Mosaic LSM: A Liquid State Machine Approach for Multimodal Longitudinal Data Analysis

Sugam Budhraja
KEDRI
Auckland University of Technology
Auckland, New Zealand
ORCID: 0000-0002-3647-8722

Balkaran Singh
KEDRI
Auckland University of Technology
Auckland, New Zealand
balkaran.singh@aut.ac.nz

Maryam Doborjeh
KEDRI
Auckland University of Technology
Auckland, New Zealand
mgholami@aut.ac.nz

Zohreh Doborjeh
School of Population Health and Center for Brain Research
University of Auckland
Auckland, New Zealand
zohreh.doborjeh@auckland.ac.nz

Samuel Tan
School of Biological Sciences
Nanyang Technological University
Singapore
MINGXUAN001@e.ntu.edu.sg

Edmund Lai
KEDRI
Auckland University of Technology
Auckland, New Zealand
edmund.lai@aut.ac.nz

Wilson Goh
LKC School of Medicine
Nanyang Technological University
Singapore
wilsongoh@ntu.edu.sg

Nikola Kasabov*
KEDRI
Auckland University of Technology
Auckland, New Zealand
nkasabov@aut.ac.nz

***Abstract*— In this paper, we present a novel Liquid State Machine (LSM) based approach for modelling of multimodal longitudinal data: the Mosaic LSM. Our model harnesses the strengths of multiple LSMs, each designed to capture the temporal patterns of a specific data modality. This temporal information is then added to the raw data to create a composite representation that encompasses both the multimodal and the longitudinal aspects of the data. We demonstrate the performance of our approach on a real-world dataset that contains clinical, cognitive, and genetic modalities with the aim of predicting the Ultra-High Risk (UHR) status in individuals, six months in advance. Our results show that the Mosaic LSM outperforms traditional machine learning models, achieving an outstanding Matthew's Correlation Coefficient of 0.84 and prediction accuracy of 92.4%. Overall, our work highlights the potential of Mosaic LSM as a powerful tool for disease prognosis, and its ability to leverage both the multimodality and temporality of the data to improve performance.**

Keywords—Spiking Neural Networks, Liquid State Machine, Multimodal Learning, Time-Series, Ultra-High Risk, Prognosis

I. INTRODUCTION

In the biomedical domain, longitudinal multimodal datasets are collections of data collected over time, across multiple modalities such as imaging, omics, clinical, etc. These datasets are used to study the progression of diseases and conditions, and the effects of interventions, treatments, and lifestyle changes on health outcomes [1]. With multiple modalities, it is possible to gain information about the individual parts of a system and its emergent behaviour as a whole for inference or modelling related problems. Integration of multiple modalities provides a more comprehensive view of biological processes and enables the analysis of complex relationships between biological and behavioural factors. This type of data can be useful for more accurate diagnosis, prognosis and the development of personalized medicine and treatment planning [2].

However, longitudinal multimodal datasets pose new difficulties as they require modelling complex temporal relationships between various modalities, which are not addressed by standard machine learning algorithms. The naïve approach to handling this kind of data is to simply concatenate the different modalities and take the average of the time-based effects. However, this approach leads to the suppression of the interaction that takes place within each modality, resulting in the loss of the temporal dependencies that are inherent to the data. To fully leverage the potential of longitudinal multimodal datasets, both modality fusion and the treatment of longitudinal data must be considered simultaneously.

The objective of fusion techniques is to effectively utilize complementary information from multiple modalities. There are two main methods of fusion, early and late fusion [3]. Early fusion simply concatenates features from different modalities into a single input for a machine learning model, allowing the model to learn relationships between and within modalities, but at a lower level of abstraction. Late fusion, on the other hand, trains separate models for each modality and combines their predictions using combination methods such as simple averaging or weighted averaging. Early fusion enables learning joint representations, but relevant features of a modality may only be recognized at higher levels of abstraction. Late fusion excels in marginal representations but does not take into account cross-modality interactions [4]. For the treatment of the longitudinal aspect, the challenge is how to find marginal representation for dealing with within modality temporal interactions and joint representations of the heterogeneous modalities.

This paper proposes a new reservoir spiking neural network architecture to tackle the aforementioned challenges. It integrates data from multiple modalities while accounting for their temporal dynamics through intermediate fusion. Reservoir spiking neural networks (SNNs) are well-suited for handling longitudinal data due to their ability to incorporate time-series information in the dynamic state of the reservoir [5]. The reservoir acts as a memory storage system that is able to preserve temporal dependencies, allowing it to process sequential data effectively. Our architecture fuses the dynamic states of different reservoirs representing different modalities to achieve longitudinal multimodal data integration. Based on the task at hand, this architecture can be used to perform early, intermediate, or late fusion of the modalities. This method presents a significant advantage over other artificial neural network-based architectures as it only trains the last layer which combines and performs a simultaneous readout of the different reservoirs representing each modality, resulting in faster training compared to other recurrent neural networks.

Research funded by Ministry of Business, Innovation and Employment, New Zealand, and Singapore Data Science Consortium

As a case study on potential biomedical applications of this method, we show how information from clinical, genetic (RNA sequencing), and cognitive modalities can be combined for prognosis of ultra-high-risk (UHR) status in patients. Previous research on UHR modelling has investigated techniques for single modality data analysis and multimodal data modelling. One study proposed a neuro-fuzzy method for discovering personalized predictive rules using vector-based gene expression data [6], while another introduced a method for early integration of cognitive and social longitudinal data based on the NeuCube SNN framework, which extracts spatio-temporal associations between variables from different modalities [7]. Building on these studies, this paper presents a multimodal LSM architecture that supports early, intermediate, and late integration approaches. These approaches leverage the strengths of each modality, overcoming the limitations of relying on a single source of information. Furthermore, this paper tests these methods on the data used in the aforementioned studies.

