within the application.
- **Shutdown Procedure:** NeuroGeMS includes a graceful shutdown process, ensuring that all subprocesses, including the Flask server, are terminated when the application closes, preventing any orphan processes or resource leaks.

The integration and communication mechanisms within NeuroGeMS, demonstrate the application's cohesive architecture. By bridging the frontend and backend through Electron's robust framework, NeuroGeMS ensures a synchronized and efficient workflow, essential for its operation as an advanced data analysis tool.

## 6.5 Case Studies and Applications

To demonstrate the practical applications and effectiveness of NeuroGeMS, two case studies utilizing the LYRIKS and ADNI (TADPOLE) biomedical datasets are presented. These case studies illustrate how NeuroGeMS handles complex datasets, integrates diverse data modalities, and provides interpretable results that can significantly enhance biomedical research.

### 6.5.1 Case Study #1: LYRIKS Dataset

The LYRIKS dataset contains longitudinal data of individuals with Ultra-High Risk (UHR) for psychosis, along with control samples. This dataset includes a battery of assessments, providing a comprehensive view of the participants' status. The objective was to classify the participants' UHR status (healthy or at-risk) using baseline data from the cognitive and social modalities.

Four machine learning models were employed: Support Vector Machine (SVM) (Cortes and Vapnik, 1995), Gradient-boosted Decision Trees (XGBoost) (Chen and

Guestrin, 2016), Logistic Regression (LogReg) (Cox, 1958), and Weighted Weighted K Nearest Neighbors (WWKNN) (Kasabov, 2007a). These models were evaluated using 5-fold cross-validation on accuracy, Matthews correlation coefficient (MCC), and area under the curve (AUC) metrics. The results are summarized in Table 6.1.

TABLE 6.1: LYRIKS UHR prediction results using cognitive and social modalities

| Modality | Model | Accuracy | MCC | AUC | F1-Score |
|---|---|---|---|---|---|
| Cognitive | SVM | 0.643 | 0.284 | 0.643 | 0.641 |
| | XGBoost | 0.650 | 0.299 | 0.650 | 0.670 |
| | LogReg | 0.645 | 0.284 | 0.642 | 0.666 |
| | WWKNN | 0.626 | 0.248 | 0.624 | 0.646 |
| Social | SVM | 0.583 | 0.156 | 0.577 | 0.577 |
| | XGBoost | 0.603 | 0.196 | 0.594 | 0.607 |
| | LogReg | 0.583 | 0.158 | 0.578 | 0.602 |
| | WWKNN | 0.572 | 0.140 | 0.570 | 0.596 |
| Early Fusion | SVM | 0.660 | 0.355 | 0.672 | 0.664 |
| | XGBoost | 0.643 | 0.282 | 0.641 | 0.641 |
| | LogReg | 0.660 | 0.319 | 0.660 | 0.662 |
| | WWKNN | 0.654 | 0.304 | 0.652 | 0.652 |
| Late Fusion | XGBoost (Both) | 0.643 | 0.328 | 0.657 | 0.641 |

The early fusion strategy, in particular, showed improved performance, highlighting the advantage of combining cognitive and social modalities at the feature level. These results demonstrate that NeuroGeMS can effectively integrate multimodal data.

### 6.5.2 Case Study #2: ADNI (TADPOLE) Dataset

The TADPOLE dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) contains data of individuals with mild cognitive impairment (MCI) and Alzheimer's disease along with control samples. Baseline data from the D1 dataset was utilized, focusing on cognitive and neuroimaging modalities to predict cognitive impairment status (healthy or impaired).

Similarly to the LYRIKS dataset, four machine learning models were employed: SVM, XGBoost, LogReg, and MLP. The performance of these models is summarized in Table 6.2.

The results from the TADPOLE dataset show that NeuroGeMS effectively integrates cognitive and neuroimaging data to produce highly accurate models for predicting impairment status. The early fusion strategy again demonstrates superior performance, highlighting the effectiveness of integrating diverse modalities at the feature level.

In conclusion, these case studies showcase the effectiveness of NeuroGeMS in handling multimodal biomedical data. By providing an intuitive interface and access to advanced machine learning algorithms, NeuroGeMS facilitates the analysis and interpretation of complex datasets, ultimately contributing to improved outcomes and advancing the field of biomedicine.

TABLE 6.2: TADPOLE Impairment prediction results using cognitive and neuroimaging modalities

| Modality | Model | Accuracy | MCC | AUC | F1-Score |
|---|---|---|---|---|---|
| Cognitive | SVM | 0.952 | 0.901 | 0.952 | 0.951 |
| | XGBoost | 0.951 | 0.899 | 0.949 | 0.959 |
| | LogReg | 0.950 | 0.897 | 0.949 | 0.957 |
| | WWKNN | 0.930 | 0.862 | 0.936 | 0.937 |
| Neuroimaging | SVM | 0.621 | 0.294 | 0.645 | 0.622 |
| | XGBoost | 0.617 | 0.220 | 0.611 | 0.622 |
| | LogReg | 0.634 | 0.283 | 0.643 | 0.635 |
| | WWKNN | 0.634 | 0.284 | 0.644 | 0.635 |
| Early Fusion | SVM | 0.965 | 0.929 | 0.961 | 0.965 |
| | XGBoost | 0.958 | 0.912 | 0.955 | 0.958 |
| | LogReg | 0.955 | 0.908 | 0.954 | 0.955 |
| | WWKNN | 0.935 | 0.869 | 0.938 | 0.935 |
| Late Fusion | SVM (Both) | 0.965 | 0.930 | 0.959 | 0.965 |

## 6.6 Conclusion

This chapter has introduced and detailed the development of NeuroGeMS, a graphical user interface (GUI) software designed to facilitate the adoption of advanced multimodal AI methods in research settings. By addressing the research objective of facilitating the adoption of multimodal AI methods and advancing research in the biomedical domain (Objective 5), NeuroGeMS aims to bridge the gap between AI innovation and practical application in mental health care.

NeuroGeMS provides a unified platform for the integration and analysis of diverse data types. The software's intuitive design and user-friendly interfaces enable researchers to perform comprehensive analyses without requiring extensive technical expertise. By incorporating advanced AI methods, NeuroGeMS ensures that users can leverage state-of-the-art tools for mental health diagnostics and prognostics.

Throughout the chapter, we reviewed existing tools and software used in mental health research, highlighting their strengths and limitations. This comparison underscored the need for a more integrated and accessible solution, which NeuroGeMS addresses through its design principles, architecture, and key features. The implementation process, including the integration of multimodal data handling capabilities and the development of user-friendly interfaces, was thoroughly described, demonstrating the meticulous approach taken to create a robust and practical tool.

Case studies and examples presented in this chapter showcased the practical applications of NeuroGeMS in real-world and research settings. These examples illustrated how NeuroGeMS could enhance the accuracy and efficiency of mental health diagnostics by combining multimodal data, support personalization and explainability, and facilitate innovative research.

Despite its strengths, NeuroGeMS is not without limitations. Future enhancements could focus on further improving its computational efficiency, expanding its data handling capabilities, and incorporating more advanced AI techniques as they emerge. While informal feedback from research collaborators guided the design of NeuroGeMS, a structured user study was not conducted during the scope of this thesis. As future work, a formal usability evaluation to assess interface usability,

workflow integration, and interpretability would provide quantitative and qualitative insights into user experience, aiding in iterative design and broader adoption of the tool in translational research contexts.

