Федеральное государственное автономное образовательное учреждение

высшего образования «Национальный исследовательский университет   
«Высшая школа экономики»

###### **Факультет компьютерных наук**

###### **ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

Классификация состояний мозга с использованием жидких нейронных сетей на EEG данных

Brain State Classification Using Liquid Neural Networks on EEG Data

по направлению подготовки 01.04.02 Прикладная математика и информатика

образовательная программа «Науки о данных»

Студент группы МНОД231

Селиверстов Алексей Владимирович

(Ф.И.О.)

Руководитель КР

Сохраби Маджид

(Ф.И.О.)

Москва, 2025

**Table of contents**

[Abstract 3](#_Toc197990840)

[Абстракт 3](#_Toc197990841)

[Brain State Classification Using Liquid Neural Networks on EEG Data 5](#_Toc197990842)

[Introduction 5](#_Toc197990843)

[Related papers 5](#_Toc197990844)

[Data 7](#_Toc197990845)

[Methodology 8](#_Toc197990846)

[Training setup 8](#_Toc197990847)

[Quality metrics 9](#_Toc197990848)

[Models 9](#_Toc197990849)

[Experimental Results 10](#_Toc197990850)

[Conclusions 17](#_Toc197990851)

[References (NOT UPDATED YET) 18](#_Toc197990852)

## Abstract

This thesis explores the application of deep learning techniques for classifying cognitive states from EEG data, focusing on confusion detection. Electroencephalography (EEG) is widely used for non-invasive monitoring of brain activity, offering real-time insights into mental states such as attention, fatigue, and confusion. Despite its potential, analyzing EEG signals remains difficult due to noise, variability among individuals, and complex brain dynamics. Recent developments in machine learning, particularly deep learning, have made it more feasible to decode EEG signals. Traditional classifiers like Support Vector Machines and Naive Bayes have achieved good results in detecting clear mental states like drowsiness. However, recognizing more subtle states, like confusion during learning, requires models capable of capturing temporal dependencies and non-linear interactions. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), are well-suited for this task since they model EEG data as time series. Nonetheless, even the most advanced RNNs often struggle with non-stationary signals and changing contexts. Liquid Neural Networks (LNNs), inspired by biological neurons, present a promising alternative. They dynamically adapt to new patterns without retraining, thanks to evolving internal states. This makes them more suitable for real-time applications, especially on low-resource devices. In this work, we compare the performance of three neural network architectures – GRUs, LSTMs, and LNNs – for classifying confusion from EEG recordings. Our data was collected from students during controlled learning sessions, using a low-cost EEG headset. The goal was to assess each model’s ability to recognize cognitive states based on temporal patterns in the recorded signals. The findings suggest that while GRUs and LSTMs perform well in capturing long-term dependencies, Liquid Neural Networks show advantages in adaptability and efficiency. The results support the potential of LNNs for real-time cognitive state monitoring in adaptive educational systems and brain-computer interfaces. Future research could explore hybrid models or integrate attention mechanisms to further enhance performance.

## Абстракт

В этой работе исследуется применение методов глубокого обучения для классификации когнитивных состояний на основе данных ЭЭГ с акцентом на выявление состояния замешательства. Электроэнцефалография (ЭЭГ) широко используется для неинвазивного мониторинга активности мозга, обеспечивая получение информации в реальном времени о таких ментальных состояниях, как внимание, усталость и замешательство. Несмотря на потенциал, анализ сигналов ЭЭГ остается сложной задачей из-за шума, межиндивидуальной изменчивости и сложной динамики мозга. Недавние достижения в области машинного обучения, особенно глубокого обучения, значительно улучшили возможности декодирования сигналов ЭЭГ. Традиционные классификаторы, такие как машины опорных векторов и наивные байесовские модели, показали хорошие результаты при распознавании очевидных ментальных состояний, например, сонливости. Однако для распознавания более тонких состояний, таких как замешательство во время обучения, требуются модели, способные учитывать временные зависимости и нелинейные взаимодействия. Рекуррентные нейронные сети (RNN), особенно сети долгой краткосрочной памяти (LSTM) и управляемые рекуррентные блоки (GRU), хорошо подходят для этой задачи, поскольку они моделируют данные ЭЭГ как временные ряды. Тем не менее, даже самые продвинутые RNN часто сталкиваются с трудностями при работе с нестационарными сигналами и меняющимися контекстами. Жидкие нейронные сети (LNN), вдохновленные биологическими нейронами, представляют собой перспективную альтернативу. Они динамически адаптируются к новым паттернам без повторного обучения благодаря эволюционирующим внутренним состояниям. Это делает их особенно подходящими для приложений в реальном времени, особенно на устройствах с ограниченными ресурсами. В данной работе проводится сравнение трех архитектур нейронных сетей – GRU, LSTM и LNN – для классификации замешательства на основе ЭЭГ-записей. Данные были собраны у студентов в ходе контролируемых обучающих сессий с использованием недорогого ЭЭГ-гарнитуры. Целью было оценить способность каждой модели распознавать когнитивные состояния на основе временных паттернов в записанных сигналах. Полученные результаты показывают, что GRU и LSTM хорошо справляются с моделированием долгосрочных зависимостей, однако жидкие нейронные сети демонстрируют преимущества в адаптивности и эффективности. Эти результаты подтверждают перспективность LNN для мониторинга когнитивных состояний в реальном времени в адаптивных образовательных системах и интерфейсах "мозг-компьютер". В дальнейшем исследовании можно будет рассмотреть гибридные модели или интеграцию механизмов внимания для повышения точности.

**Keywords:** liquid neural networks, EEG data, recurrent neural networks, cognitive state classification, confusion detection, deep learning, temporal data analysis, adaptive learning systems.

**Source code:** https://github.com/As17-01/brain\_signals

# Brain State Classification Using Liquid Neural Networks on EEG Data

## Introduction

Understanding and recognizing human cognitive states without using invasive methods has been an important goal for neuroscience, education, and human-computer interaction for many years. Out of many technologies developed for this purpose, electroencephalography is especially notable because it is affordable, portable, and can capture brain activity in real time. EEG data, which is complex both in time and frequency aspects, has shown to be useful for detecting mental states like attention, fatigue, drowsiness, and confusion. But interpreting such data is still a technical challenge, mainly because EEG signals are susceptibility to noise, vary a lot between people, and brain dynamics are very complex.

Recently, machine learning has improved our ability to analyze EEG signals. Traditional models like Support Vector Machines and Naive Bayes have already reached high accuracy in identifying obvious mental states like drowsiness or epileptic seizures. However, more subtle cognitive states – for example, confusion during learning – are much harder to detect. For this, models must take into account time-related dependencies and nonlinear interactions inside EEG signals. Deep learning models like Long Short-Term Memory networks and Gated Recurrent Units have been promising in this task, as they can model EEG as time-series data. Because they can remember and process information across longer time periods, they are good for continuous monitoring, such as in adaptive learning systems or brain-computer interfaces.

Still, even the best recurrent models have some problems with adapting and generalizing, especially when EEG signals are non-stationary or when the context changes. Liquid Neural Networks can be a good alternative to traditional approaches. They are inspired by how biological neurons work and have internal states that evolve over time, which lets them adjust in real time to new patterns without retraining. Because they have internal memory and fewer parameters, they are well-suited for real-time EEG analysis on devices with limited resources or where data is not stable.