This paper is organized as follows: In the next section, we provide a brief overview of related work for longitudinal multimodal data modelling. In Section 3, we describe the proposed model and its implementation details. In Section 4, we present the results of our experiments and a detailed analysis of the model's performance. Finally, in Section 5, we conclude the paper and discuss future work.

II. Related Works

A. Longitudinal Data

Longitudinal data, sometimes also known as panel data, refers to data collected over time from the same individuals, repeatedly. This can be very useful for biomedical research as it sheds insight into the progression of diseases and disorders. By treating each time point as a separate feature in the model, standard machine learning algorithms like support vector machines (SVMs) and random forests can be trained to predict and classify longitudinal data. However, these methods do not preserve sequential integrity of the data. A common method for accounting for time effects is the linear mixed-effects (LMEs) model. These models are commonly used to account for the within-subject correlation that is inherent in longitudinal data. [8], [9] However, strong assumptions about the distribution of the residual errors, such as normality, homoscedasticity, and independence limit the wide adoption of LMEs.

To overcome these limitations and capture both linear and non-linear patterns in the data, Recurrent neural networks (RNNs) are state-of-art methods, well-suited for modelling sequential time series data. RNNs have been adapted into various architectures like the LSTM and GRU which have been further tuned for modelling biomedical datasets [10]–[12]. Liquid state machine (LSM) is another method that can naturally handle sequential data. The LSM is a dynamic reservoir of interconnected spiking neurons. By learning the connections between the reservoir and the output layer, the LSM can learn to predict with temporal patterns in the data [13].

B. Multimodal Data

Methods for multimodal data modelling can largely be categorised into early fusion, intermediate fusion, and late fusion. Early fusion has been used for integration of gene expression and DNA methylation data for prediction of Alzheimer's disease [14], for cancer prognosis by combining multi-omics and clinical features [15] and for cancer survival analysis with multi-omics data [16]. These works performed early fusion via concatenation and feature selection and the combined features were used as input to deep learning models. Other approaches like auto encoders for learning a lower dimensional joint latent representation have also been employed [17], [18].

Intermediate fusion is when the modalities are gradually fused to learn marginal representations to capture within modality relations prior to learning joint representations or making predictions directly. This type of fusion has also been applied in biomedical literature for integrating mRNA, miRNA and methylation datasets to predict cancer survival subgroups [19]. It has also been applied to integrate multi-omics data for breast cancer survival prediction [20]. In these approaches, marginal representations of each modality were constructed via autoencoders, and the latent features were later concatenated to learn the joint representations.

Late fusion has shown effectiveness when integrating heterogeneous modalities such as imaging, medical records, and tabular data. A case study on pulmonary embolism showed that late fusion to integrate CT imaging and electronic health records resulted in better prediction accuracy compared to early and intermediate fusion [21]. However, a drawback of late fusion is its inability to learn joint interactions between features from different modalities.

C. Mix of Longitudinal and Multimodal Data

The fusion of multimodal longitudinal data is a challenging task, but recent advances in deep learning techniques have shown great potential in addressing this issue. RNNs, LSTMs, and CNNs have been successfully combined in fusion techniques to handle multimodal sequential data. For example, multimodal retinal images from longitudinal clinical studies have been accurately analysed to detect structural changes in large-scale datasets [22]. In Alzheimer's disease, the integration of MRI, PET images, cognitive scores, neuropathology, and assessment data has led to improved classification of patients from healthy controls [23]. Additionally, LSTMs and CNNs have been used to predict ICU interventions using data from multiple modalities such as vitals, labs, notes and demographics [24]. Recently, SNNs have shown promise in short-term emotion recognition using EEG and facial landmarks [25], and in enhancing EEG classification by leveraging MRI data to design the architecture [26]. Another study used a deep learning architecture paired with a linear model to learn task-specific feature representations and predict the progression of Alzheimer's disease using longitudinal multimodal neuroimaging datasets [27]. These results demonstrate the potential of deep learning techniques, including SNNs, in multimodal temporal data modelling.

III. Methodology

This section provides a description of the methodologies and techniques employed in this study. It delves into the intricacies of Liquid State Machines (LSMs), its proposed derivative for multimodal learning - the Mosaic LSM, and the experimental setup including software implementation.

The dataset and the pre-processing procedures used for the experiments are meticulously documented.

A. Liquid State Machines

A Liquid State Machine (LSM) is a spiking neural network architecture based on the reservoir computing paradigm [28]. It consists of a large number of interconnected spiking neurons also known as nodes. Each node receives a time-varying input from external sources as well as other nodes. The nodes are connected to each other at random and the connection weights are kept fixed during computation. This forms a dynamical system known as the reservoir, which transforms the input into a spatio-temporal pattern of activations in the network. These activations are passed through a linear readout layer to produce the output of the system. This readout layer associates the state (activation) of the reservoir, in response to a given input, with the expected output.

