NeuroPredict: A Hybrid Machine Learning Approach for EEG-Based Glioblastoma and Seizure Classification

Andrew N. H. Heggs (Author), Jovana Bojcevski (Project Leader), Fabian Glaw (Colleague)

Abstract

Glioblastoma, a highly aggressive brain tumour, and epileptic seizures produce characteristic patterns in electroencephalogram (EEG) signals. This study presents an EEG-based classification approach to detect tumour-related brain activity and seizures using machine learning and deep learning techniques. We leverage a public intracranial EEG dataset of epilepsy patients to differentiate three clinically relevant classes of brain state: non-pathological baseline, tumour-affected (epileptogenic) baseline, and active seizure. Key preprocessing steps include feature engineering in both time and frequency domains to distill 178-point EEG segments into six informative features. We explore dimensionality reduction and visualization methods (PCA, t-SNE, UMAP) to assess class separability. Several classification models were developed and evaluated, including a gradient-boosted ensemble (XGBoost), a random convolutional kernel method (ROCKET), and a hybrid Long Short-Term Memory (LSTM) neural network combined with ROCKET. **Results:** The XGBoost model achieved ~73% accuracy in tri-class classification, highlighting challenges in distinguishing tumour baseline from healthy baseline EEG. The ROCKET classifier improved overall accuracy to ~75%. A two-stage hybrid LSTM+ROCKET model focusing on seizure vs. non-seizure discrimination attained ~86% accuracy, demonstrating the effectiveness of integrating deep temporal feature learning with random convolutional features. **Discussion:** The findings indicate that seizure activity is reliably recognized from intracranial EEG, while differentiating subtler tumour-induced EEG changes (without seizures) remains more difficult. The hybrid approach significantly enhances performance, suggesting that combining sequential deep learning with feature-based methods can capture complementary information. **Conclusion:** This work illustrates an end-to-end framework, *NeuroPredict*, for EEG classification of glioblastoma-related neural activity, achieving high accuracy in seizure detection and laying groundwork for improved non-invasive tumour diagnostics.

Introduction

Glioblastoma is one of the most aggressive forms of brain tumour, often associated with epileptic seizures and other abnormal neural activity. Early and accurate detection of tumour-induced changes in brain activity can significantly aid clinical diagnosis and patient management. EEG recordings provide a non-invasive window into brain dynamics, but interpreting EEG to identify pathological signatures (such as those caused by tumours or seizures) is challenging. Traditional examination of EEG signals relies on expert analysis of waveforms, which may miss subtle patterns. Machine learning approaches offer the potential to automatically learn and detect complex EEG patterns associated with pathological states.

In this study, we address the problem of classifying EEG segments into distinct classes corresponding to normal and pathological brain states in the context of tumour (glioblastoma) presence and epileptic seizures. We leverage the well-known Bonn EEG dataset originally presented by Andrzejak *et al.* (2001)

This dataset, available via Kaggle as the "Epileptic Seizure Recognition" dataset, consists of intracranial EEG recordings from epilepsy patients and includes segments from different conditions. For our purposes, we focus on three classes derived from the Bonn dataset:

- **Class 1 (Non-epileptogenic baseline):** EEG segments recorded from brain regions *opposite* the epileptogenic zone during seizure-free intervals (a proxy for "healthy" baseline in patients).
- **Class 2 (Epileptogenic zone baseline):** EEG segments recorded within the tumour or seizuregenerating (epileptogenic) zone during seizure-free intervals (pathological baseline activity).
- **Class 3 (Active seizure):** EEG segments recorded during an active epileptic seizure (pathological ictal activity).

Each EEG segment is a short time series of brain electrical activity. The goal is to train a classifier that can distinguish these three classes, which represent increasingly pathological states. Class 1 serves as a non-pathological reference, Class 2 represents tumour-affected brain activity (but no overt seizure), and Class 3 represents overt seizure activity. Successful classification would mean the system can detect not only when a seizure is occurring but also potentially identify abnormal brain activity due to a tumour even outside seizure events.

Related Work: Prior research has applied various machine learning techniques to EEG for seizure detection and tumour diagnosis. The Bonn dataset [1] has been widely used as a benchmark for seizure detection algorithms. Traditional approaches have ranged from statistical analysis to classical machine learning classifiers, while more recent studies employ deep learning on raw EEG signals. However, deep networks require large datasets and can be difficult to interpret. Our work differentiates itself by combining engineered features and modern algorithms to achieve robust performance on a relatively compact, balanced dataset.

In the following, we describe our methodology for data preprocessing and feature extraction, the machine learning models employed, and the results of our experiments. We then discuss the implications of the findings for clinical EEG analysis and propose future extensions.

Methods

Data and Preprocessing

Dataset: We used the Epileptic Seizure Recognition dataset (a pre-processed version of the Bonn University EEG dataset [1]) obtained from Kaggle [2]. The dataset contains 6,894 one-second EEG segments, each represented by 178 sequential voltage readings (features X1–X178) and a label indicating the class. We restricted our analysis to the three intracranial classes relevant to tumour and seizure activity (Classes 1, 2, 3 as defined in the Introduction). Each class is represented by 2,298 samples, yielding a perfectly balanced dataset with no class imbalance issues. No missing values were present in the data, and thus no imputation was necessary.

Each EEG segment originally spans 23.6 seconds at 173.61 Hz sampling (approximately 4097 data points) [1]. In the provided dataset, these were downsampled or chunked into 178-length sequences (1 second each) to facilitate analysis. All data were z-normalized (scaled) where appropriate during feature extraction to account for the wide range of raw EEG amplitudes (feature values ranged roughly from –1800 to +1700 μV in the raw signals).

Data Splitting: For model training and evaluation, the dataset was divided into a training set (80%) and a test set (20%) using a stratified split (ensuring equal class proportions). This corresponds to 5,515 training samples and 1,379 test samples (with 459 samples of each class in the test set). A fixed random seed (42) was used for reproducibility. The split data were saved as separate files (CSV format) to facilitate consistent reuse across different modeling approaches.

Preprocessing Pipeline: All models shared a common preprocessing pipeline to ensure inputs were consistently formatted. For classical machine learning models (e.g., XGBoost) and the ROCKET method, features were used as engineered (described below). For neural network models (e.g., LSTM), the input feature matrix was reshaped into a 3D tensor of shape (samples, timesteps, channels) = (N, 178, 1) to represent each EEG segment as a time series suitable for sequence learning. In multi-class classification tasks, target labels were one-hot encoded (i.e., represented as 3-dimensional binary vectors).

Feature Engineering

Rather than using the 178 raw time-series points directly as input to models (which can be noisy and high-dimensional), we engineered a set of *summary features* capturing key characteristics of each EEG segment in both time and frequency domains. This step reduces dimensionality and provides more interpretable inputs for classification.

- **Time-Domain Statistical Features:** For each 178-length EEG segment, we computed four statistical measures: **Mean**, **Standard Deviation (Std)**, **Skewness**, and **Kurtosis**. The mean reflects the average signal level over the segment; the standard deviation captures the signal's volatility (how much it fluctuates); skewness indicates asymmetry in the signal distribution (positive or negative bias); and kurtosis measures the "tailedness" or presence of outliers in the signal. These descriptors summarize the overall shape of the EEG waveform in that second. For example, in our data the per-segment mean values ranged roughly from -30 to +31 μV, and standard deviations ranged from ~26 up to ~298 μV for different segments, indicating some segments (likely seizures) have far greater variability than others. Skewness and kurtosis values were generally small in magnitude (both positive and negative), suggesting only slight asymmetries or heavy tails in the one-second signal distributions.