This thesis explores and compares the performance of three neural architectures – GRUs, LSTMs, and Liquid Neural Networks – for the task of classifying cognitive confusion from EEG data. The dataset comes from an experiment where students were watching educational videos marked as either confusing or not confusing. EEG signals were recorded from the frontal part of the brain using a low-cost headset, and activity across eight standard frequency bands was collected. The goal is to see how well each model can recognize the student's cognitive state only from the time-based patterns in EEG data.

By doing this, the work adds to the bigger goal of building real-time, scalable systems for brain state classification that can be used in personalized education, mental workload tracking, and future BCI applications. Special focus is on the adaptability and efficiency of Liquid Neural Networks, to see if their biologically inspired design can perform better than traditional recurrent models in dealing with the complexity and variability in EEG signals related to cognitive states.

## Related papers

Electroencephalography (EEG) has already been known for a long time as non-invasive and low-cost method to observe brain activity. It helps to get important information about cognitive states like fatigue, confusion, and mental workload. With development of machine learning and deep learning methods, interpretation and classification of EEG signals became much better in terms of accuracy, scalability, and stability. The use of neural network models like GRUs, LSTMs, and Liquid Neural Networks has created new possibilities for decoding brain signals with higher precision.

Machine learning applied to EEG data has already given good results in detecting different mental and physical states. For example, Support Vector Machines showed excellent results for classifying driver drowsiness by using features from various EEG frequency bands, reaching accuracy up to 99.3%. They also worked very well in detecting epileptic seizures in controlled conditions. But for more complex cognitive states like confusion, simpler models like Gaussian Naive Bayes did not perform so well and gave only moderate accuracy. These early experiments showed that better methods are needed to model time-based complexity and noise that exists in EEG signals.

Deep learning has brought strong improvements in this field. For example, Deep Belief Networks could extract high-level features directly from raw EEG data, and gave better performance compared to older methods like PCA. In predicting driver’s cognitive state, DBNs also worked better than shallow models. Convolutional DBNs went even further, learning both spatial and temporal relations in EEG data, and helped to classify brain states more reliably. Since EEG signals are naturally sequential, Recurrent Neural Networks – especially Long Short-Term Memory models – became a good option. These models are made to handle time-series data and showed success in detecting long-term patterns, like in early diagnosis of Alzheimer’s or in real-time confusion tracking during MOOC learning. To make training faster and more stable, researchers started to use batch normalization in deep models. Applying it to recurrent structures like LSTMs helped to increase training speed and final results, which made these models more suitable for practical use with EEG data. This improvement is especially important in real-time systems like adaptive learning or brain-computer interfaces.

Lately, Liquid Neural Networks became an interesting alternative to standard RNNs and LSTMs. These models have parameters that change based on the input data, which means they can adapt to new situations without retraining. This is very useful for EEG systems, where signals often change a lot because of noise, environment, or person-specific factors. LNNs showed better results than fixed models, especially when the data changes suddenly. Their ability to deal with difficult, noisy signals – like those from aircraft magnetic fields – also shows that they can be helpful in filtering EEG noise, which is usually a big problem.

Also, the use of LNNs goes beyond just neuroscience. For example, they were used in improving urban communication systems and in aeromagnetic compensation, and they managed to keep good performance without frequent retraining. These skills are similar to what is needed in EEG processing, where high adaptability and reliability are necessary.

Some architectural upgrades of LNNs were made too. Neural Circuit Policies were created based on how real nervous systems work. They use sparse Liquid Time-Constant neurons and result in models that are more efficient and easier to understand. Also, Continuous-Time Liquid Neural Networks made training and inference simpler by removing the need for numerical solvers, which is a big advantage for real-time applications. These updates show that LNNs are becoming more relevant in EEG-related research, especially for fast adaptation and low-resource scenarios.

While deep learning models like LSTMs and GRUs showed good results for EEG time-series data, they still have problems with changing input or adjusting to new situations without retraining. On the other hand, LNNs offer a biologically inspired solution that naturally deals with data variability and signal noise. Because they can model time more flexibly, they may work better than fixed models in real-time EEG analysis and cognitive state recognition.

Besides model design, another challenge in EEG is to handle confounding factors. Techniques like Select-Additive Learning and confounder-aware training were introduced to reduce impact from unrelated data, but they often need extra models or complex changes to architecture. LNNs might have advantage here, since their dynamic design can remove unnecessary patterns more naturally, and the modeling process becomes simpler.

The integration of LSTM, GRU, and Liquid Neural Network models in EEG analysis shows the move toward more adaptive, noise-resistant, and efficient systems. Each model type has its own strong sides: LSTMs are good for long-term memory, GRUs give faster results with fewer parameters, and LNNs are great for changing and noisy environments. This thesis will continue with comparison of these models for EEG-based cognitive state classification, showing their benefits and how they can improve brain-computer interface systems and learning technologies that react to user’s mental state.

## Data

This thesis is based on a publicly available dataset originally designed to explore cognitive responses – specifically confusion – in students viewing online educational video content. The data were collected under controlled experimental conditions, with the aim of studying how EEG signals reflect varying levels of cognitive load during learning.

The dataset includes recordings from ten college students, who viewed a set of ten distinct educational videos. These videos were selected and categorized prior to the experiment into two groups:

* Non-confusing: Topics presumed to be familiar and readily understandable to the average student, such as introductory algebra or geometry.
* Confusing: Advanced topics such as quantum mechanics or stem cell research, selected for their potential to induce confusion in students unfamiliar with the material.

All students watched the same ten videos, evenly split between the two categories. Each video was approximately two minutes long, but only the central one-minute segment was used for EEG data analysis. The beginning and end of each clip were trimmed to minimize transitional and non-content-related cognitive responses.

EEG signals were recorded using a single-channel wireless EEG headset positioned to monitor activity over the frontal lobe. The headset recorded data using one electrode on the forehead and two reference electrodes placed near the ears. The device sampled neural activity every 0.5 seconds, resulting in approximately 120 EEG samples per video per participant.

Each EEG sample consists of power values across multiple standard brainwave frequency bands:

* Delta
* Theta
* Alpha1, Alpha2
* Beta1, Beta2
* Gamma1, Gamma2

These eight channels were the only features used in this study. All other available information – such as participant demographics or specific details about the videos – was excluded from the analysis. The objective was to evaluate confusion purely from the perspective of neural signal data.

The primary label used in this thesis is the predefined confusion label, assigned based on the categorization of each video as either “confusing” or “non-confusing.” This classification was made independently of the participants' self-assessments and remains consistent across all viewers. The use of this label enables an investigation into whether confusion – operationalized at the video level – can be predicted from EEG features alone.

The dataset comprises over 12,000 EEG samples, corresponding to 100 individual recording sessions (10 participants × 10 videos). Each sample includes:

* A timestamp (0.5-second intervals)
* Eight EEG frequency band values (delta through gamma2)
* The predefined confusion label (binary)

These data were used to develop and evaluate predictive models aimed at distinguishing between confusing and non-confusing learning experiences based solely on patterns in brainwave activity.

