A Multi Representation Deep Learning Approach for Epileptic Seizure Detection

Arya Tandy Hermawan [1], Ilham Ari Elbaith Zaeni [2], Aji Prasetya Wibawa [3], Gunawan [4], William Hartanto Hendrawan [5], Yosi Kristian [6]

[1, 2, 3] Department of Electrical and Informatics, Univesitas Negeri Malang, Jawa Timur, Indonesia
[1, 4, 5, 6] Department of Informatics, Institut Sains dan Teknologi Terpadu Surabaya, Jawa Timur, Indonesia
Email: [1] arya.tandy.2205349@students.um.ac.id, [2] ilham.ari.ft@um.ac.id, [3] aji.prasetya.ft@um.ac.id, [4] gunawan@stts.edu

Abstract—**Epileptic seizures, unpredictable in nature and potentially dangerous during activities like driving, pose significant risks to individual and public safety. Traditional diagnostic methods, which involve labour-intensive manual feature extraction from Electroencephalography (EEG) data, are being supplanted by automated deep learning frameworks. This paper introduces an automated epileptic seizure detection framework utilizing deep learning to bypass manual feature extraction. Our framework incorporates detailed pre-processing techniques: normalization via L2 normalization, filtering with an 80 Hz and 0,5 Hz Butterworth low-pass and high-pass filter, and a 50 Hz IIR Notch filter, channel selection based on standard deviation calculations and Mutual Information algorithm, and frequency domain transformation using FFT or STFT with Hann windows and 50% hop. We evaluated on two datasets: the first comprising 4 canines and 8 patients with 2.299 ictal, 23.445 interictal, and 32.915 test data, 400-5000Hz sampling rate across 16-72 channels; the second dataset, intended for testing, 733 icatal, 4.314 interictal, and 1908 test data, each 10 minutes long, recorded at 400Hz across 16 channels. Three deep learning architectures were assessed: CNN, LSTM, and a hybrid CNN-LSTM model-stems from their demonstrated efficacy in handling the complex nature of EEG data. Each model offers unique strengths, with the CNN excelling in spatial feature extraction, LSTM in temporal dynamics, and the hybrid model combining these advantages. The CNN model, comprising 31 layers, yielded highest accuracy, achieving 91% on the first dataset (precision 92%, recall 91%, F1-score 91%) and 82% on the second dataset using a 30-second threshold. This threshold was chosen for its clinical relevance. The research advances epileptic seizure detection using deep learning, indicating a promising direction for future medical technology. Future work will focus on expanding dataset diversity and refining methodologies to build upon these foundational results.**

Keywords—*Epileptic Seizure Detection; Ictal; Interictal; Convolutional Neural Network; Deep Learning; Long Short-Term Memory; Spectrogram.*

I. INTRODUCTION

Epileptic seizures are a prevalent neurological condition, impacting around 50 million individuals worldwide, as indicated by the World Health Organization [1]–[3]. These seizures are distinct from the isolated seizures that may sporadically occur in the general population. Epileptic seizures, which are symptomatic of chronic neurological abnormalities, often persist throughout a person's life, manifesting from birth or emerging at any stage in adulthood.

Unlike isolated seizures, which are usually singular events, epileptic seizures are characterized by their recurrent nature and can strike unpredictably, often without any forewarning, complicating the implementation of safety measures for both the individuals affected and the public. The unpredictability of epileptic seizures poses significant risks, particularly during activities that require sustained alertness, such as driving. This not only jeopardizes the patient's safety but also poses a risk to others [4].

Current diagnostic tools, such as Electroencephalography (EEG) and Magnetic Resonance Imaging (MRI), are instrumental in early seizure detection [4], [5]. However, they come with limitations. EEG's low spatial resolution can make it challenging to accurately determine the source of epileptiform activity, and interpretation often relies heavily on the expertise of medical professionals. MRI provides better spatial resolution but is less effective in detecting seizures when clear structural anomalies are absent and is limited by its low temporal resolution, which impedes its ability to track the rapid progression of epileptic events.

One of the substantial challenges in the field is the difficulty of obtaining balanced EEG datasets. The number of active seizure (ictal) episodes is typically much lower than non-seizure (interictal) episodes, reflecting the actual prevalence of such events. This discrepancy can introduce biases into model training and affect the efficacy of automated detection systems [6]. To address this, we explore mitigation strategies like weighted loss for adjusting loss functions, oversampling, undersampling, and data augmentation, aiming to enhance the performance and reliability of seizure detection models.