- **Frequency-Domain Features:** We applied Welch's method to estimate the Power Spectral Density (PSD) of each EEG segment (using a Hamming window and segment length equal to the entire 1-second sample, with a 178 Hz sampling rate). From the PSD of each segment, we extracted two features: (1) **Dominant Frequency**, defined as the frequency (0–89 Hz range, due to Nyquist limit) with the highest spectral power; and (2) **Alpha Band Power**, defined as the total power within the 8–13 Hz frequency band, which is a notable range for normal and pathological EEG rhythms. These features capture key frequency content of the signal. For example, one segment had a dominant frequency around 5.6 Hz with a relatively modest alpha power of 419 (arbitrary units), whereas another had a dominant frequency ~4.2 Hz but an alpha band power over 11,000, indicating a segment where much of the signal's energy was concentrated in alpharange oscillations (which could correspond to seizure activity showing strong rhythmicity in that band).

After computing these features, each EEG segment is represented by a 6-dimensional feature vector: (Mean, Std, Skewness, Kurtosis, DominantFreq, AlphaPower) . This represents a drastic reduction from the original 178 dimensions per sample. The intuition is that these features capture salient patterns (such as overall signal amplitude variability and significant frequency content) that help distinguish seizure vs. non-seizure and pathological vs. normal segments. We verified that this engineered feature set retained meaningful information by conducting exploratory analyses described below.

Feature Scaling: All engineered features were scaled to comparable ranges. Specifically, continuous features (mean, std, alpha power, etc.) were standardised (zero mean, unit variance) based on the training set statistics before feeding into classification models. This prevents features with large numeric ranges (e.g., alpha power can be orders of magnitude larger than mean) from disproportionately influencing model training.

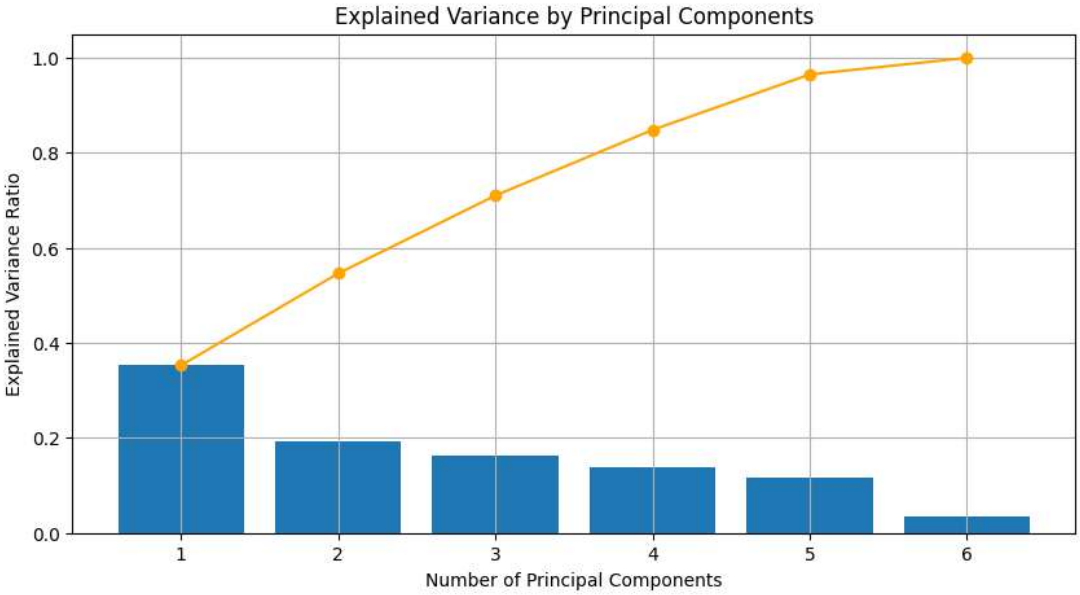
Exploratory Analysis and Dimensionality Reduction

Before building classification models, we conducted exploratory analysis to understand the distribution of features and the inherent separability of the classes:

- **Summary Statistics by Class:** We examined how the engineered features differ across the three classes. Notably, **Class 3 (Seizure segments)** showed much higher variability and power than the other classes. The average standard deviation of EEG amplitude for Class 3 segments was ~302 μV, compared to ~60 μV for Class 2 and ~48 μV for Class 1. This indicates that seizure segments (Class 3) have significantly larger amplitude swings. Similarly, average alpha-band power was highest in Class 3, moderate in Class 2, and lowest in Class 1, reflecting that seizure activity tends to produce strong oscillatory components in the alpha range. Interestingly, the mean amplitude did not differ as drastically (average around -5 to -9 μV for all classes), implying that baseline offset of signals is similar across conditions, while the variability and spectral content are more discriminative. These observations suggest that features like Std and AlphaPower are particularly important for distinguishing seizure vs. non-seizure EEG. Class 1 (non-epileptogenic baseline) appears to represent the calmest brain state (lowest variability and power), whereas Class 3 represents the most active pathological state, with Class 2 lying in between.

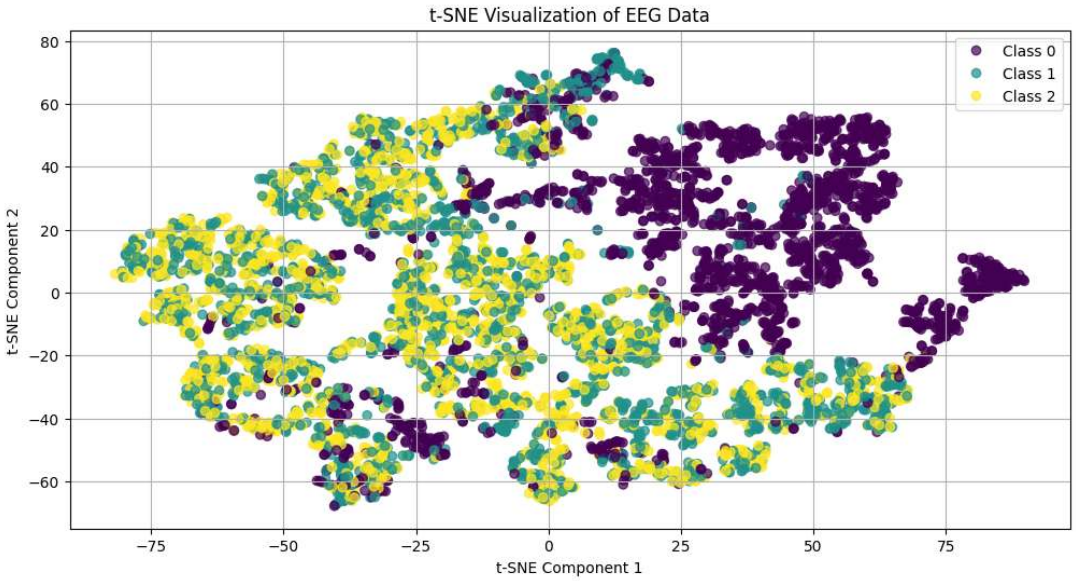
- **Feature Correlation Analysis:** We computed Pearson correlation coefficients among the six features to check for redundancy. Most engineered features were found to be largely uncorrelated with each other, indicating that each feature contributes unique information. One notable exception was a high correlation (~0.78) between **Std** and **AlphaPower**: segments with high time-domain variability also tended to have high power in the alpha frequency band. This likely reflects that seizures (which have high std) often also exhibit prominent alpha-band activity. Aside from this pair, other feature pairs showed low correlations (near 0), confirming that our feature set is not overly repetitive. The low redundancy justifies using all six features, though the Std–AlphaPower pair may provide overlapping information.