Processing the input through the reservoir network results in the application of a vast collection of nonlinear functions. Theoretically, a linear combination (using the readout units) of many such non-linear functions can be used to accomplish any mathematical operation, even complex tasks such as speech recognition or affect recognition. The word 'liquid' in liquid state machine comes from the analogy of ripples (spatio-temporal patterns) generated in the reservoir when a stone (input) is dropped into it.

B. Mosaic LSM

Our data is of the form $D = \{(X_1^1 .. X_1^m, y_1), \ldots (X_n^1 .. X_n^m, y_n)\}$, where $X_n^1 .. X_n^m$ represents n_{th} sample for the m_{th} data modality and y_n denotes the corresponding class label. Each $X_n^m \in R^{t \times d}$ represents a sequence of t time points $X_n^m = \{x_1 \ldots x_t\}$. The objective is to learn the mapping $f : \mathrm{X} \rightarrow \mathrm{Y}$, using the proposed mosaic LSM, which is composed of the components listed below.

In early fusion, all the modalities $X_1 \ldots X_m$ are concatenated to form a single dataset, which is then transformed into spike-encoded data. This combined dataset is passed through the reservoir, allowing the model to process all the modalities simultaneously. In contrast, for intermediate and late fusion, each modality has a separate reservoir to capture marginal temporal dependencies within the modality. The sparsely connected reservoir of spiking neurons projects the spike encoded data into a higher dimensional space. Depending upon the type of fusion (intermediate, or late), different strategies are employed to integrate the information from multiple reservoirs (representing each modality) before mapping it to the desired output (illustrated in Fig. 1).

1) Spike Encoder: To use LSMs for modeling, the continuous data needs to be transformed into discrete spikes, as LSMs work on event-based processing, similar to the action potentials of neurons in the brain. There are various methods to achieve this including, Bens Spiker Algorithm (BSA) and Moving window spike encoding algorithm etc. [29] Here, we have utilized the step forward algorithm [30] 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 also updated as B+th. If the input feature value at time t is less than B+th, then 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.

2) Input Layer: The input layer is sparsely connected to the reservoir neurons based on random numbers drawn from the uniform distribution. Formally, the connections are defined by the $d \times N$ matrix W_i which is drawn from $unif(a, b)$. Each modality $X_n^1 \ldots X_n^m$ has a separate input layer defined by this procedure.

3) 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 that neurons closer together have a greater probability of being connected than those farther apart. For example, the connection weight between neuron a and neuron b is defined by:

$$P_{a,b} = \begin{cases} C \cdot e^{(d_{a,b}/\lambda)^2} \; if \; d_{a,b} \leq d_{thresh} \\ 0 \qquad\qquad otherwise \end{cases} \qquad (1)$$

Where $d_{a,b}$ is the Euclidian distance between pair of neurons a and b and d_{thresh} is the distance threshold. C defines the maximum connection probability and λ defines the small world connectivity radius. These parameters can be specified to control the level of sparsity in the reservoir. For the Mosaic LSM, the synaptic connections weights for all neurons in the reservoir are stored in a $N \times N$ matrix W_j which are formed based on equation (1). Liquid neurons were either excitatory or inhibitory based on 80/20 ratio. Connections from an inhibitory neuron had negative synaptic weights which help regulate the spiking activity in the reservoir.

We used the leaky integrate and fire (LIF) neuron for modelling the network dynamics. The LIF neuron is a simplified mathematical model of the biological neuron [32], widely used in implementations of spiking neural networks. The neuron fires when its membrane potential, which accumulates over time as a result of incoming spikes, reaches a certain threshold value. After firing, the membrane potential is reset to a resting value and the neuron becomes temporarily refractory, meaning it cannot fire again for a short period of time. The 'leaky' term in the model refers to the gradual exponential decay of the membrane potential over time. This behaviour is described (2-4):

$$\mu[t] = \mu[t-1]e^{\frac{-1}{\tau}}\big(1 - \theta_j[t]\big) + I[t] \qquad (2)$$

Here, $\mu[t]$ describes (and updates) the membrane potential of a neuron at time t, τ is the membrane's time constant, θ_j represents the occurrence of a spike in the j_{th} neuron of the reservoir at time t. $(1 - \theta_j)$ is the reset mechanism embedded into update equation. In case of a spike, represented by 1, the membrane potential is reset to its resting value which is 0. $I[t]$ represents the current being injected into the neuron defined by (3).