In conclusion, this chapter has made contributions to the field of mental health diagnostics by developing and validating NeuroGeMS, a user-friendly software that integrates multimodal data for comprehensive analysis. NeuroGeMS facilitates the adoption of advanced multimodal AI methods by the broader research community. By bridging the gap between AI research and medical practice, NeuroGeMS has the potential to enhance patient outcomes, improve mental health care, and drive forward the frontier of mental health research.

# Chapter 7

# Conclusion and Future Work

## 7.1 Review

This thesis set out to address the critical challenge of improving the diagnosis and prognosis of mental health disorders through the development and application of advanced artificial intelligence (AI) and machine learning (ML) methods. Recognizing the complexity and multifactorial nature of mental health conditions, the aim was to leverage multimodal and longitudinal data to create robust and interpretable predictive models. The work presented in this thesis spans several key areas, each contributing to the overarching goal of enhancing mental health diagnostics.

Mental health disorders are a significant global health concern, affecting millions of individuals and imposing substantial social and economic burdens. Traditional diagnostic methods, which rely heavily on clinical interviews and self-reported symptoms, often lack the precision and objectivity needed for early and accurate diagnosis. Furthermore, these methods do not effectively capture the dynamic and multifaceted nature of mental health conditions, which are influenced by genetic, neurobiological, psychological, and environmental factors.

The primary objectives of this thesis were to develop methods for integrating multimodal data sources to provide a comprehensive understanding of mental health disorders, enhance predictive accuracy using longitudinal data to capture temporal dynamics and identify early markers of mental health deterioration, create explainable AI models that provide transparent and interpretable predictions, and facilitate the adoption of multimodal AI methods by researchers through the development of user-friendly tools.

Chapter 2 provided a comprehensive review of the current methodologies in mental health diagnosis and prognosis, highlighting the limitations and challenges of existing approaches. Traditional diagnostic practices often rely on subjective assessments, leading to variability in outcomes. This review underscored the need for advanced AI models that can integrate diverse data sources and provide more objective and accurate predictions.

In Chapter 3, the Filter and Wrapper Stacking Ensemble (FWSE) method for biomarker discovery using gene expression data was introduced. This novel approach was designed to address the challenge of high-dimensional genetic data and its integration with other modalities. By combining filter and wrapper techniques, the FWSE method effectively identified significant genetic markers associated with mental health disorders. This chapter emphasized the importance of integrating genetic information to enhance the robustness of diagnostic models.

Chapter 4 focused on the development of explainable AI models for analyzing longitudinal data. The Dynamic Attention Gateway (DAG) for Liquid State Machines (LSMs) was introduced as a method that balances predictive accuracy with

interpretability. This chapter demonstrated how the DAG-LSM approach could capture temporal dynamics and provide insights into the factors driving model predictions. The importance of explainability in AI models was highlighted, emphasizing the need for transparent models in clinical settings.

In Chapter 5, the Mosaic Liquid State Machine (Mosaic LSM) method for integrating multimodal longitudinal data was presented. This approach leverages the strengths of LSMs in handling temporal data and combines them with techniques for multimodal data fusion. The chapter detailed the integration of various data types, such as clinical assessments, neuroimaging, genetic data, and digital phenotyping. The Mosaic LSM method was shown to improve predictive accuracy and provide a holistic view of an individual's mental health.

Chapter 6 introduced NeuroGeMS, a graphical user interface (GUI) software developed to make advanced AI tools accessible to clinicians and researchers. The software supports the integration and analysis of diverse data types, providing an intuitive platform for performing comprehensive analyses. This chapter detailed the design and implementation of NeuroGeMS, along with case studies demonstrating its practical applications in clinical and research settings.

The rationale behind this thesis is rooted in the urgent need for more accurate, objective, and interpretable methods for diagnosing and prognosing mental health disorders. Traditional approaches are often limited by their reliance on subjective assessments and the inability to integrate diverse data sources. By developing advanced AI methods that leverage multimodal and longitudinal data, the goal was to address these limitations and provide tools that can enhance clinical decision-making.

The impact of this research is multifaceted. The methods developed in this thesis contribute to a deeper understanding of the biological, psychological, and social factors that influence mental health. The emphasis on explainability ensures that the AI models can be trusted and effectively used by clinicians, thereby facilitating their adoption in clinical practice. The development of user-friendly tools like NeuroGeMS bridges the gap between AI research and practical application, making advanced analytical capabilities accessible to a broader audience.

In conclusion, this thesis has laid a solid foundation for the integration of advanced AI methods in mental health diagnostics. By addressing the complexities of mental health disorders through multimodal and longitudinal data integration, explainable AI models, and user-friendly tools, significant strides have been made toward improving the accuracy, transparency, and practicality of mental health care.

## 7.2 Key Contributions

The key contributions of this thesis can be summarized as follows:

### 7.2.1 Integration of Multimodal Data Sources

Chapter 5 introduced the Mosaic Liquid State Machine (Mosaic LSM) method for integrating multimodal longitudinal data. This approach addresses the challenge of combining diverse data sources, such as clinical assessments, neuroimaging, genetic data, and digital phenotyping, to create a holistic and accurate representation of

an individual's mental health. The Mosaic LSM method effectively manages high-dimensional and heterogeneous datasets, enhancing the predictive power of diagnostic models and providing a comprehensive understanding of mental health disorders.

### 7.2.2 Development of Explainable AI Models

Chapter 4 detailed the importance of explainability in AI models. The DAG-LSM method exemplifies this focus by offering interpretable predictions that clinicians can understand and trust. This work contributes to the broader field of explainable AI, addressing the need for transparency in AI-driven mental health diagnostics.

### 7.2.3 Enhancement of Predictive Accuracy Using Longitudinal Data

Chapter 4 explored the development of explainable AI models for analyzing longitudinal data. The Dynamic Attention Gateway (DAG) for Liquid State Machines (LSMs) was introduced as a novel method that provides both high predictive accuracy and interpretability. By effectively capturing temporal dynamics, this approach facilitates the identification of early markers and patterns critical for the diagnosis and prognosis of mental health disorders.

### 7.2.4 Discovering Markers of Mental Health

Chapter 3 introduced the Filter and Wrapper Stacking Ensemble (FWSE) method for biomarker discovery using gene expression data. This approach addresses the challenge of integrating high-dimensional genetic data with other modalities, such as clinical assessments and neuroimaging. The FWSE method has been shown to identify significant genetic markers associated with mental health disorders, enhancing the understanding of their biological underpinnings and contributing to more robust diagnostic models.

### 7.2.5 Facilitating the Adoption of Multimodal AI Methods by Clinicians

Chapter 6 presented NeuroGeMS, a graphical user interface (GUI) software designed to make advanced AI tools accessible to clinicians and researchers. By integrating multimodal data and providing user-friendly interfaces, NeuroGeMS bridges the gap between AI innovation and clinical practice. The software's practical applications, demonstrated through case studies, highlight its potential to enhance mental health care and support clinical decision-making.

## 7.3 Future Work

While this thesis has made significant strides in advancing the field of mental health diagnostics, several areas warrant further investigation and development:

### 7.3.1 Optimizing Computational Efficiency

The computational complexity of methods such as FWSE and DAG-LSM presents a challenge for real-time clinical application. Future research should focus on optimizing these algorithms to reduce computational demands without compromising their predictive accuracy and interpretability.