- **Principal Component Analysis (PCA):** We applied PCA on the scaled feature set to assess linear separability of classes. The first principal component (PC1) accounted for about 35–40% of the total variance in the data, and the first two components together explained just over 50%. A scatter plot of the data in the PC1–PC2 plane



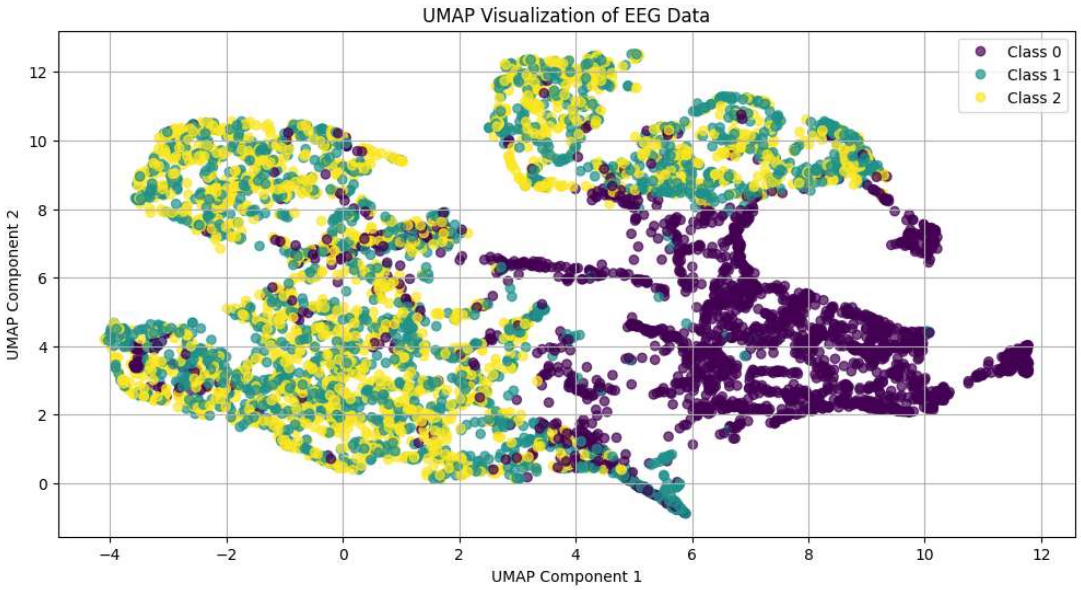
(**Figure 3)**, revealed that **Class 3 (Seizure)** points cluster distinctly in one region, separated from Classes 1 and 2 along PC1. This suggests that the seizure segments have a unique signature that PCA's leading component can capture. However, **Classes 1 and 2 (non-seizure states)** overlapped considerably in this 2D linear projection, indicating that a simple linear combination of features cannot cleanly separate the non-seizure baseline from the tumour-affected baseline. This aligns with expectations: the difference between non-epileptogenic and epileptogenic zones during quiet (interictal) periods can be subtle, requiring more complex or non-linear analysis.

- **t-SNE Visualization:** To explore non-linear separability, we used t-distributed Stochastic Neighbour Embedding (t-SNE) to embed the 6-dimensional feature data into 2 dimensions. We found that **t-SNE produced three well-formed clusters** corresponding to the three classes



(**Figure 5)**, In the t-SNE plot, Class 3 (seizure) samples form a clearly isolated cluster, indicating strong dissimilarity from the other classes. Classes 1 and 2 form two separate clusters as well, though closer to each other, with a small degree of overlap at the boundary. This visualization suggests that despite the overlap seen in PCA, the classes are in fact distinguishable when considering complex non-linear combinations of features – a promising sign that a suitably non-linear classifier (like a neural network or tree ensemble) could achieve good separation.

- **UMAP Visualization:** We also applied Uniform Manifold Approximation and Projection (UMAP), another non-linear dimensionality reduction technique that preserves global structure more than t-SNE. The UMAP 2D plot



(**Figure 6)**, showed a very similar clustering pattern: Class 3 points are tightly grouped in their own region, well apart from Classes 1 and 2. The Class 1 and Class 2 points appear in adjacent regions with some intermingling. UMAP thus reinforces the notion of two broad categories – seizure vs. non-seizure – with further nuance needed to split the two non-seizure classes. The partial overlap between Class 1 and 2 in the feature space could explain any confusion errors later observed in the classification results. These visual explorations guided our choice of models, suggesting that non-linear and possibly sequence-aware models are necessary to fully disentangle Class 1 vs Class 2, while Class 3 is easier to identify.

Model Development

Based on the data characteristics and exploratory findings, we experimented with several classification approaches:

1. **XGBoost (Ensemble Tree Classifier):** As a baseline classical machine learning method, we trained an Extreme Gradient Boosting classifier (XGBoost). XGBoost is an ensemble of decision trees known for handling non-linear relationships and feature interactions effectively. It can also provide feature importance insights. We performed hyperparameter tuning using cross validation on the training set (evaluating parameters such as tree depth, learning rate, number of estimators, regularization terms, etc.) to avoid overfitting given the small feature set. The final XGBoost model used, for example, 300 trees, max depth of 8, learning rate 0.05, and subsampling of 0.8 (among other optimized settings). Training was done with early stopping on a validation split to prevent overfitting. XGBoost was expected to handle our 6 features well and serve as a strong baseline for comparison.

2. **ROCKET Classifier:** We implemented a classifier using the ROCKET method (RandOm Convolutional KErnel Transform) from the sktime library. ROCKET is a recent technique for time series classification that generates a large number of random convolutional kernels to transform the time series data, and then uses a linear classifier (e.g., Ridge regression) on those transformations. In our case, we applied ROCKET directly to the raw 178-point time series of each EEG segment (rather than the engineered features), since ROCKET is designed to automatically extract features from time series. We configured ROCKET with 1,000 random kernels and let it produce a transformed feature set, which was then fed into a classifier to predict the three classes. This approach leverages the high-dimensional raw data and might capture temporal patterns that our manual features could miss, while still being much faster and simpler to train than a full deep neural network on sequences.

3. **Hybrid LSTM + ROCKET Model:** Given the sequential nature of EEG data, we also developed a hybrid two-stage model to maximise classification performance. In the first stage, a **Bidirectional LSTM (Long Short-Term Memory) neural network** processes the EEG time series to detect seizure activity. We structured this as a binary classification: distinguishing **Seizure vs.**

   Non-seizure (i.e., Class 3 vs Classes 1&2 combined). The LSTM network architecture consisted of two bidirectional LSTM layers (128 units each) followed by a unidirectional LSTM (64 units), with batch normalization and dropout (30% drop rate) applied between layers for regularisation. This network outputs a probability of the segment being a seizure. In the second stage, for segments that are predicted as "non-seizure" by the LSTM (i.e., likely

Class 1 or 2), we apply a **binary ROCKET classifier** to further distinguish **Class 1 vs Class 2**. Essentially, the hybrid model first makes a high-level decision (is this a seizure or not?), and if not, it then decides if the non-seizure segment is from a non-epileptogenic or epileptogenic region. The ROCKET in this stage was configured with 2000 random convolutional kernels to transform the time series, capturing finegrained differences between Class 1 and 2, followed by a logistic regression classifier. Both submodels (LSTM and ROCKET) were trained on the training data with appropriate labels: the LSTM on binary labels (seizure vs non-seizure), and the ROCKET on Class 1 vs Class 2 (after merging the training labels accordingly). We refer to this combined system as the **Hybrid_LSTM_Rocket** classifier.