## Methodology

### Training setup

The models in this study were trained using only EEG-derived features – specifically, the power values across eight standard frequency bands: delta, theta, alpha1, alpha2, beta1, beta2, gamma1, and gamma2. No participant demographic data or video metadata were included. This design choice allowed for an isolated exploration of how neural signals alone relate to cognitive confusion. If they were included, it would be easier to achieve much higher accuracy, however, it is not as valuable because it would limit the possibilities of extrapolation of the conclusions on other EEG data.

Each EEG time series corresponds to a one-minute segment, sampled at 0.5-second intervals, resulting in from 120 up to 144-time steps per recording. To ensure consistency across samples, each sequence was padded to a fixed length of 144-time steps, using zero-padding at the end of each signal. This allowed all models to accept uniform input dimensions while preserving the temporal structure of the original signals.

The dataset was split into training and testing subsets using a 70–30 ratio. For the final model evaluation, the split was performed by user, i.e. 7 users were used to train a model, the rest 3 were used for validation. To ensure robust and reliable evaluation, multiple distinct and fixed data splits were created. All models were trained and evaluated on the same series of partitions, ensuring comparability across experiments while mitigating the effect of variance due to random sampling. Each model was trained for the same number of epochs, standardizing training duration across all experiments. All experiments were run on the same machine. However, one epoch for LSNNs took approximately 5 times as long as other models. For LSTM and GRU the duration of each epoch was very similar. Also, for the best models overall I doubled the number of training epochs to observe their metric dynamics further.

In some experiments, a lagged embedding of the EEG signal was applied to enrich the input with temporal dependencies beyond the original sampling rate. This involved concatenating delayed versions of the input signal to provide the models with access to past temporal context within a given time window.

### Quality metrics

Model performance was assessed using a combination of three evaluation metrics:

* Accuracy: Measures the proportion of correctly classified samples out of all predictions. It provides a general measure of model performance across classes.

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

* AUC (Area Under the Receiver Operating Characteristic Curve): Evaluates the trade-off between true positive rate (sensitivity) and false positive rate across different classification thresholds. A higher AUC indicates better model discrimination between the classes.

TPR (True Positive Rate) =

FPR (False Positive Rate) =

* Entropy: Used as a measure of prediction uncertainty. Lower entropy indicates higher confidence in the model’s prediction. For binary classification, entropy of a prediction with probability is calculated as:

where is the Shannon entropy.

All metrics were averaged across the multiple fixed splits to obtain reliable, variance-reduced estimates of model performance.

### Models

The predictive models employed in this work consisted of three major types of neural network architectures, each with multiple internal variations:

* GRU-based networks: These models leveraged Gated Recurrent Units, which are well-suited for capturing temporal dependencies in sequential data while maintaining a relatively compact architecture. Several variations were tested with different numbers of layers and hidden dimensions.
* LSTM-based networks: Networks built with Long Short-Term Memory units were also explored due to their effectiveness in retaining long-range dependencies and mitigating vanishing gradient issues. As with the GRU models, multiple architectures were tested.
* Liquid State Neural Networks (LSNNs): Inspired by biologically plausible models, these networks incorporate liquid time-constant dynamics, allowing them to adapt their memory and response patterns over time. LSNNs were included to investigate whether models with non-static internal dynamics could outperform more traditional RNN variants on EEG time series data.

In each category, a range of architectures was evaluated to explore the design space and identify the most effective configurations for confusion detection.

All hyperparameters – including learning rate, batch size, number of layers, and hidden unit size – were manually selected based on preliminary testing and domain knowledge. These settings were held constant across similar models within each experimental group to ensure fair comparisons.

### Experimental Results

**Baseline**

As a baseline, a fully connected (FC) network architecture was implemented based on a previous study. This model flattened the EEG signal across time and frequency dimensions before feeding it into a deep multilayer perceptron. The architecture consisted of five sequential dense layers with decreasing units (200 → 100 → 50 → 16 → 1), each followed by a ReLU activation function except for the output layer, which used a sigmoid activation for binary classification.

This FC model achieved an average validation accuracy of 0.63, consistent with the source paper. However, the model contained 54,983 parameters, and its inability to account for temporal dynamics limited its effectiveness when compared to time-series-oriented approaches.

**Implemented models**

To better capture the temporal structure of the EEG signals, three categories of neural network models were evaluated: Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Liquid State Neural Networks (LSNN). All models were trained exclusively on the raw EEG frequency band features (delta, theta, alpha1, alpha2, beta1, beta2, gamma1, gamma2) and used the same training conditions:

* Adam optimizer with binary crossentropy loss
* Learning rate: 0.002
* Weight decay: 1e-7
* EMA (Exponential Moving Average): enabled
* Same number of training epochs and fixed data splits across models

Each model processed input sequences of 144-time steps per experiment, padded to a uniform length. Their architectures are described schematically below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **GRU** | **LSTM** | **LSNN** |
| **Architecture** | Input: (144 time steps, 8 EEG features)  GRU(20 units, return sequences)  GRU(20 units, return sequences)  GRU(10 units, return sequences)  GRU(10 units, return final)  Dense(1 unit, sigmoid activation) | Input: (144 time steps, 8 EEG features)  LSTM(20 units, return sequences)  LSTM(20 units, return sequences)  LSTM(10 units, return sequences)  LSTM(10 units, return final)  Dense(1 unit, sigmoid activation) | Input: (144 time steps, 8 EEG features)  LTC layer configured with an AutoNCP topology of 64 neurons and 32 interconnections (return sequences)  LTC layer configured with AutoNCP topology of 32 neurons and 16 interconnections (return final)  Dense(1 unit, sigmoid activation) |
| **Dropout** | 0.2 | 0.2 | No |
| **Best Performance\*** | ACC: 0.893  ENT: 0.237  AUC: 0.971 | ACC: 0.840  ENT: 0.303  AUC: 0.946 | ACC: 0.820  ENT: 0.239  AUC: 0.949 |
| **Best epoch** | 185 | 96 | 332 |
| **Num parameters** | 5951 | 7691 | 27105 |

\* - this is the best performing epoch for each of the models on non-embedded validation data averaged by folds. Binary cross-entropy was used for selection.

LSNN architecture leveraged the structure of NCPs to enhance temporal expressiveness while maintaining interpretability and compactness. Although the model had a relatively high parameter count, its performance remained competitive, suggesting that NCP-based LSNNs can effectively model cognitive states from EEG signals. One of the main important advantages of LSNNs is that its metrics dynamics is much more stable, compared to the other approaches, and it is much less prone to overfitting. However, there is an issue with occasional spikes in loss metric. Also, each epoch took 5 times more time, than for the rest of the models. Taking into account that it also achieved best performance after 332 epochs, it makes the network much longer to train.

Overall, all three temporal models clearly outperformed the fully connected baseline across all evaluation metrics. Among them, the GRU-based model achieved the best performance while maintaining a relatively low number of parameters, making it an optimal choice in terms of both efficiency and accuracy. The LSNN, while having the highest parameter count, still demonstrated strong performance, indicating its potential for modeling complex temporal dynamics in EEG data.

**Temporal Embeddings for Enhanced Representation**

In addition to evaluating the models on raw EEG frequency-band signals, further experiments explored the integration of temporal embeddings to enrich the input representations. These embeddings were designed to expose temporal dynamics more explicitly by incorporating historical context into each time step.