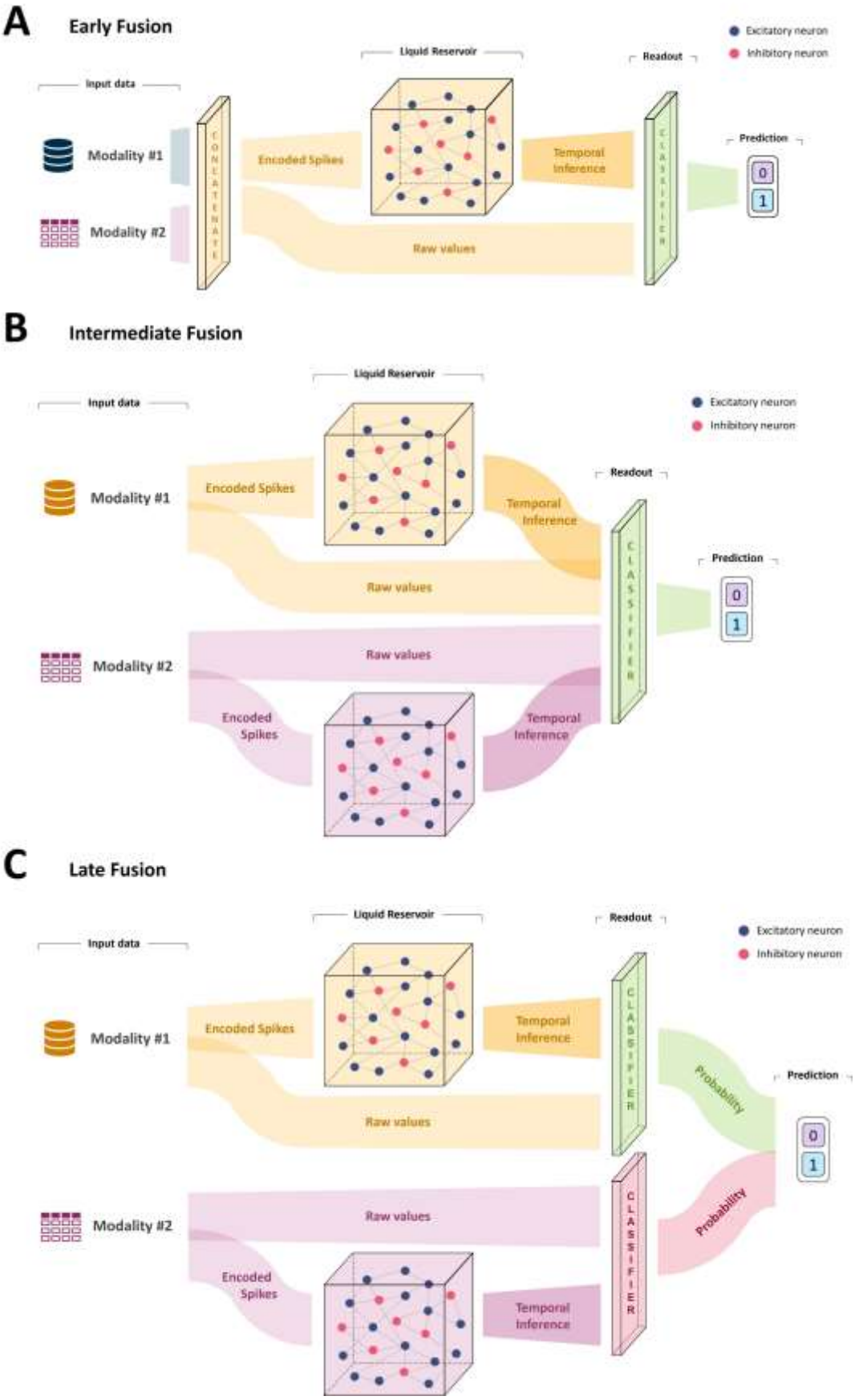

Fig. 1: The Mosaic LSM architecture for multimodal learning (A) using early fusion strategy. (B) using intermediate fusion strategy. (C) using late fusion strategy.

$$I[t] = \sum_{i=1}^{d} W_i \theta_i[t] + \sum_{j=1}^{N} W_j \theta_j[t] \quad (3)$$

$$\theta_j[t] = \begin{cases} 1 & \mu_j[t-1] \geq \mu_{th} \\ 0 & otherwise \end{cases} \quad (4)$$

The current $I[t]$ into a neuron at time t defined by (3) is computed by taking the weighted sum of incoming spikes from the input layer plus the weighted sum of incoming spikes from neighbouring neurons in the reservoir. Whether a neuron spikes or not is regulated by (4), neuron j produces a spike (indicated by 1) at time t, if its membrane potential reaches a predefined threshold μ_{th}. Each modality is passed through separate reservoirs (reservoirs for each modality can have different hyperparameters i.e., thresholds and constants.)

4) Concatenation Layer (for intermediate fusion): If the objective is to learn both the joint and marginal representations of the modalities, it is more suitable to perform intermediate fusion. This involves stacking the features and representations from each modality into a single high-dimensional vector, which is then fed as input to the subsequent readout layer.

5) Readout Layer: The readout layer acts as an interface between the highly nonlinear dynamics of the reservoirs and the final outputs of the system. For intermediate fusion, the number of times each neuron spiked for each modality and the original data is passed to Gradient Boosted Decision Trees (GBDT) for a single combined readout and classification. For late fusion, separate readouts for each modality are used to obtain separate classification probabilities for each modality.

6) Combined Probability (for late fusion): In the case of late fusion, the probabilites obtained from the separate readouts for each modality is combined by taking a weighted average. This results in marginal representations, as representations are learned independently for each modality, without considering the 'inter-modality' relationships. This can be useful and result in higher performance if the relationships between modalities are 'weak' or 'insignificant'.

C. Dataset and Data Preparation

The LYRIKS dataset [33] was used in this study to assess the capabilities of the Mosaic LSM approach. The dataset contains clinical, cognitive, and genetic data collected from Ultra-High Risk (UHR) and control groups over a period of two years. The clinical assessments in the dataset included the Clinical Assessment Interview for At-Risk Mental States (CAARMS), the Positive and Negative Syndrome Scale (PANSS), and the Calgary Depression Scale for Schizophrenia (CDSS) and were conducted every 6 months while the gene expression was recorded at an interval of 12 months and sequenced using RNASeq. Cognitive tests included Brief Assessment of Cognition in Schizophrenia (BACS), Continuous Performance Test (CPT), Snakes in Grass, and Perceptual Closure (PerClose) and were also conducted every 6 months.