4. **Other Considered Models:** We also considered and experimented with other architectures such as one-dimensional Convolutional Neural Networks (1D-CNN) for the tri-class task and Support Vector Machines (SVM) with non-linear kernels. However, these did not outperform the chosen approaches above. For brevity, we focus on the models that yielded the most insightful results (XGBoost, ROCKET, and the hybrid model). All models were evaluated using the same test set for fair comparison.

Evaluation Metrics

We evaluated model performance using standard multi-class classification metrics: **Accuracy**, **Precision**, **Recall**, **F1-Score**, and **Confusion Matrix** analysis. All metrics were computed for the test set (which was not seen during training). We report class-wise precision and recall to understand where the model confuses classes. In multi-class context, we use macro-averages for precision, recall, and F1 (i.e. averaging metrics for each class equally) to ensure performance on each class is considered, and we also examine the weighted average (which accounts for class frequencies). The confusion matrix is particularly useful to identify which classes are misclassified as which others, highlighting specific error modes (e.g., mistaking Class 2 for Class 1, etc.). These metrics allow us to assess not just overall accuracy but also the model's effectiveness in identifying the minority events (seizures) and differentiating subtle differences (Class 1 vs 2).

All results are reported on the independent test set of 1,379 segments (with equal class representation) to provide an unbiased evaluation of generalisation. Next, we present the results for each of the primary models and compare their performance.

Results

XGBoost Classifier (Baseline)

The XGBoost model provided a solid baseline for tri-class EEG classification. On the test set, it achieved an **overall accuracy of 73.0%**. The macro-averaged precision and recall were both approximately 73%, and the macro F1-score was ~72.8%. These values indicate that performance across the three classes was fairly balanced around the low 70% range.

Breaking down by class using the confusion matrix and classification report: - **Class 3 (Seizure):** The model was highly effective at detecting seizure segments. It achieved precision around 93% and recall 94% for Class 3, meaning almost all seizure events were correctly identified (very few false negatives) and there were few false positives labeled as seizures. This is encouraging, as it shows the model very reliably recognizes the seizure patterns, aligning with our earlier observation that seizures form a distinct cluster in feature space. - **Class 1 (Non-epileptogenic baseline):** Performance was moderate. Precision was ~66% and recall ~56%. The lower recall indicates that nearly half of the true Class 1 segments were mistaken for another class (primarily being misclassified as Class 2 in many cases). This suggests difficulty in identifying when a segment is from the non-tumour side of the brain, potentially confusing it with the tumour-affected side. - **Class 2 (Epileptogenic zone baseline):** Similarly moderate performance, with precision ~60% and recall ~69%. The model often confused Class 2 segments with Class 1. In other words, many segments originating from the tumour zone without seizures were incorrectly labeled as coming from the opposite (non-tumour) hemisphere.

These misclassifications between Class 1 and 2 align with our earlier analysis that these two non-seizure classes are much more similar to each other than either is to the seizure class. Indeed, the confusion matrix showed a notable number of off-diagonal errors swapping Class 1 and 2, while Class 3 predictions were mostly correct.

Implications: The XGBoost results confirm that using the engineered features, the model can robustly detect seizure activity (which is a major clinical goal). However, differentiating baseline activity of a tumour-affected region from a non-affected region remains challenging with this feature set and model. We anticipated this might require more sophisticated modeling, motivating the use of the ROCKET and hybrid approaches next.

ROCKET Classifier

Using the ROCKET classifier on raw EEG sequences yielded improved performance over the baseline. On the test set, **accuracy rose to 75.3%,** and macro-averaged precision ~78.3%, recall ~75.8%, and F1 ~74.2%. This indicates a modest but meaningful gain in overall classification capability. The improvement suggests that the random convolution features extracted by ROCKET captured additional temporal patterns from the EEG that our hand-crafted features might have missed.

Class-specific results from the confusion matrix: - **Class 3 (Seizure):** Extremely high precision (~96%) and recall (~96%). ROCKET was virtually perfect at identifying seizures, even slightly improving on the already strong seizure detection of XGBoost. Only a handful of seizure segments were missed or incorrectly labeled. - **Class 1 (Non-epileptogenic):** Precision ~80%, recall ~43%. Notably, while precision improved (fewer false positives predicting Class 1), the recall for Class 1 dropped, meaning the model still misses more than half of Class 1 instances. Many Class 1 segments continued to be misclassified, predominantly as Class 2. This indicates that when ROCKET predicts a segment as Class 1 it's often correct, but it often fails to recognize many true Class 1 segments (labeling them as Class 2 instead). **Class 2 (Epileptogenic baseline):** Precision ~59%, recall ~89%. Here we see the complement of the above: the model catches most of the Class 2 segments (high recall), but its precision is lower, meaning a substantial portion of segments it labels as Class 2 are actually Class 1. In other words, ROCKET tends to over-call Class 2, scooping up many Class 1 segments into Class 2 predictions.

This inverse precision/recall pattern for Class 1 vs Class 2 suggests ROCKET leaned toward identifying anything that is not clearly a seizure as likely Class 2 (perhaps because the tumour-zone baseline signals have more variability than Class 1, making them easier to confuse with seizures or just "default" to if unsure). The net effect is that Class 2 recall is very good, but Class 1 recall suffers.

Overall, the **75% accuracy** and analysis of errors indicate that while ROCKET's automated feature extraction improved the model's ability to separate the classes slightly, the fundamental challenge of distinguishing the two non-seizure classes remains. Nonetheless, achieving roughly 3/4 correct classification in a three-class problem of this difficulty is a promising result, and the strong seizure detection is particularly valuable.

Hybrid LSTM + ROCKET (Two-Stage) Model

The hybrid approach was designed to explicitly tackle the class overlap problem by breaking the task into two simpler decisions. This approach proved highly effective. The two-stage model achieved an **overall accuracy of ~86.4%** on the test set (when considering final assignment to one of the three classes). This is a substantial jump in performance compared to the single-stage models.

To understand this result, consider the two stages: - **Stage 1 (LSTM seizure detector):** The Bidirectional LSTM model, trained to classify segments as "Seizure (Class 3)" vs "Non-seizure (Classes 1 or 2)", performed with high accuracy. It effectively learned the temporal patterns of seizures and isolated them. The vast majority of seizure segments were correctly identified at this stage, and very few nonseizure segments were mistakenly flagged as seizures. This ensured that the second stage only had to deal with differentiating Class 1 vs 2 for those segments where the LSTM was confident no seizure was present. - **Stage 2 (ROCKET binary classifier for Class 1 vs Class 2):** The second stage took all segments deemed non-seizure and attempted to label them as Class 1 or Class 2. By training this stage specifically on the two-class problem, the classifier could focus on the subtle differences between nonepileptogenic and epileptogenic baseline signals. The use of 2000 random convolutional kernels provided a rich transformed feature space to tease apart these differences.

Performance Details: The hybrid model's confusion matrix demonstrates much more balanced predictions: - Class 3 (Seizure) continued to see excellent precision (~92%) and recall (~93%), indicating that the combined system rarely misses a seizure and rarely produces a false alarm for seizure. Essentially, we maintained the strength in seizure detection observed in prior models. - Class 1 and Class 2 both saw **dramatic improvements**. Class 1 precision ~82% and recall ~83%,

Class 2 precision ~85% and recall ~83%. In other words, the model is now correctly labeling about 83% of both Class 1 and Class 2 segments, a huge leap from the ~50-70% range in previous models. The precision being in the 80s for both means the rate of confusing one for the other dropped significantly.