The following types of lag-based embeddings were considered:

* Lag-1 embedding: Each timestamp was augmented with the feature values from the immediately preceding time step.
* Lag-2 embedding: The EEG signal at each timestamp was additionally augmented with the values from two steps before.
* Lag-1 + Lag-2 embedding: Both first and second-order lagged values were concatenated to each current timestamp, resulting in a broader temporal window for each input vector.
* First-order difference (Δlag-1): Rather than using raw lagged values, this embedding encoded the difference between the current timestamp and the one immediately before it, highlighting the change in brain activity over time rather than the absolute value.

These augmentations provided the models with richer temporal features, potentially aiding in the identification of subtle signal shifts indicative of cognitive states such as confusion.

**Convolutional Embedding**

In a parallel approach, temporal feature engineering was explored via causal convolution. A 1D convolutional layer with 50 filters, a kernel size of 3, ReLU activation, and causal padding was applied to the raw EEG signals. This method created new time-dependent embeddings that preserved the sequence’s chronological order while capturing short-term local patterns. The convolutional layer effectively summarized nearby temporal trends within the signal and passed the enriched representation to subsequent model layers.

To assess the impact of different temporal embeddings, performance metrics – accuracy, loss, and AUC – were monitored across training epochs for each embedding type and model architecture. These evaluations were conducted across GRU, LSTM, and LSNN-based models, allowing for a comprehensive comparison of embedding effectiveness. In the case of the best-performing LSNN architecture, not all embeddings were included in the final evaluation phase. This decision was based on earlier results, where certain embeddings consistently underperformed relative to others, and were therefore omitted to streamline experimentations. Also, the best experiments were extended to 400 epochs to observe the metrics dynamics further.

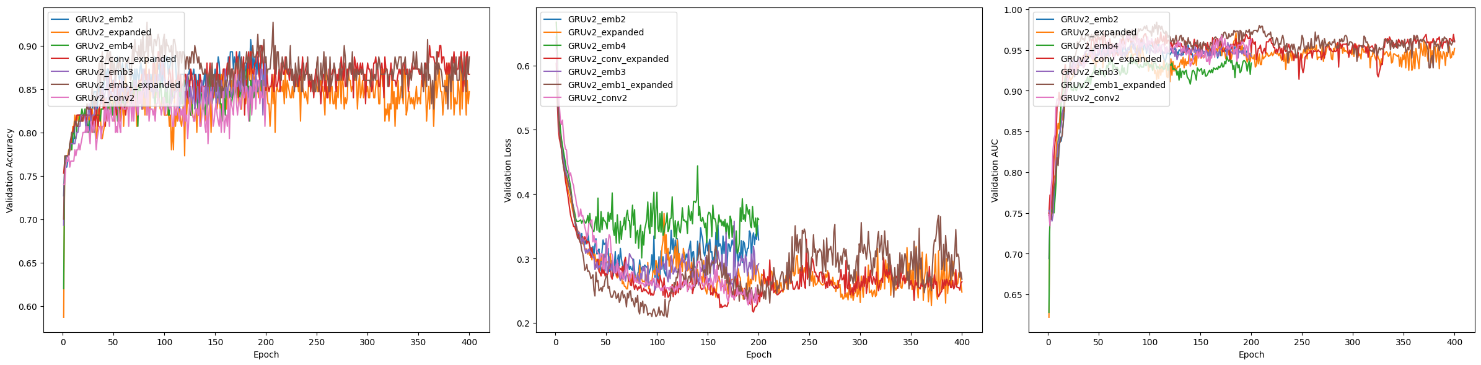


Figure 1. Best GRU performance

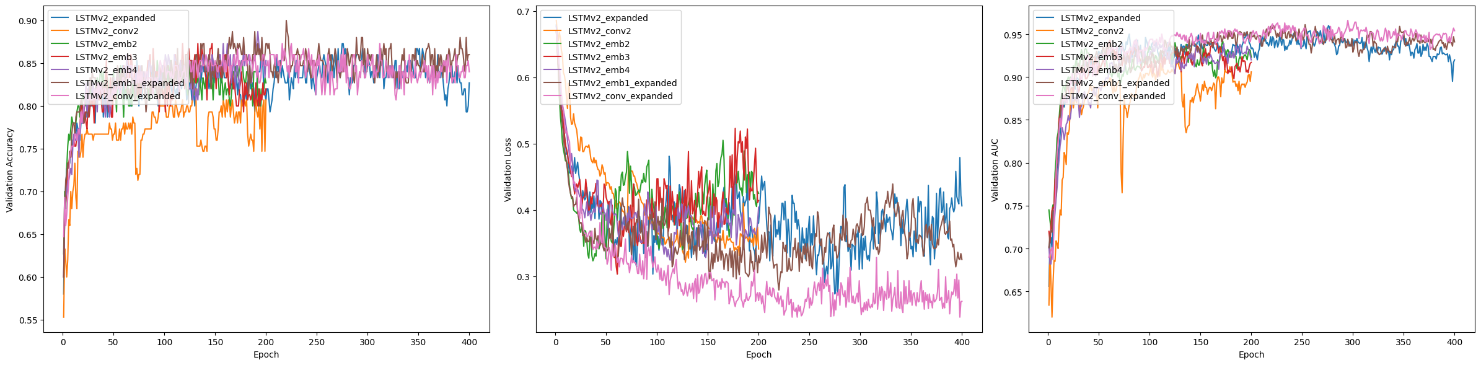


Figure 2. Best LSTM performance

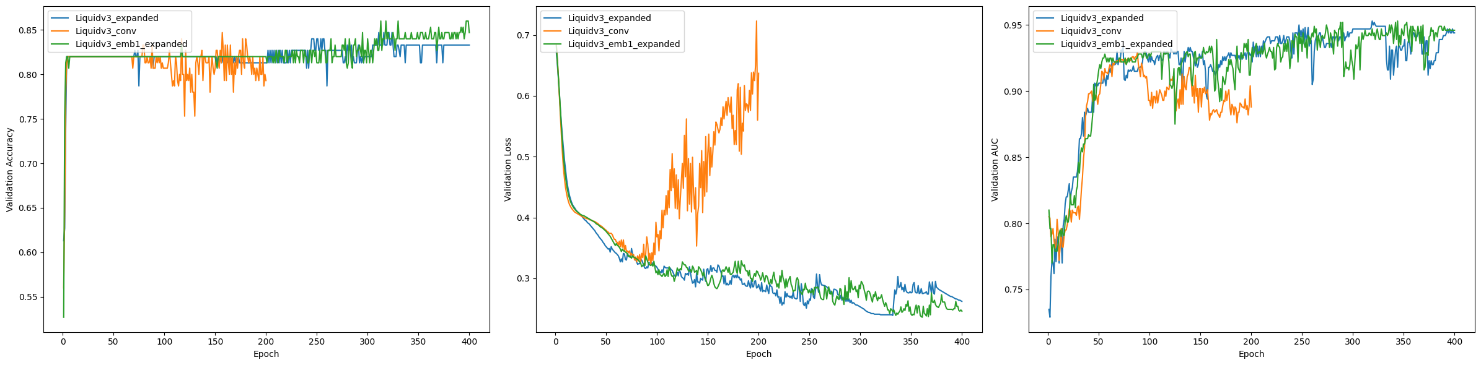


Figure 3. Best LNNS performance

Among all embedding strategies, the lag-1 embedding and the convolutional embedding consistently led to the most significant improvements across models. These two approaches not only increased predictive accuracy but also tended to stabilize the training process, reducing variance and preventing overfitting in several configurations.