### 7.3.2 Expanding Data Handling Capabilities

As the availability of diverse data types continues to grow, expanding the data handling capabilities of NeuroGeMS and other AI tools will be crucial. This includes incorporating new data sources, such as wearable devices and electronic health records, to further enhance the comprehensiveness of multimodal models.

### 7.3.3 Broader Validation and Generalizability

Further validation studies across diverse clinical populations and settings are essential to ensure the generalizability and robustness of the developed models. These studies will help to identify potential biases and refine the models for broader application.

### 7.3.4 Incorporating Advanced AI Techniques

The field of AI is rapidly evolving, with new techniques and methodologies continuously emerging. Future work should explore the integration of advanced AI methods, such as deep reinforcement learning and federated learning, to further enhance the capabilities of mental health diagnostics.

## 7.4 Final Remarks

This thesis has made substantial contributions to the field of mental health diagnostics by developing and validating innovative AI methods that integrate multimodal and longitudinal data. The emphasis on explainability and usability ensures that these models can be effectively applied, bridging the gap between AI research and practical mental health care. The insights gained from this work lay a solid foundation for future research and development, ultimately aiming to improve patient outcomes and advance the field of mental health.

# Appendix A

# Sensitivity Analysis of FWSE's Pruning Factor

In this appendix, we aim to analyse the sensitivity of the pruning factor, which plays an important role in the feature elimination process in our FWSE method. Although a pruning factor of 0.5 was used in the study, this section explores how variations in this parameter influence the performance metrics of accuracy and stability.

The pruning factor specifies the proportion of features that are discarded during the initial filter-based feature selection phase of the FWSE algorithm. It is expressed as a fraction and takes on values within the range of 0 to 1. To conduct this analysis, we consider multiple values for the pruning factor—0, 0.2, 0.33, 0.5, 0.66, 0.8, and 1.

We evaluate the impact of changing the pruning factor on two distinct datasets: LYRIKS, which contains microarray gene expression data, and PDAC which has proteomics data and is characterized by high separability. For each dataset and pruning factor setting, we calculate the mean accuracy and mean stability that FWSE achieves across the top 50 features. The results of this analysis are presented in Table A.1.

TABLE A.1: Impact of varying pruning factor on accuracy and stability, averaged across top 50 features

| Pruning Factor | LYRIKS | | PDAC | |
|---|---|---|---|---|
| | Accuracy | Stability | Accuracy | Stability |
| 0 | 88.41 | 0.325 | 94.88 | 0.367 |
| 0.2 | 88.19 | 0.339 | 94.77 | 0.355 |
| 0.33 | 89.17 | 0.340 | 94.65 | 0.356 |
| 0.5 | 90.29 | 0.346 | 94.53 | 0.349 |
| 0.66 | 89.20 | 0.340 | 94.11 | 0.351 |
| 0.8 | 87.77 | 0.322 | 93.97 | 0.338 |
| 1 | 88.08 | 0.371 | 93.09 | 0.577 |

On LYRIKS, the accuracy of FWSE appears to peak at the pruning factor of 0.5, achieving a mean accuracy of 90.29%. The accuracy tends to decrease as the pruning factor deviates from this optimal value in either direction. A similar trend can be observed with FWSE's stability, except when the pruning factor is set to 1. This suggests that a pruning factor of 0.5 provides a balanced feature selection that maximizes classification accuracy for this particular dataset.

For PDAC, the accuracy tends to increase as the pruning factor decreases. When the pruning factor is set to 0, the features are ranked only using wrapper methods. Hence, we see higher accuracy on PDAC. Whereas when the pruning factor is set to

1, the features are ranked purely using filter methods. Hence, higher stability can be observed on both datasets.

In summary, this sensitivity analysis reveals that the pruning factor has a nuanced impact on the performance of the FWSE method. It implies that the optimal setting may depend on the specific characteristics of the dataset and the priorities of the study (e.g., maximizing accuracy vs. stability). Future work will include developing methods to automatically determine the most appropriate pruning factor for different types of datasets.

# Appendix B

# How Does Self-Attention Work?

## B.1 Mathematical Foundation of Self-Attention

Self-attention is a core component of transformer architectures. This appendix dives into the mathematical underpinnings of self-attention, as well as provides an intuitive example to explain why self-attention works.

### B.1.1 Definition and Process

Self-attention, a sequence-to-sequence operation, processes a sequence of input vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t$ to produce a corresponding sequence of output vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t$. Each vector, whether input or output, resides in a $k$-dimensional space.

The computation of an output vector $\mathbf{y}_i$ involves taking a weighted sum of all input vectors:

$$\mathbf{y}_i = \sum_j w_{ij} \mathbf{x}_j \tag{B.1}$$

Here, $w_{ij}$ represents the weight assigned to the input vector $\mathbf{x}_j$ when computing the output vector $\mathbf{y}_i$.

### B.1.2 Weight Calculation

The weights $w_{ij}$ are not predefined but are calculated dynamically through the self-attention mechanism. Initially, a set of unnormalized weights $w'_{ij}$ is computed using the dot product of the input vectors:

$$w'_{ij} = \mathbf{x}_i^T \mathbf{x}_j \tag{B.2}$$

This dot product measures the similarity between two vectors, serving as a basis for the weight assignment. The higher the dot product, the more one vector contributes to the other's context representation.

### B.1.3 Normalization

The raw weights $w'_{ij}$ are then normalized using the softmax function to ensure they sum to 1, enabling a probabilistic interpretation:

$$w_{ij} = \frac{\exp(w'_{ij})}{\sum_k \exp(w'_{kj})} \tag{B.3}$$

This normalization step is crucial for the model's ability to manage the influence of each input vector on the output.

## B.2 Understanding Why Self-Attention Works

Self-attention's effectiveness in transformer architectures is not immediately intuitive, despite its simplicity. To grasp why self-attention is so powerful, let us take an example.

### B.2.1 Meal Recommendation: An Analogy

Imagine operating a meal recommendation service where you aim to suggest dishes that align with individual user preferences. Each meal can be characterized by multiple features (e.g., spiciness, sweetness, calorie count, vegetarian) and users have distinct tastes and dietary requirements. To recommend meals, one can start by defining feature vectors for both meals and users. For a meal, the vector could include values representing its spiciness, vegetarian-friendly level, or richness in certain ingredients. Correspondingly, user vectors would reflect their preferences for these attributes.

The dot product between a meal's feature vector and a user's preference vector provides a relevance score, indicating how well the meal aligns with the user's tastes. A positive alignment (e.g., a dish with high spiciness levels recommended to someone who likes spicy things) contributes positively to the score, while a mismatch (e.g., a spicy dish recommended to someone who dislikes spice) detracts from it.

However, manually defining the features for the dishes and users can be a challenging task. Instead, we can allow the model to determine these features itself. We collect a small set of meals that users have indicated they like or dislike. The model can learn feature vectors for both meals and users such that the dot product between them aligns with the users' preferences. What usually tends to happen is, over time, even without explicit instructions on what each feature should represent, the model develops feature vectors that capture meaningful aspects of the meals. For example, one feature might end up corresponding to the spiciness of the meal, while another might relate to its vegetarian status, even though these labels were never explicitly provided.