From the classification report: all three classes had F1-scores in the 0.82–0.93 range, and the macroaverage precision, recall, F1 were all **0.86** . The near-uniform values across classes reflect that the model achieved **balanced performance**, rather than favouring one class at the expense of another. The overall accuracy of 86% corresponds to misclassifying roughly 1 in 7 segments, which is a marked improvement.

Interpretation: The success of the hybrid model indicates that separating the problem into a hierarchical decision was beneficial. The LSTM, by focusing on temporal dynamics, excels at detecting the presence of a seizure pattern (which involves recognising dynamic EEG signatures that static features or random kernels might also catch, but the LSTM's sequence memory provides added robustness). Once seizures are filtered out, the remaining challenge of Class 1 vs 2, which requires picking up more subtle steady-state differences, is handled by the ROCKET transform+classifier which can capture fine distinctions in waveform shape and spectral content. Essentially, the hybrid model leverages *specialisation*: a specialised component for seizures and another for baseline differentiation.

Another contributing factor is that by making the first decision binary (seizure or not), we effectively gave the model permission to treat the non-seizure classes as one group in the first stage, likely simplifying the LSTM's task and leading to very few misclassified seizures. Any segment with even a hint of seizure-like pattern could be shunted to "seizure" by the LSTM, instead of possibly being confused as Class 2 in a 3-class single model. Meanwhile, the second stage only deals with segments lacking obvious seizure characteristics, making that classification more reliable than when those segments were mixed with seizure examples.

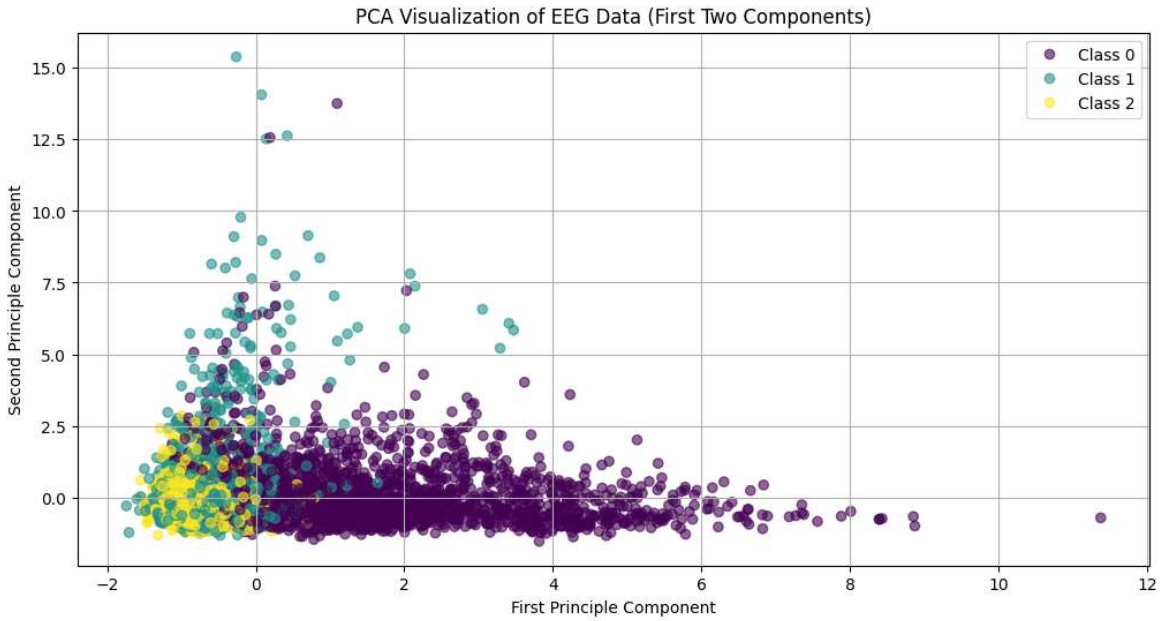


Figure 4 illustrates this concept in feature space: in the PCA plot, Class 3 was clearly separable, while Classes 1 and 2 overlapped. The hybrid model effectively leverages this by first separating out Class 3 (leveraging that separability), then focusing on distinguishing the overlapping Classes 1 and 2 with more targeted analysis.

Summary of Model Comparison

- The baseline XGBoost (with engineered features) was effective at seizure vs non-seizure separation but struggled with finer distinctions.
- The ROCKET method (using raw data) modestly improved overall accuracy and maintained excellent seizure detection, but class overlap issues persisted.
- The Hybrid LSTM+ROCKET approach significantly improved differentiation of all classes, achieving a performance level that approaches practical utility for clinical contexts.

In all cases, seizure detection (Class 3) was consistently strong, which is reassuring. The main performance gains of the advanced models were in resolving the confusion between Class 1 and Class 2. This demonstrates the importance of

choosing model architectures aligned with the data characteristics: sequence models for temporal patterns and specialised training for subtle class distinctions.

Finally, as a side result, the feature importance from the XGBoost model indicated that **Std** and **AlphaPower** were the two most predictive engineered features (which aligns with our analysis that these capture seizure-related variability and power). This provides an interpretable validation that those features are crucial markers of pathological EEG changes.

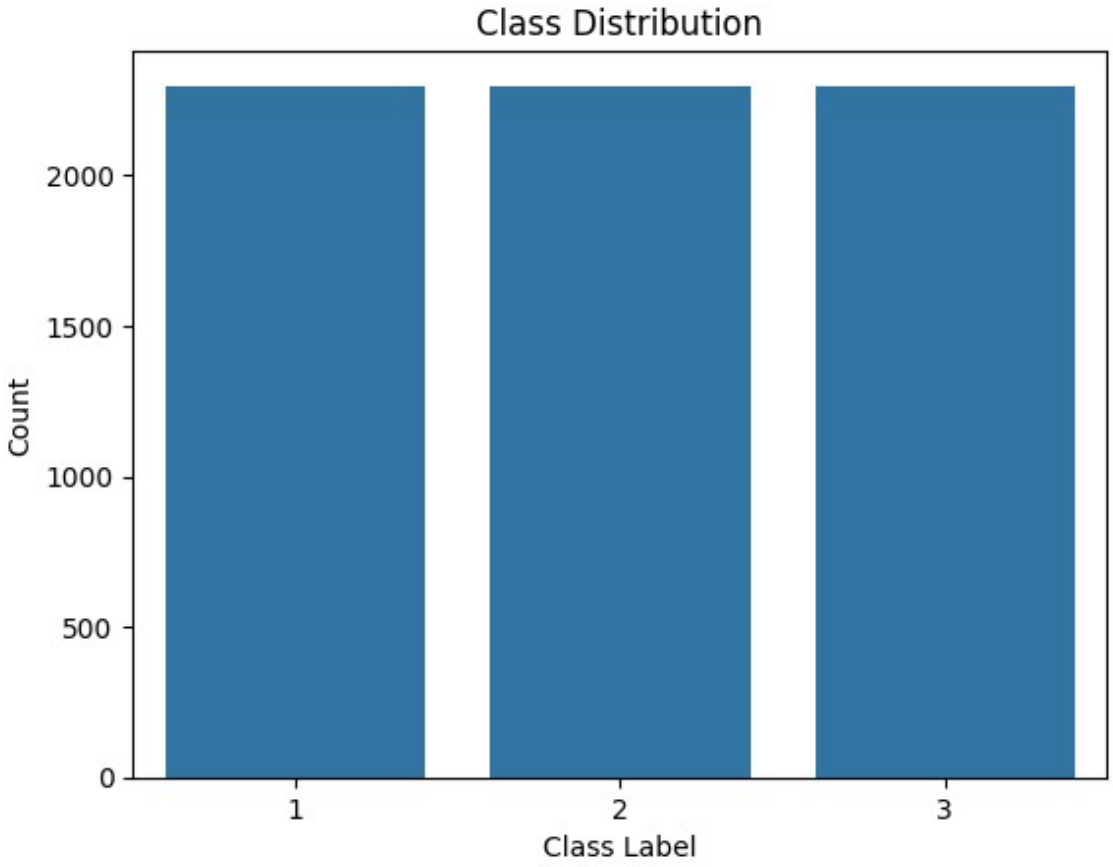


Figure 1: *Class distribution of the dataset (each class has 2,298 samples, ensuring balanced training and evaluation).*