The lag-1 embedding provided a minimal but meaningful temporal context, enriching the model’s ability to recognize transitions in brain signal patterns associated with cognitive states. Similarly, the convolutional embedding, by leveraging local temporal filters, effectively captured short-range dependencies that were often missed by raw features alone.

Overall, both of these embedding methods generally enhanced model performance, suggesting that even lightweight temporal transformations of EEG data can provide substantial benefits when predicting cognitive states like confusion.

**Summary of Final Model Performance**

To provide a comprehensive comparison, a summary graph is included that illustrates the final validation metrics – accuracy, AUC, and loss – for the best-performing configuration of each model type: GRU, LSTM, and LSNN. These models were selected based on their performance across all previous experiments, including variations with and without temporal embeddings.

The plotted results clearly show that GRU-based models outperform the others across all metrics. Notably, the GRU model enhanced with lag-1 embedding achieved the best overall results, with accuracy exceeding 0.90 and AUC surpassing 0.95. This performance reflects a substantial improvement over the baseline fully connected model from the referenced paper, which reported an average validation accuracy of only 0.63. It is worth mentioning, however, that the model quickly became overfitted, and other networks gradually became better. Towards the end of the training LSNN with lag-1 embedding even became the best model by performance.

These findings emphasize the importance of sequence-aware architectures and temporal embeddings when working with EEG signals. The GRU’s ability to model temporal dependencies effectively, especially when coupled with lightweight lagged context, appears to be particularly well-suited for confusion detection tasks using frontal EEG data. Liquid neural networks, on the other hand, require some optimizations to train with a comparable speed to its counterparts. However in the long run they show some potential when we deal with EEG data.

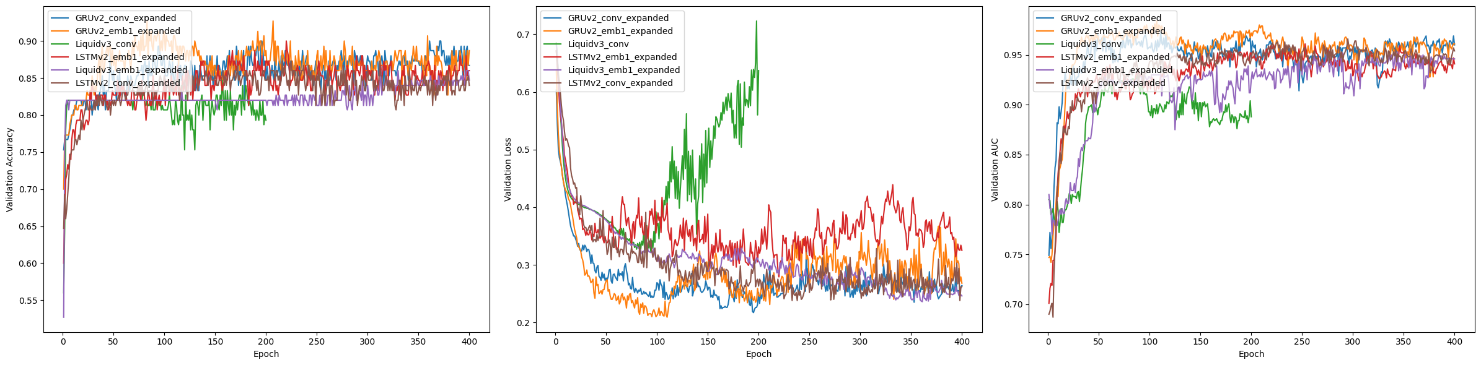


Figure 4. Final model performance

**Model Interpretability with SHAP Analysis**

To explore the interpretability of the trained models, I employed SHAP (SHapley Additive exPlanations) values, a widely adopted method for explaining predictions made by machine learning models. SHAP assigns an importance value to each feature for individual predictions, enabling insight into how specific EEG signals contribute to the final output.

In the SHAP summary plots, each point represents a feature contribution for a single sample. The size of the point corresponds to the absolute magnitude of the SHAP value, indicating the strength of that feature's influence. Red-colored points indicate positive SHAP values, meaning that the feature increased the model’s predicted probability of confusion, while blue-colored points reflect negative contributions, lowering the predicted probability.

To maintain clarity across signal types, the same EEG channel was consistently represented with the same shade of red or blue, allowing visual tracking of individual signal behavior across samples. This consistent color-coding helped highlight the relative importance of each frequency band (e.g., delta, theta, alpha1, etc.) throughout the dataset.

In experiments involving temporal embeddings, such as lagged values, I aggregated the SHAP values of the lagged features with their corresponding original features. This allowed for a more interpretable comparison across models and provided a clearer understanding of the role each signal type played in both its original and embedded forms.

A particularly striking observation from the SHAP summary plots is the presence of specific timestamps where all EEG signal features exhibit notably high importance. At these moments, the absolute SHAP values across all frequency bands increase sharply, suggesting that the model places disproportionate weight on certain temporal segments within the one-minute recording. This implies that cognitive confusion may manifest more strongly at specific points in the video, potentially aligning with moments of conceptual difficulty or topic transitions. While this observation is consistent across multiple samples and models, it warrants further investigation to verify whether these signal spikes consistently correspond to semantically confusing segments in the educational content itself.

Another consistent finding across experiments is the dominant importance of the theta band in predicting confusion. SHAP values frequently show that the theta signal contributes more significantly – both positively and negatively – to the model's decision-making process compared to other frequency bands. This trend is in agreement with prior research, which has linked increased theta activity in the frontal lobe with elevated cognitive load and attentional engagement, particularly in tasks that involve comprehension or problem-solving. The model's reliance on this band therefore provides additional validation for both the data and the model’s alignment with neuroscientific understanding of EEG signal interpretation.

While the GRU and LSTM models yielded interpretable patterns through SHAP analysis, LSNN models unfortunately did not provide useful insights into feature importance. The SHAP value distributions for LSNNs were generally less consistent, lacked clear structure, and often failed to highlight meaningful temporal or spectral patterns. This may be attributed to the architectural sparsity and internal dynamics of neural circuit policies, which, although biologically inspired, can obscure direct attribution of output predictions to individual input features. As a result, while LSNNs demonstrated reasonable predictive performance, their interpretability remains limited, reducing their utility for detailed signal-level or time-resolved analysis in this context.