To ensure the comparability amongst the data modalities, all of them were normalized using min-max scaling. The objective set forth in this study was predicting 6 months in advance, whether an individual will be UHR for psychosis or not. The data from 0-18 months was used to predict the UHR status at 24 months. Of the total participants in the LYRIKS dataset, 66 individuals had data for all three modalities, at all the timesteps, with minimal missing values. Out of these 66 participants, 40 were classified as no-risk at 24 months, including 20 healthy individuals and 20 individuals who were previously UHR but remitted. The remaining 26 participants were classified as at-risk, including 6 individuals who developed full-blown psychosis.

The missing values were imputed using temporal interpolation i.e., using the available previous or next timesteps. The step-forward encoding was used to translate the longitudinal data into spike trains for the Liquid State Machine (LSM) reservoir, with the 0-18 months data being interpolated into 52 timesteps to ensure sufficient spiking activity in the reservoir.

D. Experimental Setup

The Mosaic LSM was implemented in Python and the snnTorch library was used to handle the spike trains. The experiments were carried out on Google Colab, a cloud-based computing resource for jupyter notebooks. GPU acceleration was used for faster and more efficient computations. Since the number of samples in the data was less, 5-fold cross validation was used to obtain a robust measure of model performance. To ensure a fair comparison, the same 5 folds were used for all the models. The hyperparameters of LSM reservoir were tuned within each fold, on the training set, using grid search. For the genetic modality, feature selection was performed within each fold, to select the top 10 genes relevant to UHR prognosis.

The Mosaic LSM model was compared with four other popular and well-established machine learning models for time-series analysis, namely Gradient Boosted Decision Trees (GBDT), Support Vector Machines (SVM), Long Short-Term Memory (LSTM), and simple Liquid State Machine (LSM). The GBDT implementation was taken from the XGBoost library, the SVM from scikit-learn, and LSTM from TensorFlow Keras. The LSTM model architecture consisted of two LSTM layers of size 32, two dense layers of size 64 and 32, and a final SoftMax layer. The adam optimizer was employed to tune the weights with the learning rate set to 5×10^{-4}. The performance of these models was assessed using three metrics: accuracy, Matthew's correlation coefficient (MCC), and confusion matrix.

IV. Results

In this section, we present the results of our study on the prognosis of Ultra-High Risk (UHR) for psychosis, including an ablation analysis where we evaluated the efficacy of each modality separately. We first report the results of prognosis using single modalities, clinical, cognitive, and genetic. The objective here is to understand the performance of each modality by itself. Next, we present the results of multimodal integration of these modalities. The aim is to evaluate the performance improvement

achieved by fusing the multiple modalities together and the effectiveness of our proposed Mosaic LSM architecture.

A. Unimodal Prognosis

Clinical assessments serve as the established method for diagnosis of UHR in individuals. However, they suffer from a few limitations such as subjective interpretation of answers by assessors, potential biases in answering and inability to capture the complete status of an individual's mental health. Despite these limitations, clinical assessments remain the most commonly used modality for prognosis due to the lack of alternative modalities with sufficient validation. Therefore, we begin by evaluating the efficacy of the clinical modality first and then proceed to compare the efficacy of the other modalities, genetic and cognitive, with that of the clinical modality. The results of the unimodal prognosis are presented in Table 1 and visualised in Fig. 2.

1) Clinical data: The results show that the Mosaic LSM achieves the highest accuracy 87.9% followed by SVM and GBDT with 86.4%. LSM achieves 84.8% while LSTM achieves the lowest accuracy 78.8%. The MCC, which accounts for true and false positives and negatives, also reflects a similar ranking. The optimal paramters for the reservoir were usually found to be the ones that allowed a significant amount of spiking activity and SVM was found to be the appropriate readout classifier.

2) Genetic data: For the genetic modality, Mosaic LSM outperforms the other models with an accuracy of 86.4% and MCC of 0.74. Comparatively, the basic LSM only achieves an accuracy of 80.3% while SVM performs the second best with an accuracy of 83.3%. The optimal parameters for the reservoir were found to be ones that severely constrained the spiking activity in the reservoir. GBDT was found to be the optimal readout classifier which aligns with the nature of the rna data because it contains counts. Overall, the genetic modality showed promising results, achieving accuracies close to the clinical modality while also being an objective measure of an individual's state of health. However, it is important to note that only a limited number of samples were used in this study and more research and improvement on the genetic modality is needed to match the accuracy of the clinical modality for UHR prognosis.

3) Cognitive data: The results show that Mosaic LSM again achieves the highest accuracy (72.7%) and MCC (0.42). GBDT and SVM had MCC close to zero indicating they failed to learn anything from the data while LSM and LSTM showed weak correlation of 0.35 and 0.22 respectively. The optimal parameters for the reservoir were found to be close to the default parameters, where regular spiking was sustained and SVM was found to be the optimal readout classifier. While the accuracies were lower using the cognitive modality, by achieving a MCC of 0.42, the Mosaic LSM demonstrated its potential as a complementary modality, which when combined with other modalities, can provide additional insights into the mental health status of individuals.