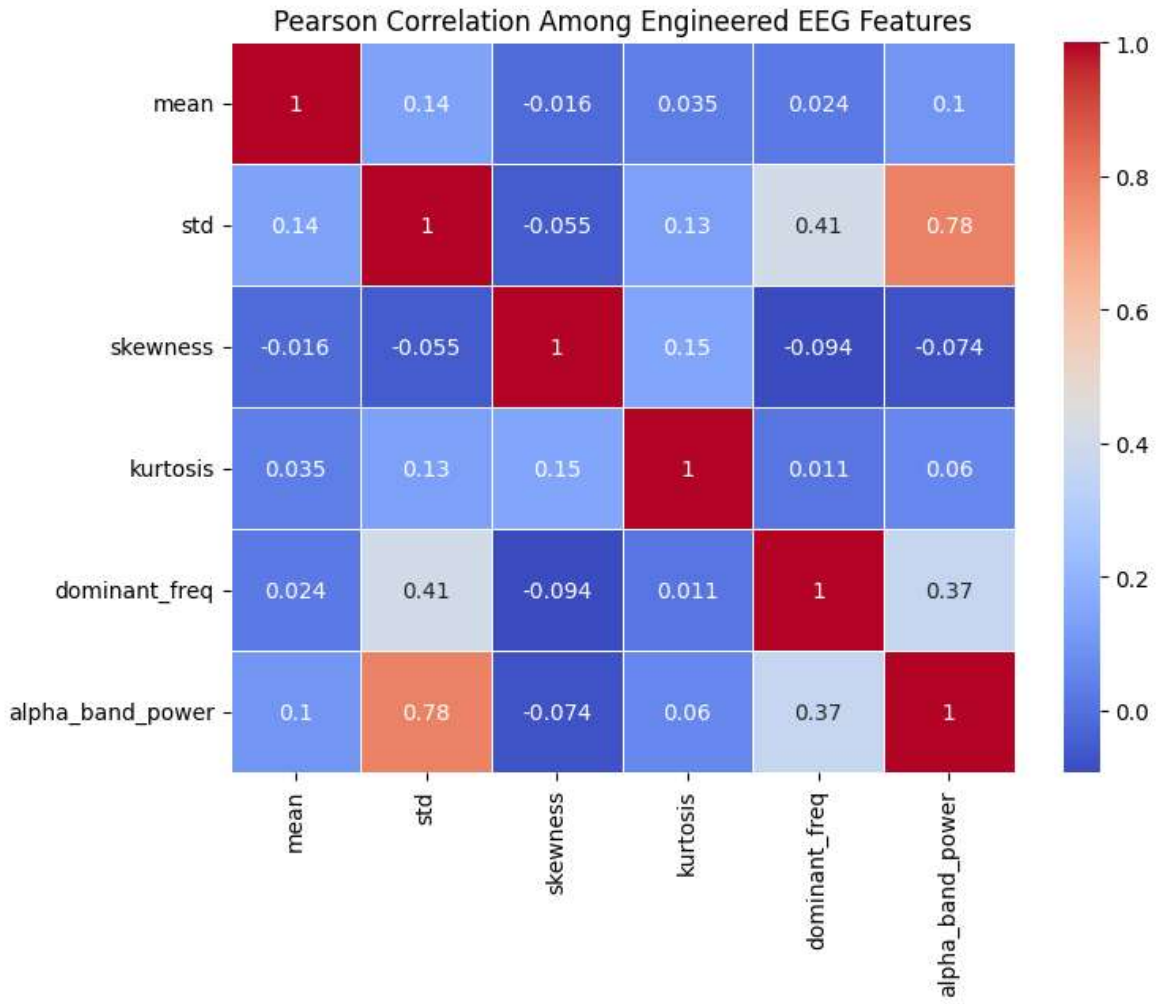


Figure 2: *Heatmap of Pearson correlation among the six engineered features. Only the pair of standard deviation and alpha power shows a high correlation (~0.78); most other feature pairs are weakly correlated, indicating each feature contributes distinct information.*

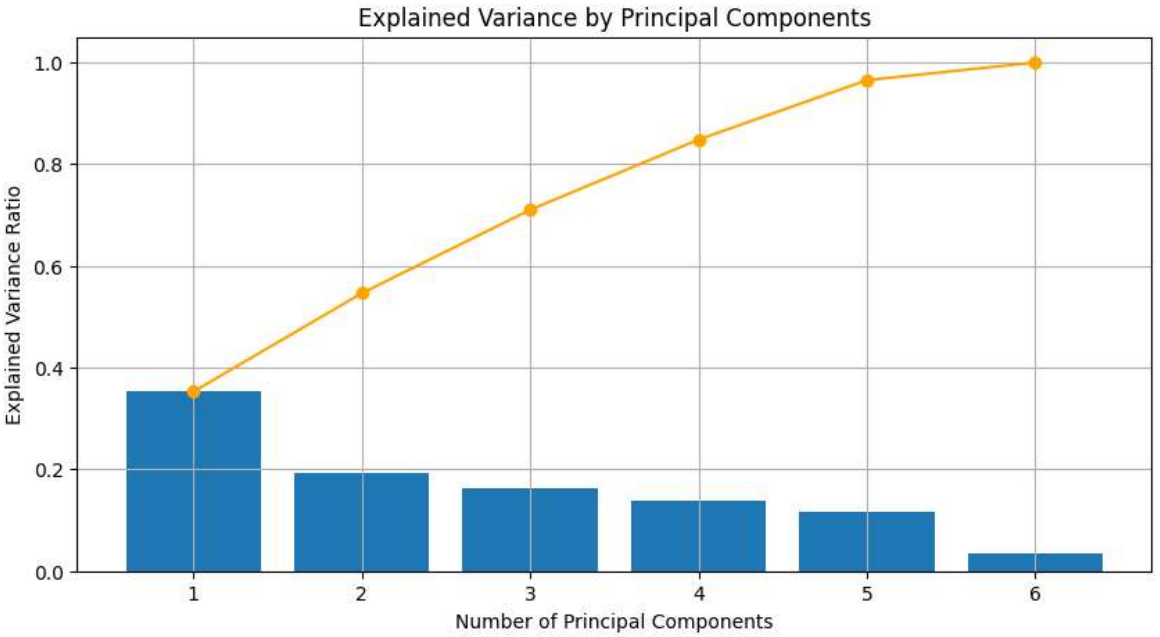


Figure 3: *Explained variance ratio by principal components (PC1–PC6). PC1 captures ~35–40% variance; the first two PCs together >50% (indicated by the cumulative variance curve). Additional components each contribute progressively less variance.*