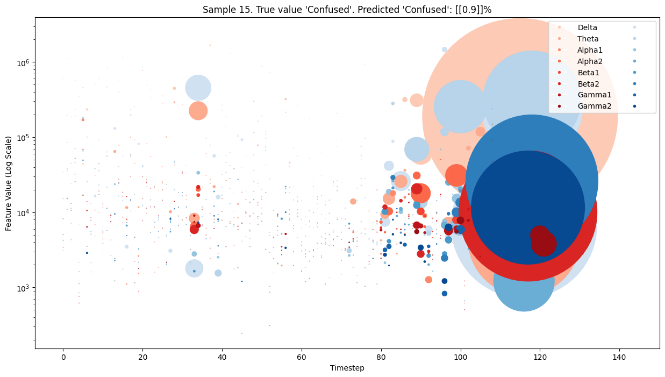
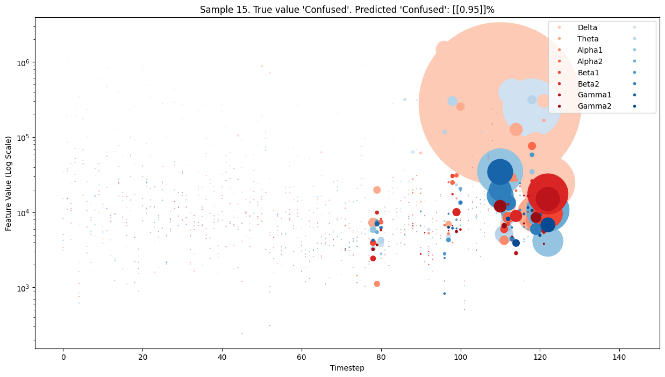
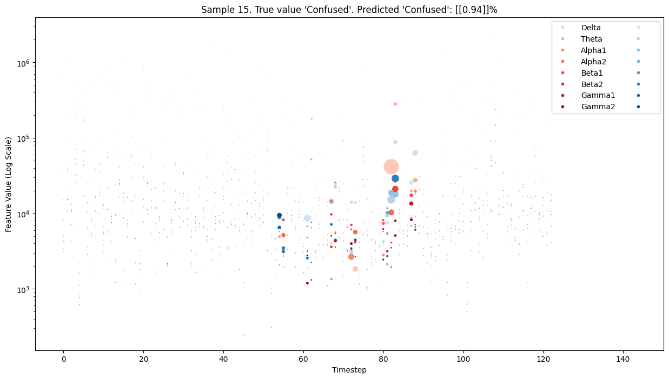
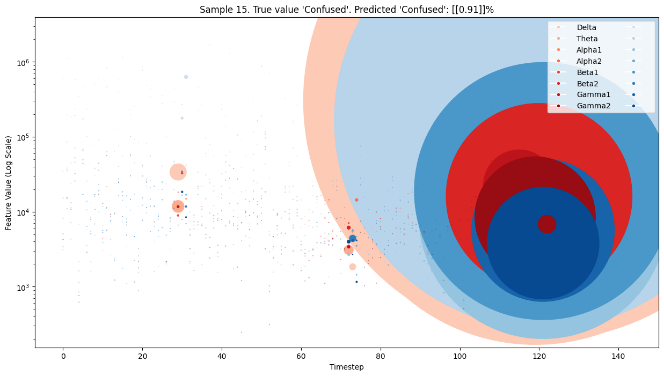
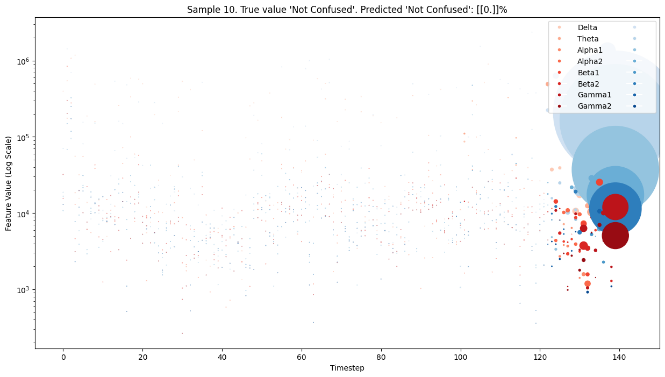
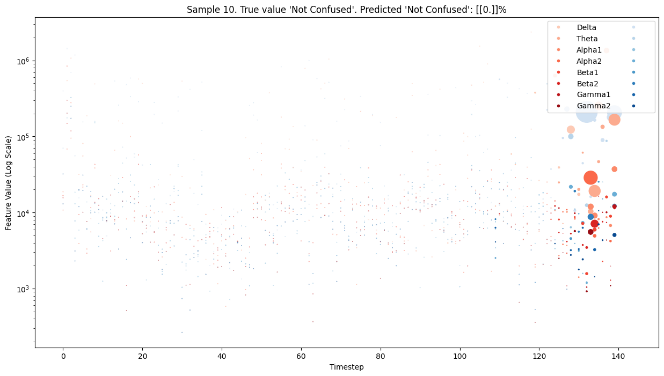
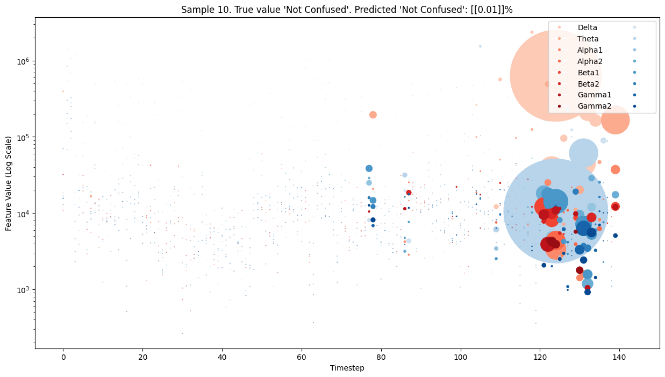
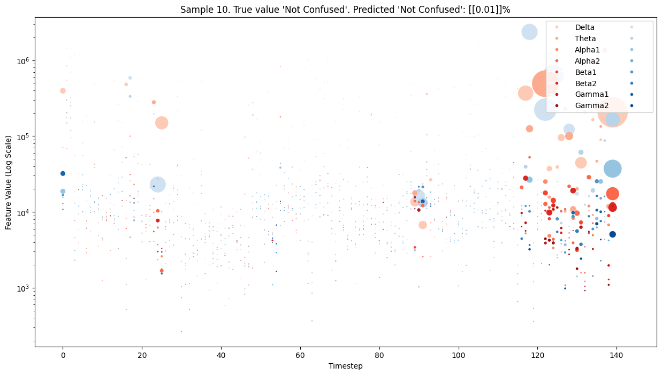
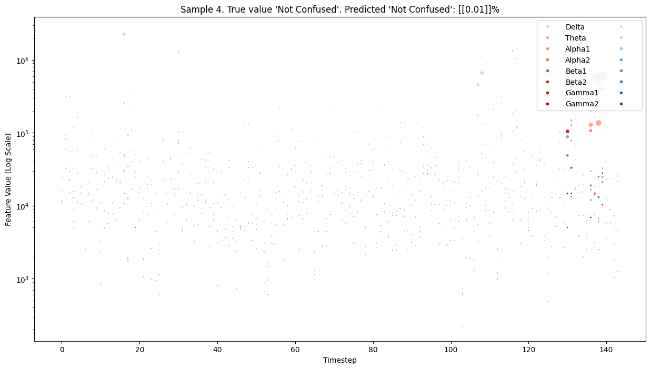
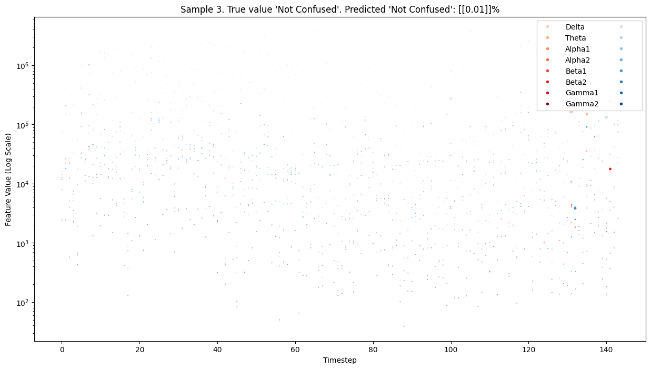
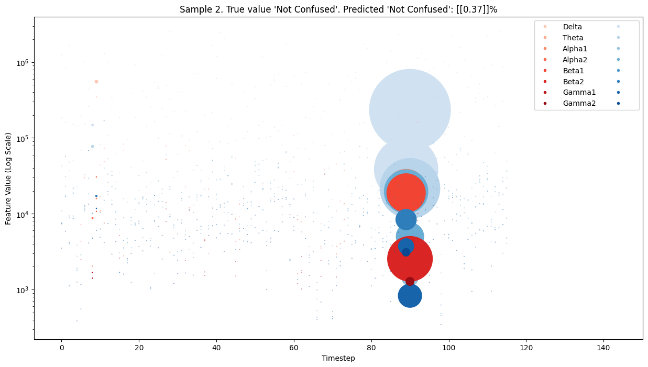
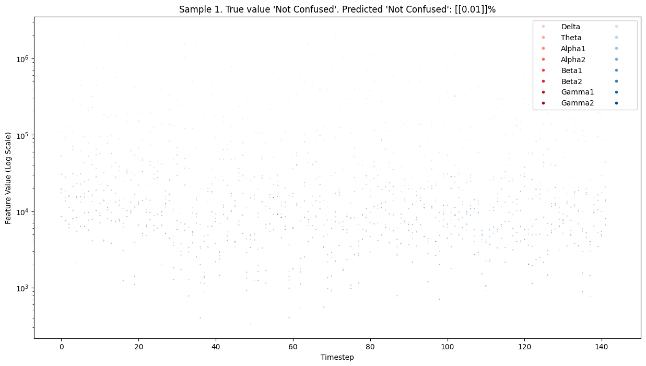
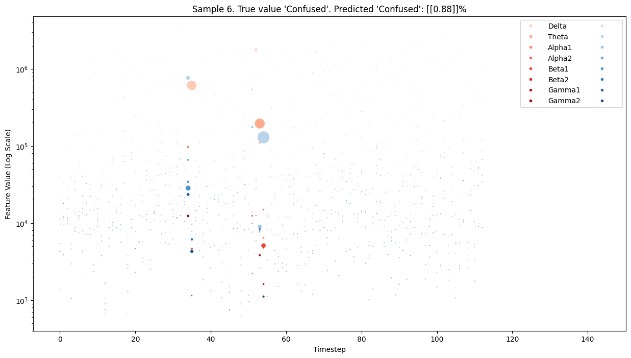
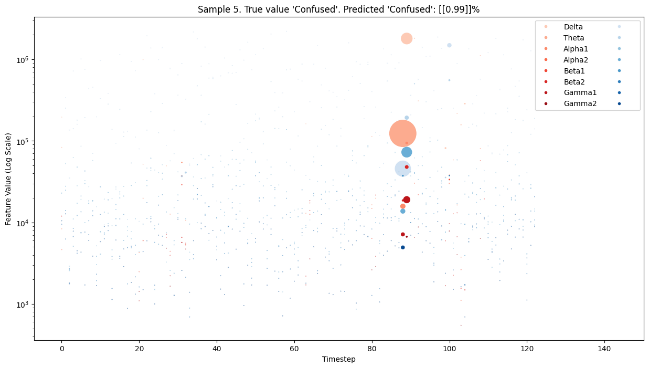