## B.3 Conclusion

The self-attention mechanism's ability to dynamically weigh and integrate information across a sequence is foundational to the success of transformer models. This appendix has dissected the mathematical underpinnings of self-attention and illustrated its effectiveness through a meal recommendation analogy. The dynamic weighting (attention) allows models to focus on important information and adapt to the context, thereby enhancing predictive capabilities. The balance of global and local perspective is what drives the versatility and efficiency of transformer models, powered by self-attention, making them a key advancement in AI.

# Bibliography


Abraham, Alexandre et al. (2014). "Machine learning for neuroimaging with scikit-learn". In: *Frontiers in neuroinformatics* 8, p. 71792.

Achenbach, Thomas M and Leslie Rescorla (2007). *Multicultural understanding of child and adolescent psychopathology: Implications for mental health assessment*. Guilford Press.

Addington, Donald, Jean Addington, and Eleanor Maticka-Tyndale (1993). "Assessing depression in schizophrenia: the Calgary Depression Scale". In: *The British journal of psychiatry* 163.S22, pp. 39–44.

Afgan, Enis et al. (2018). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update". In: *Nucleic acids research* 46.W1, W537–W544.

Amrhein, Valentin, Sander Greenland, and Blake McShane (2019). "Scientists rise up against statistical significance". In: *Nature* 567.7748, pp. 305–307.

Anaissi, Ali et al. (2016). "Ensemble feature learning of genomic data using support vector machine". In: *PloS one* 11.6, e0157330.

Andrew, Galen et al. (2013). "Deep canonical correlation analysis". In: *International conference on machine learning*. PMLR, pp. 1247–1255.

Aragaki, Masato et al. (2011). "Characterization of a Cleavage Stimulation Factor, 3' pre-RNA, Subunit 2, 64 kDa (CSTF2) as a Therapeutic Target for Lung CancerC-STF2 Activation in Lung Cancer". In: *Clinical Cancer Research* 17.18, pp. 5889–5900.

Atrey, Pradeep K et al. (2010). "Multimodal fusion for multimedia analysis: a survey". In: *Multimedia systems* 16, pp. 345–379.

Bailey, Dale L et al. (2005). *Positron emission tomography*. Vol. 2. Springer.

Baltrusaitis, Tadas, Chaitanya Ahuja, and Louis-Philippe Morency (2019). "Multimodal machine learning: A survey and taxonomy". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.2, pp. 423–443.

Barzaman, Khadijeh et al. (2020). "Breast cancer: Biology, biomarkers, and treatments". In: *International immunopharmacology* 84, p. 106535.

Bauer, Eric and Ron Kohavi (1999). "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants". In: *Machine learning* 36.1, pp. 105–139.

Beck, Aaron T et al. (1993). "Beck anxiety inventory". In: *Journal of consulting and clinical psychology*.

Biesanz, Jeremy C et al. (2004). "The role of coding time in estimating and interpreting growth curve models." In: *Psychological methods* 9.1, p. 30.

Biewald, Lukas et al. (2020). "Experiment tracking with weights and biases". In.

Birchwood, Max et al. (1990). "The social functioning scale the development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients". In: *The British Journal of Psychiatry* 157.6, pp. 853–859.

Bolker, Benjamin M et al. (2009). "Generalized linear mixed models: a practical guide for ecology and evolution". In: *Trends in ecology & evolution* 24.3, pp. 127–135.

Box, George EP and David A Pierce (1970). "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models". In: *Journal of the American statistical Association* 65.332, pp. 1509–1526.

Box, George EP et al. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brahim, Afef Ben and Mohamed Limam (2013). "Robust ensemble feature selection for high dimensional data sets". In: *2013 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE, pp. 151–157.

Breiman, Leo (1996). "Bagging predictors". In: *Machine learning* 24.2, pp. 123–140.

— (1999). "Pasting small votes for classification in large databases and on-line". In: *Machine learning* 36.1, pp. 85–103.

— (2001). "Random forests". In: *Machine learning* 45.1, pp. 5–32.

Budhraja, Sugam et al. (2020). "Sleep stage classification using neucube on spinnaker: a preliminary study". In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8.

Budhraja, Sugam et al. (2023a). "Filter and wrapper stacking ensemble (FWSE): a robust approach for reliable biomarker discovery in high-dimensional omics data". In: *Briefings in Bioinformatics* 24.6, bbad382.

Budhraja, Sugam et al. (2023b). "Mosaic LSM: A liquid state machine approach for multimodal longitudinal data analysis". In: *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8.

— (2024). "NeuroGeMS: An open-source GUI software for multimodal modelling in biomedical research and applications". In: *2024 International Conference on Neural Information Processing (ICONIP)*. APNNS.

Butterfield, D Allan, Debra Boyd-Kimball, and Alessandra Castegna (2003). "Proteomics in Alzheimer's disease: insights into potential mechanisms of neurodegeneration". In: *Journal of neurochemistry* 86.6, pp. 1313–1327.

Bzdok, Danilo and Andreas Meyer-Lindenberg (2018). "Machine learning for precision psychiatry: opportunities and challenges". In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3.3, pp. 223–230.

Cao, Liwei et al. (2021). "Proteogenomic characterization of pancreatic ductal adenocarcinoma". In: *Cell* 184.19, pp. 5031–5052.

Caruana, Rich (1997). "Multitask learning". In: *Machine learning* 28, pp. 41–75.

Chen, Kai et al. (2022). "Single cell RNA-seq reveals the CCL5/SDC1 receptor-ligand interaction between T cells and tumor cells in pancreatic cancer". In: *Cancer Letters* 545, p. 215834.

Chen, Tianqi and Carlos Guestrin (2016). "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Cho, Han-Jun et al. (2018). "Association of specific gene mutations derived from machine learning with survival in lung adenocarcinoma". In: *PLoS One* 13.11, e0207204.

Clarke, Christopher J and John N Haselden (2008). "Metabolic profiling as a tool for understanding mechanisms of toxicity". In: *Toxicologic Pathology* 36.1, pp. 140–147.

Cnaan, Avital, Nan M Laird, and Peter Slasor (1997). "Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data". In: *Statistics in medicine* 16.20, pp. 2349–2380.

Cohen, Alex S et al. (2020). "Digital phenotyping using multimodal data". In: *Current behavioral neuroscience reports* 7, pp. 212–220.

Cornblatt, Barbara A et al. (1988). "The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families". In: *Psychiatry research* 26.2, pp. 223–238.

Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine learning* 20.3, pp. 273–297.

Cover, Thomas and Peter Hart (1967). "Nearest neighbor pattern classification". In: *IEEE transactions on information theory* 13.1, pp. 21–27.

Cox, David R (1958). "The regression analysis of binary sequences". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 20.2, pp. 215–232.

— (1972). "Regression models and life-tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2, pp. 187–202.

Cummins, Nicholas et al. (2015). "A review of depression and suicide risk assessment using speech analysis". In: *Speech communication* 71, pp. 10–49.

Czeisler, Mark É (2020). "Mental health, substance use, and suicidal ideation during the COVID-19 pandemic—United States, June 24–30, 2020". In: *MMWR. Morbidity and mortality weekly report* 69.

Dalman, Mark R et al. (2012). "Fold change and p-value cutoffs significantly alter microarray interpretations". In: *BMC bioinformatics* 13.2, pp. 1–4.

De Hoffmann, Edmond and Vincent Stroobant (2007). *Mass spectrometry: principles and applications*. John Wiley & Sons.