TABLE I. SIX-MONTH AHEAD PREDICTION OF UHR USING UNIMODAL PROGNOSTIC MODELS. MODEL PERFORMANCE ARE EVALUATED USING ACCURACY, MCC AND CONFUSION MATRICES

Data	**Model**	**Performance**		
		Accuracy	***MCC***	***Confusion Matrix***
Clinical	XGBoost	86.4	0.71	[36 4] [5 21]
	SVM	86.4	0.72	[35 5] [4 22]
	LSTM	78.8	0.57	[31 9] [5 21]
	LSM	84.8	0.69	[34 6] [4 22]
	Mosaic LSM	**87.9**	**0.74**	**[37 3] [5 21]**
Genetic	XGBoost	81.8	0.62	[34 6] [6 20]
	SVM	83.3	0.66	[33 7] [4 22]
	LSTM	78.8	0.56	[32 8] [6 20]
	LSM	80.3	0.58	[35 5] [8 18]
	Mosaic LSM	**86.4**	**0.72**	**[35 5] [4 22]**
Cognitive	XGBoost	56	0	[32 8] [21 5]
	SVM	53	-0.1	[30 10] [21 5]
	LSTM	65.2	0.22	[35 5] [18 8]
	LSM	69.7	0.35	[32 8] [12 14]
	Mosaic LSM	**72.7**	**0.42**	**[33 7] [11 15]**

The best results are highlighted.

B. Multimodal Prognosis

In recent years, the healthcare industry has witnessed a growing trend of utilizing multiple data modalities for disease diagnosis and prognosis. The use of multiple modalities leads to a more holistic view of an individual's health and has been shown to improve diagnostic and prognostic accuracy. Hence, in this study, we investigated the feasibility of combining clinical, genetic, and cognitive modalities for UHR prognosis. All three multimodal integration strategies of the Mosaic LSM – early fusion, intermediate fusion, and late fusion, as described in Section 2B and shown in Fig. 1, were tested on this data.

In the simple LSM, a separate reservoir was employed for each modality with a shared SVM readout classifier. For the remaining models, an early fusion approach was used where the three modalities were concatenated before passing on to the model. The Mosaic LSM used SVM as its readout classifier for early and intermediate fusion, whereas for late fusion, the most optimal readout classifiers were employed based on the results of the unimodal analysis. For the clinical and cognitive modalities, SVM was utilized, while GBDT was utilized for the genetic modality. The class probabilities of the three readout classifiers were combined using the ratio of 1:0.25:0.1 for the clinical, genetic, and cognitive modalities, respectively. The results of the experiment are presented in Table 2 and Fig.2.

The performance of the models varied when tested on the multimodal data. Most models showed marked improvement in accuracy and MCC, with GBDT being the only exception. The LSTM model showed marginal improvements in accuracy, however, still had the lowest

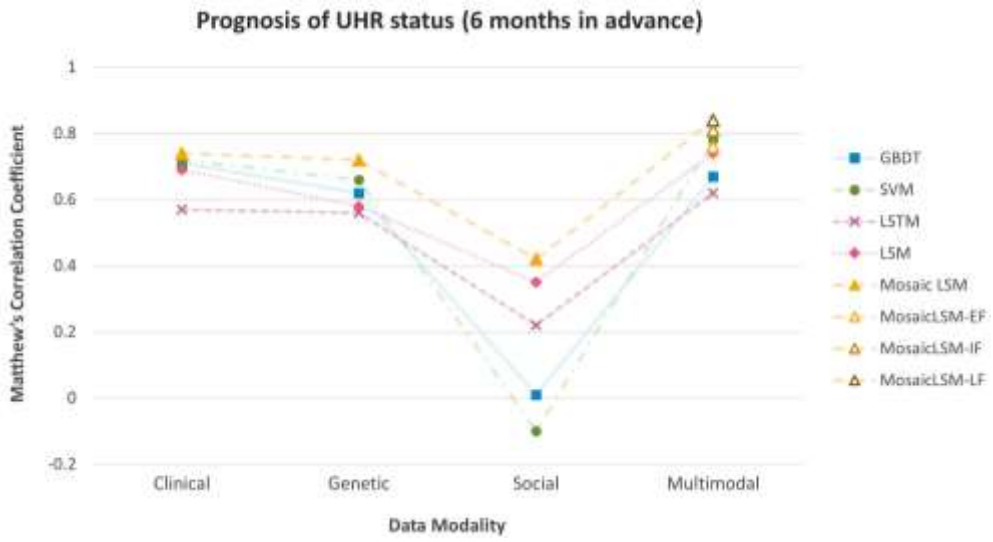

Figure 2: Comparison of MCC achieved by various classifiers using the different data modalities for UHR prognosis.

accuracy among all the models. The Mosaic LSM outperformed all models, achieving a remarkable 92.4% accuracy and 0.84 MCC with the late fusion strategy. The SVM model followed closely behind, achieving 89.4% accuracy and 0.78 MCC. Even the simple LSM showed a notable improvement with 87.9% accuracy and 0.74 MCC.

The superior performance of the late fusion version of Mosaic LSM on this heterogeneous UHR dataset can be attributed to the adaptive weighting of modalities. Intermediate fusion tends to assign equal importance to all modalities, which can negatively impact the model's performance when some modalities underperform. In contrast, early fusion merges the modalities at the beginning of the learning process, potentially leading to the loss of modality-specific temporal information. Late fusion, by assigning more weight to well-performing modalities, achieves superior accuracy. The results obtained from the Mosaic LSM demonstrate its potential to learn from multiple sources of information. The LSM in Mosaic LSM adeptly captures temporal patterns in the data, which when combined with the raw values, provides more information to the readout classifier, and improves accuracy. Notably, both intermediate and late integration strategies of Mosaic LSM achieved an MCC exceeding 0.8, which is highly significant for early-stage disease prognosis.