The limitations of manual EEG analysis in epilepsy diagnosis are further discussed in [6] which highlights the time-consuming and error-prone nature of this approach. The paper advocates for the adoption of graph-theory-based methods to automate epilepsy detection. These advanced network approaches offer deeper insights into the intricate dynamics of EEG signals, thereby enhancing the accuracy of diagnosis. This serves as a valuable resource for neurologists and researchers aiming to develop intelligent epilepsy detection systems.

In the field of EEG-based emotion and attention detection, significant advancements have been made. For example, a 2022 study by Cui et al. [7] employed Gated Recurrent Units and Minimum Class Confusion to improve subject-independent emotion recognition. In a similar vein, research in [8] work utilized probability-based features in

EEG signals to detect confused students in online educational settings. Additionally, Anala and Bhumireddy's in [9] study conducted a comparative analysis of machine learning algorithms aimed at identifying student confusion during Massive Open Online Courses (MOOCs). These contributions collectively highlight the emerging importance of EEG-based techniques in emotion recognition and educational applications.

There's also other work that show deep learning effectiveness on EEG based data in general, which include fatigue detection [10]–[14], person identification [15], and also mental illness [16]–[19]. Other than EEG, deep learning has also show impressive result on time series task such as gender classification using ECG [20], crude oil price forecasting [21], and also general problem such as music genre classification [22] sentiment analysis [23], and pose classification [24].

Collectively, these studies affirm the effectiveness of EEG-based deep learning techniques in identifying complex patterns in neurological data. They contribute to a foundation from which methods for automated epileptic seizure detection can be developed, leveraging the signal characteristics and deep learning methodologies proven in adjacent fields of emotion recognition and educational assessment. This research uniquely addresses the gap in current seizure detection methods by focusing on the integration of advanced deep learning techniques with EEG data, which has not been extensively explored in existing literature.

Building upon these advancements, our research introduces novel improvements in the realm of deep learning applications for epileptic seizure detection. We have meticulously tailored our preprocessing methods to align with the unique characteristics of EEG datasets, ensuring that the data fed into our models optimally reflects the intricate patterns associated with seizures. Furthermore, we have innovatively modified both the architecture and input models of our deep learning systems. By fine-tuning the hyperparameters, our approach significantly enhances model performance, setting a new benchmark in the accuracy and efficiency of epileptic seizure detection using EEG data. Our approach significantly advances the field by not only applying deep learning to EEG data but also by innovatively modifying the preprocessing and model architecture to specifically cater to the distinctive patterns seen in epileptic seizures.

This paper aims to address this issue by employing deep learning methodologies [25], [26], obviating the need for manual feature extraction. We focus on three main deep learning paradigms: Convolutional Neural Networks (CNN)[27]–[31] as described in [2], [32], [33]. Long Short-Term Memory (LSTM) networks [34], following the works in [1], [35]; and a hybrid approach that utilizes CNN for feature extraction and LSTM for classification, as in [36]. The study categorizes the results into two labels: 'ictal' for detected seizure episodes and 'interictal' for non-seizure episodes. We evaluate three distinct architectural configurations:

1. A hybrid model incorporating CNN for feature extraction and LSTM for classification,

2. A standalone CNN model serving both as a feature extractor and classifier, (Fig. 1 provides an illustrative pipeline of this approach).

3. A standalone LSTM model also functioning as both feature extractor and classifier.

Our research contributions are, the development of preprocessing methods tailored to the unique characteristics of epileptic seizure data, thereby enhancing its accuracy for seizure detection, and we have not only modified existing architectures but also introduced a novel hybrid model that combines the strengths of CNNs for spatial feature extraction and LSTMs for temporal pattern recognition.
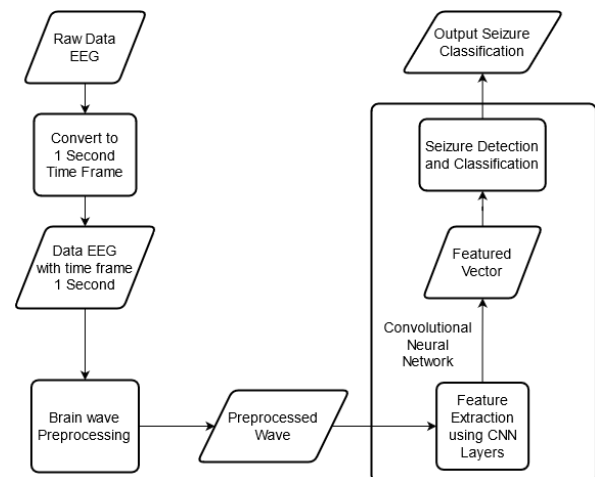

Fig. 1. Architecture system for epileptic seizure detection

II. RELATED WORKS

In the specific area of epileptic seizure detection using Electroencephalogram (EEG) data, many studies utilizing machine learning methodologies have been conducted. These earlier studies serve as the foundational support for the innovations and contributions of the current study. The available literature provides a range of approaches that have been effective in tackling the challenges inherent to EEG-based detection or classification. These approaches vary from traditional machine learning techniques to advanced neural network architectures. This chapter is organized into four distinct sub-chapters to comprehensively cover the state-of-the-art methodologies.