De Pablo, Gonzalo Salazar et al. (2021). "Probability of transition to psychosis in individuals at clinical high risk: an updated meta-analysis". In: *JAMA psychiatry* 78.9, pp. 970–978.

Dix, A et al. (2016). "Use of systems biology to decipher host–pathogen interaction networks and predict biomarkers". In: *Clinical Microbiology and Infection* 22.7, pp. 600–606.

Doborjeh, Maryam et al. (2019). "Personalised modelling with spiking neural networks integrating temporal and static information". In: *Neural Networks* 119, pp. 162–177.

Doborjeh, Maryam et al. (2021a). "Deep learning of explainable EEG patterns as dynamic spatiotemporal clusters and rules in a brain-inspired spiking neural network". In: *Sensors* 21.14, p. 4900.

Doborjeh, Maryam et al. (2021b). "Personalised predictive modelling with brain-inspired spiking neural networks of longitudinal MRI neuroimaging data and the case study of dementia". In: *Neural Networks* 144, pp. 522–539.

Doborjeh, Maryam et al. (2022). "Personalized spiking neural network models of clinical and environmental factors to predict stroke". In: *Cognitive Computation* 14.6, pp. 2187–2202.

Doborjeh, Zohreh et al. (2020). "Interpretability of spatiotemporal dynamics of the brain processes followed by mindfulness intervention in a brain-inspired spiking neural network architecture". In: *Sensors* 20.24, p. 7354.

Doborjeh, Zohreh et al. (2023). "Investigation of social and cognitive predictors in non-transition ultra-high-risk'individuals for psychosis using spiking neural networks". In: *Schizophrenia* 9.1, p. 10.

Doborjeh, Zohreh et al. (2024). "A generalisability theory approach to quantifying changes in psychopathology among ultra-high-risk individuals for psychosis". In: *Schizophrenia* 10.1, p. 87.

Doniger, Glen M et al. (2000). "Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes". In: *Journal of cognitive neuroscience* 12.4, pp. 615–621.

Doniger, Glen M et al. (2001). "Impaired sensory processing as a basis for object-recognition deficits in schizophrenia". In: *American Journal of Psychiatry* 158.11, pp. 1818–1826.

Doshi-Velez, Finale and Been Kim (2017). "Towards a rigorous science of interpretable machine learning". In: *arXiv preprint arXiv:1702.08608*.

Drotár, Peter, Matej Gazda, and Liberios Vokorokos (2019). "Ensemble feature selection using election methods and ranker clustering". In: *Information Sciences* 480, pp. 365–380.

Drumm, Mitchell L et al. (2005). "Genetic modifiers of lung disease in cystic fibrosis". In: *New England Journal of Medicine* 353.14, pp. 1443–1453.

Du, Jing et al. (2016). "A new feature evaluation algorithm and its application to fault of high-speed railway". In: *International Conference on Intelligent Transportation*. Springer, pp. 1–14.

Duffy, Michael J (2001). "Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful?" In: *Clinical chemistry* 47.4, pp. 624–630.

Dwork, Cynthia et al. (2001). "Rank aggregation methods for the web". In: *Proceedings of the 10th international conference on World Wide Web*, pp. 613–622.

Efron, Bradley and Robert J Tibshirani (1994). *An introduction to the bootstrap*. CRC press.

Elman, Jeffrey L (1991). "Distributed representations, simple recurrent networks, and grammatical structure". In: *Machine learning* 7, pp. 195–225.

Engler, Henry et al. (2006). "Two-year follow-up of amyloid deposition in patients with Alzheimer's disease". In: *Brain* 129.11, pp. 2856–2866.

Ennajdaoui, Hanane et al. (2016). "IGF2BP3 modulates the interaction of invasion-associated transcripts with RISC". In: *Cell reports* 15.9, pp. 1876–1883.

Finn, Chelsea, Pieter Abbeel, and Sergey Levine (2017). "Model-agnostic meta-learning for fast adaptation of deep networks". In: *International conference on machine learning*. PMLR, pp. 1126–1135.

First, Michael B and Miriam Gibbon (2004). "The structured clinical interview for DSM-IV axis I disorders (SCID-I) and the structured clinical interview for DSM-IV axis II disorders (SCID-II)." In: *Handbook of Psychological Assessment* 2, pp. 134–143.

Fix, Evelyn and Joseph Lawson Hodges (1989). "Discriminatory analysis. Nonparametric discrimination: Consistency properties". In: *International Statistical Review/Revue Internationale de Statistique* 57.3, pp. 238–247.

Frank, Richard and Richard Hargreaves (2003). "Clinical biomarkers in drug discovery and development". In: *Nature reviews Drug discovery* 2.7, pp. 566–580.

Friedman, Jerome H (2002). "Stochastic gradient boosting". In: *Computational statistics & data analysis* 38.4, pp. 367–378.

Fusar-Poli, Paolo et al. (2012). "Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk". In: *Archives of general psychiatry* 69.3, pp. 220–229.

Fusar-Poli, Paolo et al. (2020). "Real-world long-term outcomes in individuals at clinical risk for psychosis: the case for extending duration of care". In: *EClinicalMedicine* 28.

Galle, Peter R et al. (2019). "Biology and significance of alpha-fetoprotein in hepatocellular carcinoma". In: *Liver international* 39.12, pp. 2214–2229.

Gammerman, Alex, Volodya Vovk, and Vladimir Vapnik (2013). "Learning by transduction". In: *arXiv preprint arXiv:1301.7375*.

Gentleman, Robert C et al. (2004). "Bioconductor: open software development for computational biology and bioinformatics". In: *Genome biology* 5, pp. 1–16.

Ghemrawi, Rose and Mostafa Khair (2020). "Endoplasmic reticulum stress and unfolded protein response in neurodegenerative diseases". In: *International journal of molecular sciences* 21.17, p. 6127.

Gibson, Clare M et al. (2010). "Social skill and social cognition in adolescents at genetic risk for psychosis". In: *Schizophrenia research* 122.1-3, pp. 179–184.

Goh, Wilson Wen Bin and Limsoon Wong (2016). "Evaluating feature-selection stability in next-generation proteomics". In: *Journal of bioinformatics and computational biology* 14.05, p. 1650029.

— (2019). "Advanced bioinformatics methods for practical applications in proteomics". In: *Briefings in bioinformatics* 20.1, pp. 347–355.

Goh, Wilson Wen Bin et al. (2017). "Can peripheral blood-derived gene expressions characterize individuals at ultra-high risk for psychosis?" In: *Computational Psychiatry*, pp. 168–183.

Goodwin, Sara, John D McPherson, and W Richard McCombie (2016). "Coming of age: ten years of next-generation sequencing technologies". In: *Nature Reviews Genetics* 17.6, pp. 333–351.

Grande, Iria et al. (2016). "Bipolar disorder". In: *The Lancet* 387.10027, pp. 1561–1572.

Green, Jennifer Greif et al. (2010). "Childhood adversities and adult psychiatric disorders in the national comorbidity survey replication I: associations with first onset of DSM-IV disorders". In: *Archives of general psychiatry* 67.2, pp. 113–123.

Guyon, Isabelle et al. (2002). "Gene selection for cancer classification using support vector machines". In: *Machine learning* 46.1, pp. 389–422.

Hall, Mark et al. (2009). "The WEKA data mining software: an update". In: *ACM SIGKDD explorations newsletter* 11.1, pp. 10–18.