TABLE II. SIX-MONTH AHEAD PREDICTION OF UHR USING MULTIMODAL PROGNOSTIC MODELS. MODEL PERFORMANCE ARE EVALUATED USING ACCURACY, MCC AND CONFUSION MATRICES

Data	**Model**	**Performance**		
		Accuracy	***MCC***	***Confusion Matrix***
Clinical + Genetic + Cognitive	XGBoost	83.3	0.67	[32 8] [3 23]
	SVM	89.4	0.78	[37 3] [4 22]
	LSTM	81.8	0.62	[34 6] [6 20]
	LSM	87.9	0.74	[37 3] [5 21]
	MosaicLSM-EF	87.9	0.75	[38 2] [6 20]
	MosaicLSM-IF	90.9	0.81	[36 4] [2 24]
	MosaicLSM-LF	**92.4**	**0.84**	**[37 3]** **[2 24]**

The best results are highlighted.

V. CONCLUSION

This paper presents a novel and promising approach for multimodal learning from longitudinal data. The field of AI has seen a surge in the use of multimodal data for various tasks such as natural language understanding, human activity recognition, sentiment analysis, speech recognition, and disease diagnosis and prognosis. Despite the existence of many methods for multimodal data analysis, there is a lack of approaches that can effectively model temporal data, especially in smaller datasets.

The proposed Mosaic LSM architecture addresses the challenges of modelling complex temporal relationships between various modalities by incorporating time-series information in the dynamic state of the reservoir. This information is then combined with the raw dataset to create a joint representation (early and intermediate fusion) or marginal representations (late fusion), depending on the chosen approach, prior to being fed into the readout classifier. By exploiting the strengths of each modality, this approach overcomes the limitations of relying on a single source of information. The results of the study on the LYRIKS dataset demonstrate that the Mosaic LSM outperforms traditional models such as Support Vector Machines (SVM), Gradient Boosting Decision Trees (GBDT), and Long Short-Term Memory Networks (LSTM) for UHR prognosis, with a significant MCC, greater than 0.8.

This work has significant implications for the future of healthcare and personalized medicine, particularly in the integration of multimodal longitudinal data, such as EEG and fMRI. The Mosaic LSM has the potential to provide a more accurate diagnosis, prognosis, and treatment planning by leveraging multiple modalities and accounting for their temporal dynamics. However, as with other artificial neural networks, the initial weights of the connections can greatly impact the results, particularly in smaller datasets. Further research is needed to develop a more robust approach for initializing the connectivity in LSMs. To enhance the interpretability of the reservoir, future work will draw

inspiration from NeuCube. Overall, our study represents an important step forward in the application of spiking neural networks for multimodal learning and analysis of longitudinal data.

ACKNOWLEDGMENTS

This research is supported by the MBIE Catalyst: Strategic New Zealand-Singapore Data Science Research Programme Funding and the National Research Foundation, Singapore under its Industry Alignment Fund – Pre-positioning (IAF-PP) Funding Initiative. The LYRIKS study was supported by the National Research Foundation Singapore under the National Medical Research Council Translational and Clinical Research Flagship Programme (NMRC/TCR/003/2008). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

REFERENCES

[1] S. M. Schüssler-Fiorenza Rose *et al.*, "A longitudinal big data approach for precision health," *Nat Med*, vol. 25, no. 5, 2019, doi: 10.1038/s41591-019-0414-6.

[2] A. Kline *et al.*, "Multimodal machine learning in precision health: A scoping review," *NPJ Digit Med*, vol. 5, no. 1, p. 171, 2022, doi: 10.1038/s41746-022-00712-8.

[3] S. C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, "Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines," *npj Digital Medicine*, vol. 3, no. 1. 2020. doi: 10.1038/s41746-020-00341-z.

[4] S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, "Multimodal deep learning for biomedical data fusion: A review," *Briefings in Bioinformatics*, vol. 23, no. 2. 2022. doi: 10.1093/bib/bbab569.

[5] W. Maass and H. Markram, "On the computational power of circuits of spiking neurons," *J Comput Syst Sci*, vol. 69, no. 4, 2004, doi: 10.1016/j.jcss.2004.04.001.

[6] B. Singh *et al.*, "Constrained neuro fuzzy inference methodology for explainable personalised modelling with applications on gene expression data," *Scientific Reports 2023 13:1*, vol. 13, no. 1, pp. 1–15, Jan. 2023, doi: 10.1038/s41598-022-27132-8.

[7] Z. Doborjeh *et al.*, "Investigation of social and cognitive predictors in non-transition ultra-high-risk' individuals for psychosis using spiking neural networks," *Schizophrenia 2023 9:1*, vol. 9, no. 1, pp. 1–10, Feb. 2023, doi: 10.1038/s41537-023-00335-2.

[8] K. E. Langbehn *et al.*, "Behavioral deficits in juvenile onset huntington's disease," *Brain Sci*, vol. 10, no. 8, 2020, doi: 10.3390/brainsci10080543.

[9] J. C. Morris *et al.*, "Pittsburgh compound B imaging and prediction of progression from cognitive normality to symptomatic Alzheimer disease," *Arch Neurol*, vol. 66, no. 12, 2009, doi: 10.1001/archneurol.2009.269.