Figure 5. SHAP for Samples 10 and 15 for GRU and LSTM with lag-1 embeddings and convolution





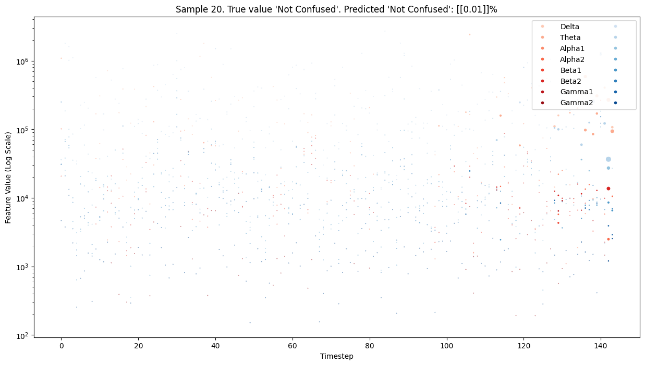
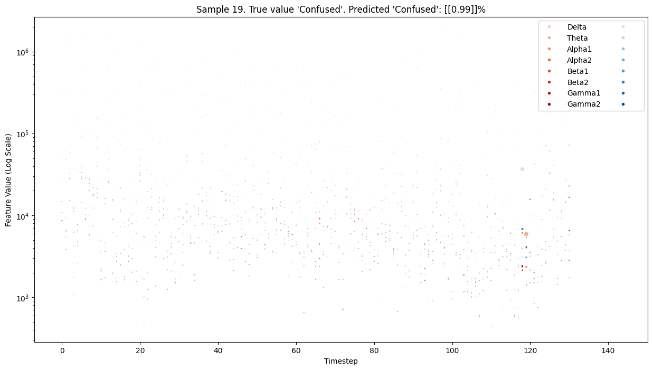
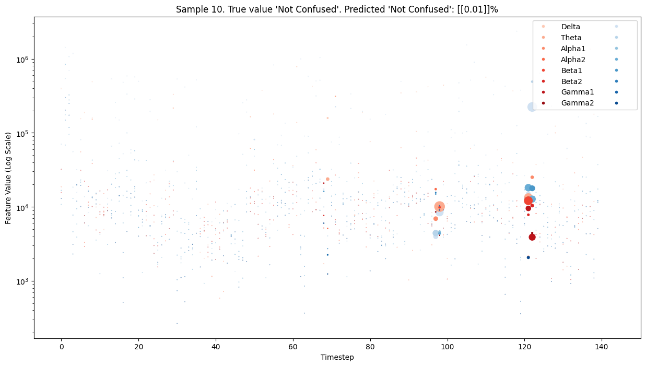
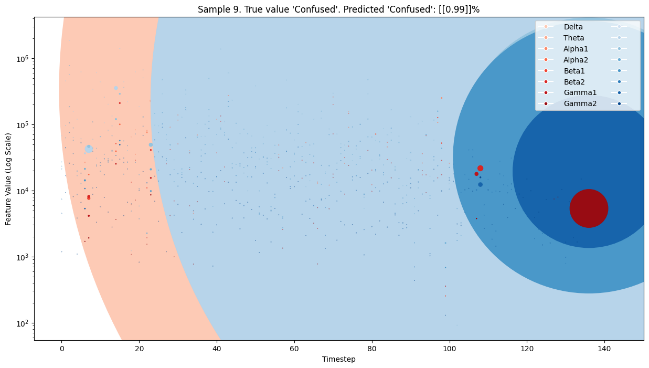


Figure 6. SHAP for GRU with embeddings (best model)

## Conclusions

This thesis has studied how well three deep learning models – Gated Recurrent Units, Long Short-Term Memory networks, and Liquid Neural Networks – can classify confusion states from EEG data. By conducting detailed experiments on EEG recordings collected during controlled learning tasks, we analyzed how each model deals with temporal, noisy, and non-stationary aspects of brain signals.