Halsey, Lewis G et al. (2015). "The fickle P value generates irreproducible results". In: *Nature methods* 12.3, pp. 179–185.

Hanke, Michael et al. (2009). "PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data". In: *Neuroinformatics* 7, pp. 37–53.

Hartmann, Jessica A et al. (2016). "Declining transition rates to psychotic disorder in "ultra-high risk" clients: investigation of a dilution effect". In: *Schizophrenia research* 170.1, pp. 130–136.

Hasin, Yehudit, Marcus Seldin, and Aldons Lusis (2017). "Multi-omics approaches to disease". In: *Genome biology* 18.1, pp. 1–15.

Hatz, Maximilian HM, Katharina Schremser, and Wolf H Rogowski (2014). "Is individualized medicine more cost-effective? A systematic review". In: *Pharmacoeconomics* 32, pp. 443–455.

He, Zengyou and Weichuan Yu (2010). "Stable feature selection for biomarker discovery". In: *Computational biology and chemistry* 34.4, pp. 215–225.

Ho, Tin Kam (1998). "The random subspace method for constructing decision forests". In: *IEEE transactions on pattern analysis and machine intelligence* 20.8, pp. 832–844.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Holzinger, Andreas et al. (2017). "What do we need to build explainable AI systems for the medical domain?" In: *arXiv preprint arXiv:1712.09923*.

Hong, Chao-Qun et al. (2015). "Elevated C1orf63 expression is correlated with CDK10 and predicts better outcome for advanced breast cancers: a retrospective study". In: *BMC cancer* 15.1, pp. 1–12.

Horgan, Richard P and Louise C Kenny (2011). "'Omic'technologies: genomics, transcriptomics, proteomics and metabolomics". In: *The Obstetrician & Gynaecologist* 13.3, pp. 189–195.

Hu, Hai et al. (2019). "The C/EBP homologous protein (CHOP) transcription factor functions in endoplasmic reticulum stress-induced apoptosis and microbial infection". In: *Frontiers in immunology* 9, p. 3083.

Huang, Shih-Cheng et al. (2020). "Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection". In: *Scientific reports* 10.1, p. 22147.

Huang, Xinyang et al. (2022). "Up-regulated misp is associated with poor prognosis and immune infiltration in pancreatic ductal adenocarcinoma". In: *Frontiers in Oncology* 12, p. 827051.

IJzendoorn, David GP van et al. (2019). "Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas". In: *PLoS computational biology* 15.2, e1006826.

Insel, Thomas R (2014). "The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry". In: *American journal of psychiatry* 171.4, pp. 395–397.

Ioannidis, John PA et al. (2005). "Microarrays and molecular research: noise discovery?" In: *Lancet* 365.9458, pp. 454–454.

Jaeger, Herbert (2001). "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note". In: *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* 148.34, p. 13.

Jain, Kewal K and Kewal K Jain (2010). *The handbook of biomarkers*. Springer.

Jansson, Inger and A Birgitta Gunnarsson (2018). "Employers' views of the impact of mental health problems on the ability to work". In: *Work* 59.4, pp. 585–598.

Joachims, Thorsten (2003). "Transductive learning via spectral graph partitioning". In: *Proceedings of the 20th international conference on machine learning (ICML-03)*, pp. 290–297.

Joachims, Thorsten et al. (1999). "Transductive inference for text classification using support vector machines". In: *Icml*. Vol. 99, pp. 200–209.

Kasabov, Nikola (2007a). "Global, local and personalised modeling and pattern discovery in bioinformatics: An integrated approach". In: *Pattern Recognition Letters* 28.6, pp. 673–685.

Kasabov, Nikola et al. (2013). "Dynamic evolving spiking neural networks for online spatio-and spectro-temporal pattern recognition". In: *Neural Networks* 41, pp. 188–201.

Kasabov, Nikola K (2007b). *Evolving connectionist systems: the knowledge engineering approach*. Springer Science & Business Media.

— (2014). "NeuCube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data". In: *Neural Networks* 52, pp. 62–76.

Kasabov, Nikola K et al. (2023). "Transfer learning of fuzzy spatio-temporal rules in a brain-inspired spiking neural network architecture: a case study on spatio-temporal Brain Data". In: *IEEE Transactions on Fuzzy Systems* 31.12, pp. 4542–4552.

Kay, Stanley R, Abraham Fiszbein, and Lewis A Opler (1987). "The positive and negative syndrome scale (PANSS) for schizophrenia". In: *Schizophrenia bulletin* 13.2, pp. 261–276.

Keefe, Richard SE et al. (1999). "Brief assessment of cognition in schizophrenia". In: *Schizophrenia Research*.

Kessler, Ronald C et al. (2005). "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication". In: *Archives of general psychiatry* 62.6, pp. 593–602.

Khaire, Utkarsh Mahadeo and R Dhanalakshmi (2019). "Stability of feature selection algorithm: A review". In: *Journal of King Saud University-Computer and Information Sciences*.

Kidder, Benjamin L et al. (2017). "SMYD5 Controls Heterochromatin and Chromosome Integrity during Embryonic Stem Cell DifferentiationSMYD5 Regulates Genome Stability". In: *Cancer research* 77.23, pp. 6729–6745.

Kimhy, D et al. (2007). "Visual form perception: a comparison of individuals at high risk for psychosis, recent onset schizophrenia and chronic schizophrenia". In: *Schizophrenia Research* 97.1-3, pp. 25–34.

Kiros, Ryan, Ruslan Salakhutdinov, and Richard S Zemel (2014). "Unifying visual-semantic embeddings with multimodal neural language models". In: *arXiv preprint arXiv:1411.2539*.

Klein, John P, Melvin L Moeschberger, et al. (2003). *Survival analysis: techniques for censored and truncated data*. Vol. 1230. Springer.

Kolch, Walter and Andrew Pitt (2010). "Functional proteomics to dissect tyrosine kinase signalling pathways in cancer". In: *Nature Reviews Cancer* 10.9, pp. 618–629.

Koprinkova-Hristova, Petia et al. (2023). "On-line learning, classification and interpretation of brain signals using 3D SNN and ESN". In: *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–6.

Koutsouleris, Nikolaos et al. (2021). "Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression". In: *JAMA psychiatry* 78.2, pp. 195–209.

Krebs, Catharine E et al. (2020). "Whole blood transcriptome analysis in bipolar disorder reveals strong lithium effect". In: *Psychological medicine* 50.15, pp. 2575–2586.

Kupfer, David J, Michael B First, and Darrel A Regier (2008). "A research agenda for DSM V". In.

La Thangue, Nicholas B and David J Kerr (2011). "Predictive biomarkers: a paradigm shift towards personalized cancer medicine". In: *Nature reviews Clinical oncology* 8.10, pp. 587–596.

Lahat, Dana, Tülay Adali, and Christian Jutten (2015). "Multimodal data fusion: an overview of methods, challenges, and prospects". In: *Proceedings of the IEEE* 103.9, pp. 1449–1477.

Lan, Zhen-zhong et al. (2014). "Multimedia classification and event detection using double fusion". In: *Multimedia tools and applications* 71, pp. 333–347.

Lee, Jason E et al. (2020). "Endoplasmic reticulum contact sites regulate the dynamics of membraneless organelles". In: *Science* 367.6477, eaay7108.

Lee, Jimmy et al. (2013). "The longitudinal youth at risk study (LYRIKS)—an Asian UHR perspective". In: *Schizophrenia research* 151.1-3, pp. 279–283.