A. Based on CNN Approach

In a recent paper we reviewed, J. Cho and H. Hwang introduced a Convolutional Neural Network (CNN) model in 2020 that they called the C3D model [33]. This model has 31 layers, making it relatively deep. It employs EEG data and also incorporates a 3D CNN via a residual architecture [28]. The model was initially designed for emotion recognition tasks and showed promising results. It uses EEG data from 32 channels, recorded over a one-minute period. In our study, we aim to adapt this model to our specific challenges by adjusting some hyperparameters. Additionally, CNNs have been applied to text-based emotion recognition as well, as cited in paper [37].

In 2018, another paper [2] also employed a CNN model for EEG data analysis. However, this model, introduced by Mengni Zhou et al., is considerably simpler than Cho's, comprising just three layers: one convolutional layer, one average pooling layer, and one fully connected layer with a sigmoid activation function. The data for this model came from intracranial EEG (iEEG) recordings at the University Hospital of Freiburg, Germany, captured using a Neurofile NT digital video EEG system. The design of this simpler model was motivated by two main factors. First, the dataset they used had an imbalance between ictal and interictal samples. Second, the limited number of electrodes used in the recordings constrained the number of layers they could include in the neural network.

The third paper we reference is [32] by Ihsan Ullah et. al. This paper is using the CNN model with a pyramidal shape. Same as Mengni Zhou et. al., the dataset came from the Epilepsy Center at the University Hospital of Freiburg, Germany. The model is consisting of 15 layers with multiple batch normalization [38] layers and ReLU [39] layers. Batch normalization is used to perform normalization just like in the preprocessing but applied along with the learning process. Ihsan Ullah et. al. model could perform on 99% validation accuracy with 10-fold validation.

There are other papers, worthy to mention, that use CNN model approach to do epilepsy recognition by using CNN. Automatic seizure detection using raw amplitude at each sample and channel as the pixel attribute of a 2D image [40] by Gómez C et. al. The best model reached average accuracy and specificity values of 99.3% and 99.6%, respectively. Epileptic seizure detection using CNN model with a Sequential layer, Three Convolutional 2D layers, and Two dense layers with L2 regularizers (l2=0.001) [41] by Gramacki A et. al. Seizure classification using CNN model and FCNN with RFC [42] model as feature selection [43] by Caffarini J et. al. Detecting seizures in pediatric patients using raw multichannel EEG signal recordings that are minimally pre-processed with two-dimensional deep convolutional neural network(2D-DCNN), as an autoencoder, and connect the output with MLP [44] by Abdelhameed A et. al. Seizure detection using Reconstructed Phase Space image and Alex-Net [45] pretrained model reached binary class accuracy (98.5±1.5)% [46] by Ilakiyaselvan N et. al. Using 1D CNN for Seizure Detection on EEG signals that has been processed using Butterworth Filter [47] and Discrete Wavelet Transform [48] by Hassan F et. al.

Essentially, CNN or Convolutional Neural Network is one of the deep learning methods that are frequently used on a picture type of data, whether it is for object recognition, object detection, and anything picture-related [49]–[53]. CNN is built based on how humans see something. The smart feature of CNN itself is training the kernels or windows to recognize something. Kernels or windows will split or divide large pictures into smaller picture sections while the windows are sliding. A CNN will be something like Fig. 2.

The black square is an input image while Cn are convolutional layers and Sn are pooling layers. The convolutional layers will learn some features from the previous input. On the other hand, pooling layers will have a

purpose to reduce the shape of the layers. After many convolutions and pooling layers, there is a fully connected layer to classify or produce label output. Mathematically, equation (1) will describe how the convolution layer works.

$$conv(I,K)_{x,y} = \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} \sum_{k=1}^{n_D} K_{i,j,k} I_{x+i-1,y+j-1,k} \qquad (1)$$
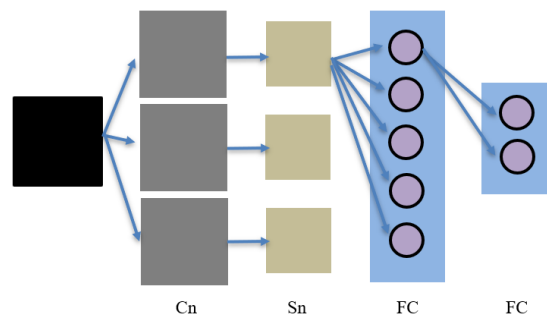