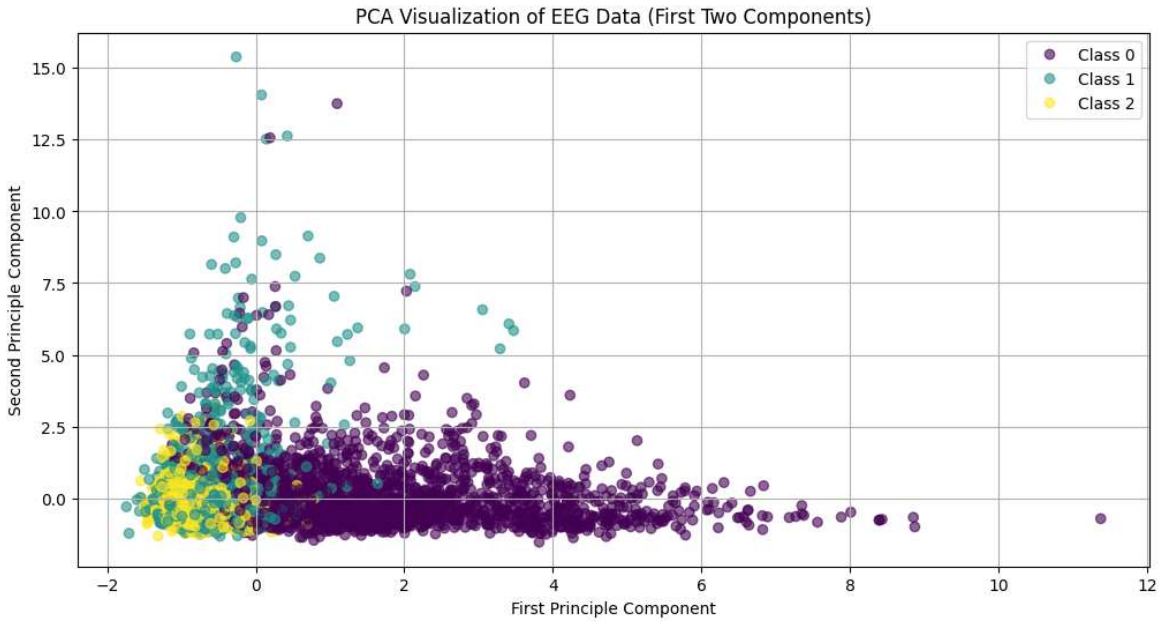


Figure 4: *PCA scatter plot of EEG segments projected onto PC1 vs PC2. Purple = Class 1 (Non-epileptogenic baseline), Green = Class 2 (Epileptogenic baseline), Yellow = Class 3 (Seizure). Class 3 points cluster separately (high PC1 values) whereas Classes 1 and 2 overlap along these axes, indicating linear PCA cannot fully separate Class 1 vs 2.*

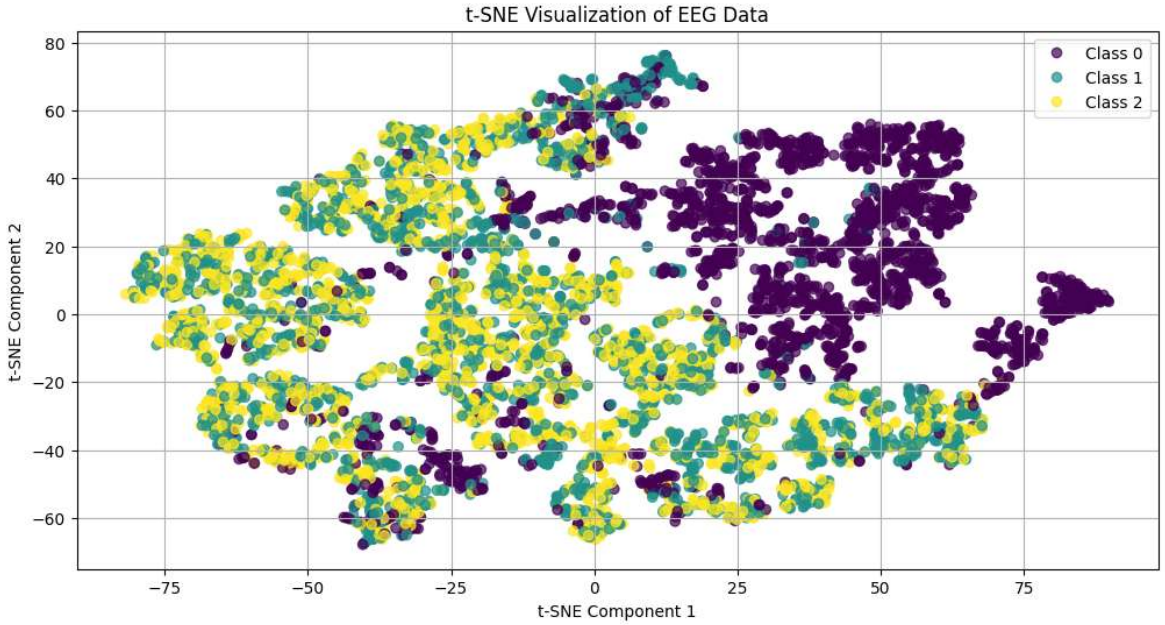


Figure 5: *t-SNE visualization of EEG samples in 2D. Each point is coloured by class (Purple = Class 1, Green = Class 2, Yellow = Class 3). Three distinct clusters form, showing strong separability of Class 3 (seizure) and noticeable, though not perfect, separation between Class 1 and Class 2 in the nonlinear embedding.*

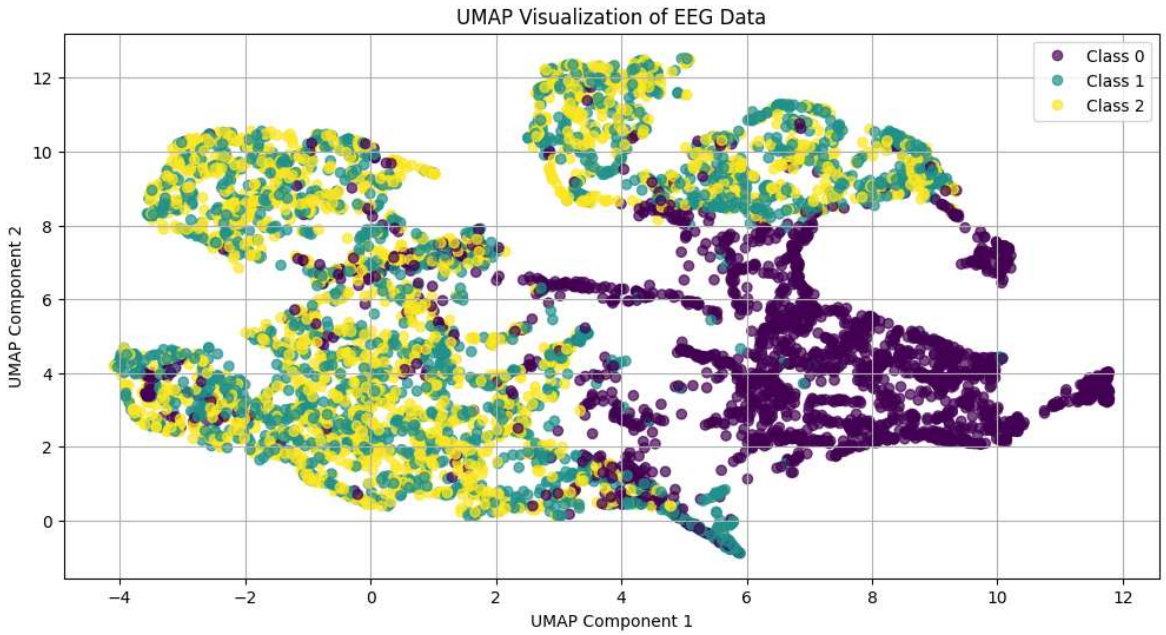


Figure 6: *UMAP visualization of EEG data in 2D. Purple = Class 1, Green = Class 2, Yellow = Class 3. Class 3 forms a well-separated cluster. Classes 2 and 3 (non-seizure classes) appear in adjacent clusters with some overlap, reflecting their greater similarity to each other than to Class 3. UMAP preserves more of the global structure, reinforcing the observation that Classes 1 and 2 are closer in the feature space.*

Discussion

The results demonstrate that machine learning methods can effectively classify EEG segments according to pathological brain states, with certain approaches yielding high accuracy. Several important observations emerge:

Seizure Detection: All models showed that Class 3 (seizure activity) is readily identifiable from even a one-second EEG snippet. This underscores that seizures have distinct signatures (large amplitude swings, rhythmic spikes or oscillations) that are captured by both engineered features and automated feature learning (ROCKET, LSTM). In a clinical sense, this confirms that automated seizure detection from intracranial EEG can be highly reliable, which aligns with prior research in seizure detection. The near-perfect precision and recall for seizure segments in the advanced models is notable – false alarms were minimal and very few seizures went undetected. This level of performance is encouraging for realworld applications like an EEG-based monitoring system for epileptic seizures.