Leuchter, Andrew F et al. (2010). "Biomarkers to predict antidepressant response". In: *Current psychiatry reports* 12, pp. 553–562.

Liang, Jiaqi et al. (2023). "VSOLassoBag: a variable-selection oriented LASSO bagging algorithm for biomarker discovery in omic-based translational research". In: *Journal of Genetics and Genomics*.

Liu, Hongcheng et al. (2016). "Up-regulation of SRPK1 in non-small cell lung cancer promotes the growth and migration of cancer cells". In: *Tumor Biology* 37, pp. 7287–7293.

Liu, Jianfang et al. (2018a). "An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics". In: *Cell* 173.2, pp. 400–416.

Liu, Yuqing et al. (2018b). "Unconventional myosin VIIA promotes melanoma progression". In: *Journal of Cell Science* 131.4, jcs209924.

Liu, Zhuohang, Hang Li, and Shuyi Pan (2021). "Discovery and validation of key biomarkers based on immune infiltrates in Alzheimer's disease". In: *Frontiers in genetics* 12, p. 658323.

Loscalzo, Steven, Lei Yu, and Chris Ding (2009). "Consensus group stable feature selection". In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 567–576.

Louppe, Gilles and Pierre Geurts (2012). "Ensembles on random patches". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 346–361.

Lundberg, Scott M and Su-In Lee (2017). "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30.

Lyons-Weiler, James, Satish Patel, and Soumyaroop Bhattacharya (2003). "A classification-based machine learning approach for the analysis of genome-wide expression data". In: *Genome research* 13.3, pp. 503–512.

Maass, Wolfgang, Thomas Natschläger, and Henry Markram (2002). "Real-time computing without stable states: A new framework for neural computation based on perturbations". In: *Neural computation* 14.11, pp. 2531–2560.

Mardin, Wolf Arif et al. (2016). "SERPINB5 promoter hypomethylation differentiates pancreatic ductal adenocarcinoma from pancreatitis". In: *Pancreas* 45.5, pp. 743–747.

Marinescu, Razvan V et al. (2018). "TADPOLE challenge: prediction of longitudinal evolution in Alzheimer's disease". In: *arXiv preprint arXiv:1805.03909*.

Marquand, Andre F et al. (2016). "Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders". In: *Biological psychiatry: cognitive neuroscience and neuroimaging* 1.5, pp. 433–447.

McDermott, Jason E et al. (2013). "Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data". In: *Expert opinion on medical diagnostics* 7.1, pp. 37–51.

McGrath, John J et al. (2023). "Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries". In: *The Lancet Psychiatry* 10.9, pp. 668–681.

Miotto, Riccardo et al. (2018). "Deep learning for healthcare: review, opportunities and challenges". In: *Briefings in bioinformatics* 19.6, pp. 1236–1246.

Monkaresi, Hamed, M Sazzad Hussain, and Rafael A Calvo (2012). "Classification of affects using head movement, skin color features and physiological signals". In: *2012 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, pp. 2664–2669.

Mueller, Susanne G et al. (2005). "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI)". In: *Alzheimer's & Dementia* 1.1, pp. 55–66.

Nasreddine, Ziad S et al. (2005). "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment". In: *Journal of the American Geriatrics Society* 53.4, pp. 695–699.

Nemeroff, Charles B et al. (2013). "DSM-5: a collection of psychiatrist views on the changes, controversies, and future directions". In: *BMC medicine* 11.1, p. 202.

Ngiam, Jiquan et al. (2011). "Multimodal deep learning". In: *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 689–696.

Öhman, Arne, Anders Flykt, and Francisco Esteves (2001). "Emotion drives attention: detecting the snake in the grass." In: *Journal of experimental psychology: general* 130.3, p. 466.

Organization, World Health (2022). *World mental health report: Transforming mental health for all*. World Health Organization.

Organization, World Health et al. (2017). *Depression and other common mental disorders: global health estimates*. Tech. rep. World Health Organization.

O'Halloran, Kay L et al. (2012). "Interactive software for multimodal analysis". In: *Visual Communication* 11.3, pp. 363–381.

Park, Chihyun, Jihwan Ha, and Sanghyun Park (2020). "Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset". In: *Expert Systems with Applications* 140, p. 112873.

Park, Kyoung Sik et al. (2021). "Highly accurate diagnosis of papillary thyroid carcinomas based on personalized pathways coupled with machine learning". In: *Briefings in Bioinformatics* 22.4, bbaa336.

Paszke, Adam et al. (2019). "Pytorch: An imperative style, high-performance deep learning library". In: *Advances in neural information processing systems* 32.

Pearl, Judea et al. (2000). "Models, reasoning and inference". In: *Cambridge, UK: CambridgeUniversityPress* 19.2, p. 3.

Petrucelli, Nancie, Mary B Daly, and Gerald L Feldman (2010). "Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2". In: *Genetics in Medicine* 12.5, pp. 245–259.

Piccart-Gebhart, Martine J et al. (2005). "Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer". In: *New England Journal of Medicine* 353.16, pp. 1659–1672.

Preacher, Kristopher J (2008). *Latent growth curve modeling*. 157. Sage.

Quante, Anne S et al. (2016). "Projections of cancer incidence and cancer-related deaths in Germany by 2020 and 2030". In: *Cancer medicine* 5.9, pp. 2649–2656.

Rabinowitz, JD et al. (2011). "Metabolomics in drug target discovery". In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 76. Cold Spring Harbor Laboratory Press, pp. 235–246.

Racine, Nicole et al. (2021). "Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis". In: *JAMA pediatrics* 175.11, pp. 1142–1150.

Radhakrishnan, Rajiv, Muzaffer Kaser, and Sinan Guloksuz (2017). "The link between the immune system, environment, and psychosis". In: *Schizophrenia bulletin* 43.4, pp. 693–697.

Rasmussen, Carl Edward (2003). "Gaussian processes in machine learning". In: *Summer school on machine learning*. Springer, pp. 63–71.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016). """ Why should i trust you?" Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144.

Roberts, Lee D, Albert Koulman, and Julian L Griffin (2014). "Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome". In: *The lancet Diabetes & endocrinology* 2.1, pp. 65–75.

Robin, Fabien et al. (2020). "Molecular profiling of stroma highlights stratifin as a novel biomarker of poor prognosis in pancreatic ductal adenocarcinoma". In: *British Journal of Cancer* 123.1, pp. 72–80.

Rohlfing, Curt L et al. (2002). "Defining the relationship between plasma glucose and HbA1c: analysis of glucose profiles and HbA1c in the Diabetes Control and Complications Trial". In: *Diabetes care* 25.2, pp. 275–278.

Rosas, Verónica Pérez, Rada Mihalcea, and Louis-Philippe Morency (2013). "Multimodal sentiment analysis of spanish online videos". In: *IEEE intelligent Systems* 28.3, pp. 38–45.

Rosenblat, Joshua D and Roger S McIntyre (2017). "Bipolar disorder and immune dysfunction: epidemiological findings, proposed pathophysiology and clinical implications". In: *Brain sciences* 7.11, p. 144.

Rosenblatt, Frank (1961). *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*. Tech. rep. Cornell Aeronautical Lab Inc Buffalo NY.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1985). *Learning internal representations by error propagation*. Tech. rep. California Univ San Diego La Jolla Inst for Cognitive Science.