The results show that all three models are capable of learning useful patterns from EEG time series and can effectively distinguish between cognitive states. GRUs and LSTMs, with their gating mechanisms, performed strongly in modeling long-term dependencies, providing stable and consistent baseline results. Among them, GRU models showed slightly better performance due to their computational efficiency and ability to capture temporal patterns without excessive parameter overhead. On the other hand, Liquid Neural Networks demonstrated clear advantages in adaptability, efficiency, and robustness to subtle changes in brain signals, without requiring extensive retraining. Their biologically inspired architecture and continuous-time modeling make them highly suitable for real-time brain state detection, especially in low-resource or dynamic environments.

However, the study also revealed some important limitations of the proposed approach. One major drawback of using Liquid Neural Networks is the increased computational time during training compared to GRU and LSTM models. Each epoch took approximately five times longer, making the training process significantly more resource-intensive. Moreover, while LNNs demonstrated stable performance during longer training sessions, they were prone to occasional spikes in the loss metric, which might indicate overfitting to specific segments of the data. Additionally, the model interpretability of LNNs was somewhat limited. Unlike GRUs and LSTMs, which allowed for relatively clear SHAP value analysis, LNNs produced less consistent and less interpretable feature importance patterns. This limits their utility in applications where understanding the decision-making process is crucial.

Another limitation is related to the dataset itself. The data was collected from a relatively small number of participants (ten students) and consisted of recordings taken only during educational video sessions. While this controlled environment allowed for systematic analysis, it may not fully represent more diverse real-world scenarios, where cognitive confusion might arise from a broader range of tasks and environments. Consequently, the model's generalizability to other contexts remains uncertain.

Future work could address these limitations by exploring several promising directions. First, improving the training efficiency of LNNs would make them more practical for real-world applications. Optimizing their architecture, possibly by reducing the number of neurons or implementing more efficient training techniques, could help reduce training time. Additionally, exploring hybrid architectures that combine the temporal modeling capabilities of GRUs and LSTMs with the adaptive features of LNNs might result in models that balance accuracy, efficiency, and interpretability.

Another important direction would be integrating attention mechanisms into LNN-based architectures. Attention can help focus the model on the most informative parts of the EEG signals, potentially increasing accuracy while reducing the computational burden. Moreover, using transfer learning could allow the model to generalize better when applied to different EEG datasets or cognitive states beyond confusion, such as stress, concentration, or mental fatigue.

Furthermore, expanding the dataset to include multimodal signals, such as combining EEG with eye-tracking or physiological signals like heart rate variability, could enhance model accuracy by providing richer context. Conducting experiments with a more diverse participant group and varied cognitive tasks would also make the findings more generalizable.

In conclusion, this thesis contributes to the growing field of cognitive state classification by demonstrating that Liquid Neural Networks can offer substantial advantages over traditional RNNs for real-time EEG analysis. Despite some limitations related to training time, interpretability, and data diversity, the promising results suggest that LNNs are worth further exploration and development. As brain-computer interfaces and adaptive educational systems continue to evolve, incorporating more flexible and efficient neural architectures will be essential to improving user experience and cognitive monitoring accuracy.

## References

1. **Ayoub, O., Andreoletti, D., Knapińska, A., Goścień, R., Lechowicz, P., Leidi, T., Giordano, S., Rottondi, C., & Walkowiak, K.** (2024). Liquid Neural Network-based Adaptive Learning vs. Incremental Learning  for Link Load Prediction amid Concept Drift due to Network Failures. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2404.05304>
2. **Bidollahkhani, M., Atasoy, F., & Abdellatef, H.** (2023). LTC-SE: Expanding the potential of Liquid Time-Constant Neural Networks for scalable AI and embedded systems. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2304.08691>
3. **Hasani, R., Lechner, M., Amini, A., Liebenwein, L., Ray, A., Tschaikowski, M., Teschl, G., & Rus, D.** (2022). Closed-form continuous-time neural networks. Nature Machine Intelligence, 4(11), 992–1003. <https://doi.org/10.1038/s42256-022-00556-7>
4. **Hasani, R., Lechner, M., Amini, A., Rus, D., & Grosu, R.** (2021). Liquid Time-constant Networks. Proceedings of the AAAI Conference on Artificial Intelligence, 35(9), 7657-7666. <https://doi.org/10.1609/aaai.v35i9.16936>
5. **Huang, Z., Contreras, L. F. H., Leung, W. H., Yu, L., Truong, N. D., Nikpour, A., & Kavehei, O.** (2024). Efficient Edge-AI models for robust ECG abnormality detection on Resource-Constrained hardware. PubMed. <https://doi.org/10.1007/s12265-024-10504-y>
6. **Lechner, M., Hasani, R., Amini, A., Henzinger, T. A., Rus, D., & Grosu, R.** (2020). Neural circuit policies enabling auditable autonomy. Nature Machine Intelligence, 2(10), 642–652. <https://doi.org/10.1038/s42256-020-00237-3>
7. **Lechner, M., Hasani, R. M., & Grosu, R.** (2018). Neuronal circuit policies. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.1803.08554>
8. **Miah, M. O., Habiba, U., & Kabir, M. F.** (2024). ODL-BCI: Optimal deep learning model for brain-computer interface to classify students' confusion via hyperparameter tuning. *Brain Disorders*, 13, 100121. <https://doi.org/10.1016/j.dscb.2024.100121>
9. **Nerrise, F., Sosanya, A. S., & Neary, P.** (2024). Physics-Informed Calibration of Aeromagnetic Compensation in Magnetic Navigation Systems using Liquid Time-Constant Networks. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2401.09631>
10. **Ni, Z., Yuksel, A. C., Ni, X., Mandel, M. I., & Xie, L.** (2017). Confused or not confused? Disentangling brain activity from EEG data using bidirectional LSTM recurrent neural networks. *ACM-BCB '17: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 241–246. <https://doi.org/10.1145/3107411.3107513>
11. **Nye, L.** (2023). Digital twins for patient care via knowledge graphs and Closed-Form Continuous-Time liquid neural networks. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2307.04772>
12. **Wang, H., Wu, Z., & Xing, E. P.** (2019). Removing confounding factors associated weights in deep neural networks improves the prediction accuracy for healthcare applications. *Pacific Symposium on Biocomputing*. <https://doi.org/10.48550/arXiv.1803.07276>
13. **Wang, H., Li, Y., Hu, X., Yang, Y., Meng, Z., & Chang, K.-M.** (2013). Using EEG to improve massive open online courses feedback interaction. *CEUR Workshop Proceedings*, 1009, 59–66.
14. **Yang, J., Shi, R., & Ni, B.** (2021). MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis. In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 191-195). Nice, France. doi: <https://doi.org/10.1109/ISBI48211.2021.9434062>
15. **Zheng, Z., & Jia, X.** (2023). Complex Mixer for MEDMNIST Classification Decathlon. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2304.10054>
16. **Zhu, F., Wang, X., Huang, C., Jin, R., Yang, Q., Alhammadi, A., Zhang, Z., Yuen, C., & Debbah, M.** (2024). Robust Continuous-Time Beam Tracking with Liquid Neural Network. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2405.00365>