[10] R. Gao *et al.*, "Distanced LSTM: Time-Distanced Gates in Long Short-Term Memory Models for Lung Cancer Detection," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019. doi: 10.1007/978-3-030-32692-0_36.

[11] J. Andjelkovic *et al.*, "Sequential machine learning in prediction of common cancers," *Inform Med Unlocked*, vol. 30, 2022, doi: 10.1016/j.imu.2022.100928.

[12] A. Guo, R. Beheshti, Y. M. Khan, J. R. Langabeer, and R. E. Foraker, "Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models," *BMC Med Inform Decis Mak*, vol. 21, no. 1, 2021, doi: 10.1186/s12911-020-01345-1.

[13] A. Patiño-Saucedo, H. Rostro-González, T. Serrano-Gotarredona, and B. Linares-Barranco, "Liquid State Machine on SpiNNaker for Spatio-Temporal Classification Tasks," *Front Neurosci*, vol. 16, 2022, doi: 10.3389/fnins.2022.819063.

[14] C. Park, J. Ha, and S. Park, "Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset," *Expert Syst Appl*, vol. 140, 2020, doi: 10.1016/j.eswa.2019.112873.

[15] G. Xie, C. Dong, Y. Kong, J. F. Zhong, M. Li, and K. Wang, "Group lasso regularized deep learning for cancer prognosis from multi-omics and clinical features," *Genes (Basel)*, vol. 10, no. 3, 2019, doi: 10.3390/genes10030240.

[16] L. Zhao *et al.*, "DeepOmix: A scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis," *Comput Struct Biotechnol J*, vol. 19, 2021, doi: 10.1016/j.csbj.2021.04.067.

[17] J. YU *et al.*, "A model for predicting prognosis in patients with esophageal squamous cell carcinoma based on joint representation learning," *Oncol Lett*, vol. 20, no. 6, 2020, doi: 10.3892/ol.2020.12250.

[18] L. Zhang *et al.*, "Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma," *Front Genet*, vol. 9, no. OCT, 2018, doi: 10.3389/fgene.2018.00477.

[19] O. B. Poirion, K. Chaudhary, and L. X. Garmire, "Deep Learning data integration for better risk stratification models of bladder cancer.," *AMIA Jt Summits Transl Sci Proc*, vol. 2017, 2018.

[20] L. Tong, J. Mitchel, K. Chatlin, and M. D. Wang, "Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis," *BMC Med Inform Decis Mak*, vol. 20, no. 1, 2020, doi: 10.1186/s12911-020-01225-8.

[21] S. C. Huang, A. Pareek, R. Zamanian, I. Banerjee, and M. P. Lungren, "Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection," *Sci Rep*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-78888-w.

[22] T. de Silva, E. Y. Chew, N. Hotaling, and C. A. Cukras, "Deep-learning based multi-modal retinal image registration for the longitudinal analysis of patients with age-related macular degeneration," *Biomed Opt Express*, vol. 12, no. 1, 2021, doi: 10.1364/boe.408573.

[23] S. El-Sappagh, T. Abuhmed, S. M. Riazul Islam, and K. S. Kwak, "Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, 2020, doi: 10.1016/j.neucom.2020.05.087.

[24] H. Suresh *et al.*, "Proceedings of Machine Learning for Healthcare 2017 Clinical Intervention Prediction and Understanding with Deep Neural Networks," *Ml4H*, vol. 68, 2017.

[25] C. Tan, M. Šarlija, and N. Kasabov, "NeuroSense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns," *Neurocomputing*, vol. 434, pp. 137–148, Apr. 2021, doi: 10.1016/J.NEUCOM.2020.12.098.

[26] S. A. Saeedinia, M. R. Jahed-Motlagh, A. Tafakhori, and N. Kasabov, "Design of MRI structured spiking neural networks and learning algorithms for personalized modelling, analysis, and prediction of EEG signals," *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–14, Jun. 2021, doi: 10.1038/S41598-021-90029-5.

[27] G. Lee, B. Kang, K. Nho, K. A. Sohn, and D. Kim, "MildInt: Deep learning-based multimodal longitudinal data integration framework," *Front Genet*, vol. 10, no. JUN, 2019, doi: 10.3389/fgene.2019.00617.

[28] W. Maass, "Liquid state machines: Motivation, theory, and applications," in *Computability in Context: Computation and Logic in the Real World*, 2011. doi: 10.1142/9781848162778_0008.

[29] D. Auge, J. Hille, E. Mueller, and A. Knoll, "A Survey of Encoding Techniques for Signal Processing in Spiking Neural Networks," *Neural Processing Letters*, vol. 53, no. 6. 2021. doi: 10.1007/s11063-021-10562-2.

[30] N. Kasabov *et al.*, "Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: Design methodology and selected applications," *Neural Networks*, vol. 78, 2016, doi: 10.1016/j.neunet.2015.09.011.

[31] H. Hazan and L. M. Manevitz, "Topological constraints and robustness in liquid state machines," in *Expert Systems with Applications*, 2012. doi: 10.1016/j.eswa.2011.06.052.

[32] L. F. Abbott, "Lapicque's introduction of the integrate-and-fire model neuron (1907)," *Brain Research Bulletin*, vol. 50, no. 5–6. 1999. doi: 10.1016/S0361-9230(99)00161-6.

[33] J. Lee *et al.*, "The longitudinal youth at risk study (LYRIKS)—an Asian UHR perspective," *Schizophr Res*, vol. 151, no. 1–3, pp. 279–283, 2013.