Sawada, Kingo et al. (2017). "Identifying neurocognitive markers for outcome prediction of global functioning in individuals with first-episode and ultra-high-risk for psychosis". In: *Psychiatry and Clinical Neurosciences* 71.5, pp. 318–327.

Seijo-Pardo, Borja et al. (2017). "Ensemble feature selection: homogeneous and heterogeneous approaches". In: *Knowledge-Based Systems* 118, pp. 124–139.

Shi, Yi-Hua et al. (2020). "Predicting miRNA targets for hepatocellular carcinoma with an integrated method". In: *Translational Cancer Research* 9.3, p. 1752.

Singh, Balkaran et al. (2023). "Constrained neuro fuzzy inference methodology for explainable personalised modelling with applications on gene expression data". In: *Scientific Reports* 13.1, p. 456.

Song, Qun and Nikola Kasabov (2006). "TWNFI—a transductive neuro-fuzzy inference system with weighted data normalization for personalized modeling". In: *Neural Networks* 19.10, pp. 1591–1596.

Strimbu, Kyle and Jorge A Tavel (2010). "What are biomarkers?" In: *Current Opinion in HIV and AIDS* 5.6, p. 463.

Sui, Jing et al. (2012). "A review of multivariate methods for multimodal fusion of brain imaging data". In: *Journal of neuroscience methods* 204.1, pp. 68–81.

Sun, Yan V and Yi-Juan Hu (2016). "Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases". In: *Advances in genetics* 93, pp. 147–190.

Sung, Hyuna et al. (2021). "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: *CA: a cancer journal for clinicians* 71.3, pp. 209–249.

Suzuki, Miho M and Adrian Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics". In: *Nature reviews genetics* 9.6, pp. 465–476.

Tan, Ming, Ove B Schaffalitzky de Muckadell, and Maiken Thyregod Joergensen (2020). "Gene expression network analysis of precursor lesions in familial pancreatic cancer". In: *Journal of Pancreatic Cancer* 6.1, pp. 73–84.

Tan, Samuel Ming Xuan et al. (2023). "RNA-sequencing of peripheral whole blood of individuals at ultra-high-risk for psychosis–A longitudinal perspective". In: *Asian Journal of Psychiatry* 89, p. 103796.

Tang, Deping et al. (2018). "Strand-specific RNA-seq analysis of the Acidithiobacillus ferrooxidans transcriptome in response to magnesium stress". In: *Archives of microbiology* 200, pp. 1025–1035.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288.

Tjoa, Ernesta and Cuntai Guan (2020). "A survey on explainable artificial intelligence (XAI): Towards medical XAI". In: *IEEE Transactions on Neural Networks and Learning Systems* 32.11, pp. 4793–4813.

Tombaugh, Tom N and Nancy J McIntyre (1992). "The mini-mental state examination: a comprehensive review". In: *Journal of the American Geriatrics Society* 40.9, pp. 922–935.

Toth, Reka et al. (2019). "Random forest-based modelling to detect biomarkers for prostate cancer progression". In: *Clinical epigenetics* 11, pp. 1–15.

Trautmann, Sebastian, Jürgen Rehm, and Hans-Ulrich Wittchen (2016). "The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders?" In: *EMBO reports* 17.9, pp. 1245–1249.

Tremblay, Johanne and Pavel Hamet (2013). "Role of genomics on the path to personalized medicine". In: *Metabolism* 62, S2–S5.

Vapnik, Vladimir Naumovich, Vlamimir Vapnik, et al. (1998). "Statistical learning theory". In.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30.

Vlaardingerbroek, Marinus T and Jacques A Boer (2013). *Magnetic resonance imaging: theory and practice*. Springer Science & Business Media.

Wang, Huazhen, Fan Yang, and Zhiyuan Luo (2016). "An experimental study of the intrinsic stability of random forest variable importance measures". In: *BMC bioinformatics* 17.1, pp. 1–18.

Weng, Stephen F et al. (2017). "Can machine-learning improve cardiovascular risk prediction using routine clinical data?" In: *PloS one* 12.4, e0174944.

Wenk, Markus R (2005). "The emerging field of lipidomics". In: *Nature reviews Drug discovery* 4.7, pp. 594–610.

Wolpert, David H (1992). "Stacked generalization". In: *Neural networks* 5.2, pp. 241–259.

Wu, Xun et al. (2022). "Genetic analysis of potential biomarkers and therapeutic targets in ferroptosis from coronary artery disease". In: *Journal of Cellular and Molecular Medicine* 26.8, pp. 2177–2190.

Wu, Zhiyong, Lianhong Cai, and Helen Meng (2005). "Multi-level fusion of audio and visual features for speaker identification". In: *Advances in Biometrics: International Conference, ICB 2006, Hong Kong, China, January 5-7, 2006. Proceedings*. Springer, pp. 493–499.

Xu, Kelvin et al. (2015). "Show, attend and tell: Neural image caption generation with visual attention". In: *International conference on machine learning*. PMLR, pp. 2048–2057.

Ye, Zhen et al. (2021). "Comprehensive analysis of alteration landscape and its clinical significance of mitochondrial energy metabolism pathway-related genes in lung cancers". In: *Oxidative Medicine and Cellular Longevity* 2021.

Yung, A et al. (2002). "Comprehensive assessment of at-risk mental states (CAARMS)". In: *Melbourne, Australia, University of Melbourne, Department of Psychiatry, Personal Assessment and Crisis Evaluation Clinic*.

Yung, Alison R and Barnaby Nelson (2013). "The ultra-high risk concept—a review". In: *The Canadian Journal of Psychiatry* 58.1, pp. 5–12.

Yung, Alison R et al. (2003). "Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group". In: *Schizophrenia research* 60.1, pp. 21–32.

Yung, Alison R et al. (2005). "Mapping the onset of psychosis: the comprehensive assessment of at-risk mental states". In: *Australian & New Zealand Journal of Psychiatry* 39.11-12, pp. 964–971.

Zadeh, Amir et al. (2016). "Mosi: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos". In: *arXiv preprint arXiv:1606.06259*.

Zaharia, Matei et al. (2018). "Accelerating the machine learning lifecycle with MLflow." In: *IEEE Data Eng. Bull.* 41.4, pp. 39–45.

Zeng, Yuan et al. (2018). "SERINC2-knockdown inhibits proliferation, migration and invasion in lung adenocarcinoma". In: *Oncology Letters* 16.5, pp. 5916–5922.

Zhang, Fan et al. (2019). "Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease". In: *Neurocomputing* 361, pp. 185–195.

Zhang, Huiqin et al. (2021). "Interaction between A$\beta$ and tau in the pathogenesis of Alzheimer's disease". In: *International journal of biological sciences* 17.9, p. 2181.

Zhang, Xiaokang, Inge Jonassen, and Anders Goksøyr (2021). "Machine learning approaches for biomarker discovery using gene expression data". In: *Exon Publications*, pp. 53–64.

Zhao, Yanbin et al. (2013). "ABCC3 as a marker for multidrug resistance in non-small cell lung cancer". In: *Scientific reports* 3.1, pp. 1–6.

Zhu, Hongwei et al. (2021). "S100A14 promotes progression and gemcitabine resistance in pancreatic cancer". In: *Pancreatology* 21.3, pp. 589–598.

Zhu, Xiaojin Jerry (2005). "Semi-supervised learning literature survey". In.