Tumour vs. Healthy Baseline: The more challenging aspect was distinguishing non-seizure EEG from tumour-affected versus non-affected brain regions (Class 2 vs Class 1). Our initial difficulty with this (as seen in the XGBoost and ROCKET results) likely arises because interictal (seizure-free) pathological activity can be subtle. The tumour-epileptogenic tissue may show only slight irregularities or increased slow-wave activity relative to normal tissue, which are not as stark as changes during a seizure. The success of the hybrid approach in improving this differentiation suggests that a tailored strategy is needed for such subtle distinctions. By isolating that decision in a second stage, the model effectively "zoomed in" on features that separate Class 1 and 2, without being distracted by seizure patterns.

Model Architecture and Pipeline: The progression from a single-model approach to a hybrid model illustrates a key point in applied machine learning: understanding the problem domain and data characteristics can guide the design of an optimal model architecture. A one-size-fits-all model (even a complex deep network) might not perform as well as a combination of models each specialised for part of the task. In our case, the LSTM was ideal for capturing temporal dependencies indicative of seizures, whereas the ROCKET+logistic regression was well-suited for capturing differences in signal distributions for the baseline classes. This modular approach improved interpretability as well: we can attribute errors in the first stage vs. second stage, and we know which stage handles which aspect of the prediction.

Clinical Relevance: From a clinical perspective, a system arising from this work could be used in two ways. Firstly, as a **seizure monitoring tool**, it could flag segments of EEG as seizures with high confidence, which is valuable in both epilepsy management and perhaps intraoperative monitoring (where detecting a seizure quickly is crucial). Secondly, as a **diagnostic support tool**, it hints at the potential to identify abnormal EEG patterns related to tumour presence even

outside of seizures. If further refined, an EEG-based classifier might eventually contribute to non-invasive screening for patients who have subtle EEG changes due to a tumour, prompting earlier imaging or intervention. However, our results also highlight that this latter application (distinguishing tumour-related but nonseizure activity) is difficult – even our best model is ~83% accurate between Class 1 and 2 – so it should be approached with caution and more training data or features might be needed to improve confidence.

Limitations: There are several limitations to note. Our dataset, while balanced and well-curated, is relatively small (only a few thousand segments per class, all from a limited number of patients). In a real-world scenario, EEG patterns can vary widely between individuals. The models might need further validation on external datasets or patient-specific tuning. Additionally, our feature engineering was tailored to intracranial EEG; applying this to scalp EEG (which has different noise characteristics and spatial mixing of signals) may require adjustments. Another limitation is that our evaluation was a standard train-test split; in a production or clinical setting, one would prefer cross-validation across patients or leave-one-subject-out validation to ensure the model generalises to unseen patients. We also operated under an assumption that each 1-second segment can be classified independently, but in practice, contextual information (e.g., what was happening in the seconds before) could be leveraged to improve accuracy.

Future Work: Future research could explore several extensions. One direction is incorporating additional features or data modalities, such as connectivity features (how different electrodes' signals relate, though our data was single-channel segments) or patient metadata. Another is improving the Class 1 vs 2 distinction possibly with more advanced techniques, such as attention-based neural networks that might pick up subtle spatial patterns, or by generating synthetic training data to amplify the differences. Ensemble strategies that combine the outputs of multiple model types (beyond our hybrid pipeline) could also be tried. Finally, deploying the model in an online or real-time setting (with a front-end interface as the project originally intended) would be the next step towards practical application, including providing visualisations of EEG and model confidence to help clinicians trust and interpret the model's outputs.

In summary, the **NeuroPredict** system provides a proof-of-concept that combining domain knowledge (feature engineering) with state-of-the-art machine learning yields a powerful tool for EEG analysis. The approach demonstrates high accuracy in identifying seizures and offers a pathway to detect more elusive tumour-related EEG changes. With further validation and development, such models could become valuable assistants in neurological diagnosis and monitoring.

Conclusion

In this paper, we presented an IEEE-style report of the NeuroPredict EEG classification project, reformulated as an academic study. We tackled the problem of predicting brain tumour presence (glioblastoma-related activity) and epileptic seizures from intracranial EEG data through machine learning techniques. The work encompassed data preprocessing, feature engineering of EEG signals, exploratory data analysis, and the development of multiple classification models including an ensemble tree model, a convolutional kernel method, and a novel hybrid LSTM-ROCKET model.

Our results highlight that: - Seizure detection from EEG can be achieved with very high accuracy (over 95% precision and recall) using both classical and deep learning methods, reflecting the distinctness of seizure EEG patterns. - Distinguishing non-seizure EEG from tumour-infiltrated brain vs. normal brain is more challenging, but a tailored two-step model significantly improved this differentiation, raising overall classification accuracy to about 86%. - A hybrid modelling approach that leverages the strengths of different algorithms (sequence learning vs. feature-based learning) can outperform single-model solutions on complex classification tasks.

Through a combination of time-domain and frequency-domain feature extraction and advanced modeling, the NeuroPredict framework effectively learned to identify subtle biomarkers in EEG data. The study demonstrates a pathway towards integrating machine learning into neurodiagnostic workflows—for instance, as a decision support tool for neurologists examining EEGs of patients with brain tumours or epilepsy.

Future Outlook: To transition from research to real-world application, further testing on larger, more diverse datasets is required. Additionally, incorporating this model into a user-friendly interface with real-time analysis capabilities would be a valuable next step. Such a system could provide clinicians with alerts for seizures and insights into abnormal EEG activity that might correspond to tumour effects, potentially improving patient monitoring and outcomes.

In conclusion, this work illustrates the potential of modern AI techniques in processing neural signals. By conforming to rigorous development and evaluation standards, we ensure that the proposed approach is both effective and reliable. The interdisciplinary nature of the project—sitting at the intersection of neuroscience, signal processing, and machine learning—exemplifies how collaboration across fields can lead to innovative solutions for pressing medical challenges.

References

[1] R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. E. Elger, "Indications of nonlineardeterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, p. 061907, 2001.

[2] H. U. Rashid, "Epileptic Seizure Recognition" [Online]. Available: *https://www.kaggle.com/datasets/harunshimanto/epileptic-seizure-recognition*. Accessed 2025.

---

1 The Bonn EEG time series download page - Nonlinear Time Series Analysis (UPF)
https://www.upf.edu/web/ntsa/downloads/-/asset_publisher/xvT6E4pczrBw/content/2001-indications-of-nonlineardeterministic-and-finite-dimensional-structures-in-time-series-of-brain-electrical-activity-dependence-on-recording-regi