Fig. 2. Simple convolutional neural network

B. Based on LSTM Approach

There are few papers that we observed and reference with LSTM Model. The first paper is [1] by Sirwan Jaafar et. al. In that paper, the researcher using multiple preprocessing techniques. The introduced model is consisting of 1 LSTM layer with 100 cells to learn about the signal feature spike on each channel, one time distributed layer with 500 units, and lastly with softmax function to classify the data between normal and seizure. This model has achieved a total accuracy of 97.75%, which is very high. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit the use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

The second paper is [54] by Ibrahim Aliyu et. al. This paper consists of 1 model LSTM from EEG signal. He also uses DWT preprocessing technique to cut noises and extract features manually into 20 eigenvalues for training and testing. The introduced models are using 3 and 4 LSTM layers which best on 3 layers only with adam optimizer [55] of around 97% accuracy.

The Third paper is [56], they proposed analyzing multivariate time-series using robust LSTM models with attention mechanisms [57], [58]. The paper introduces a new model for time-series forecasting that combines LSTM with tuned Particle Swarm Optimization (PSO) [59] and Bifold-Attention mechanisms. The model outperforms various baselines, including ARIMA and other LSTM configurations, in terms of accuracy. It is tested on multiple environmental and traffic datasets, proving its adaptability and robustness for time-series analysis.

In addition to the aforementioned papers, another notable work in epileptic seizure detection with LSTM-based models is presented by [60] Xinmei Hu et al that employ a band-pass filter to preprocess the original EEG signal and calculate linear features. Subsequently, they utilize a deep neural network comprising a bidirectional long-short term memory network (Bi-LSTM) to train and classify these features. The

results on EEG data of Epilepsy Center of the Bonn University Hospital of Freiburg demonstrate impressive accuracy of 98.56%, showcasing the effectiveness of their proposed model in accurately identifying epileptic EEG signals. There's also work of [61]Mona Nasseri et al that employ LSTM on seizure forecasting using a noninvasive wrist-worn research-grade physiological sensor device, which is tested in patients with epilepsy in an ambulatory setting and ale to achieves forecasting performance significantly better than a random predictor for 5 of 6 patients studied.

LSTM is based on RNN or Recurrent Neural Network. RNN is used for a certain type of problem that needs sequence data like sentences and sound waves, but in this paper, brain waves could be used as well. Different from RNN, LSTM will have an additional gate like forget gate and additional activation function. Because of that, there will be more computation on LSTM than on RNN. Each LSTM cell will be represented in Fig. 3. The first yellow circle means forget gate with the purpose to forget or remember St-1 or previous context combined with input Xt and sigmoid. The second and third yellow circles are input gates that will combine sigmoid and tanh function to make the current cell state by addition from the output of forgetting gate which will be affected by the previous cell state. This is where the sequence comes in handy. The last yellow circle is the output gate to make a new cell state from the current cell state and also a new context for the next LSTM cell. Mathematically, LSTM operation is fairly straightforward and shown in equation (2) to (7).

In other problem [62] Saputra et. al. investigates the impact of varying the number of hidden layers and neurons on LSTM forecasting performance. Using RMSE as the performance metric, the study finds that the most effective architecture utilizes two hidden layers and 64 neurons, achieving an RMSE value of 0.699. The research suggests that increasing the number of hidden layers significantly improves forecasting accuracy, particularly when using 16 and 32 neurons.
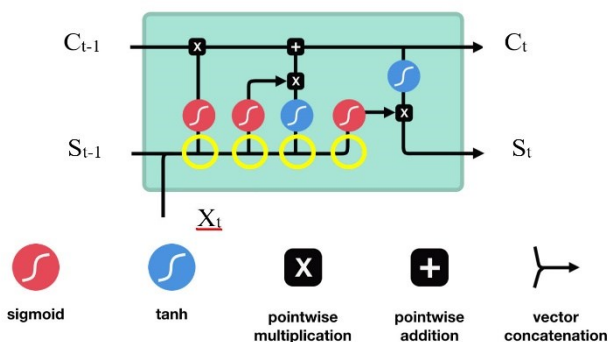

Fig. 3. Long short-term memory cell

$$f_t = \sigma(W_f.[S_{t-1}, X_t] + b_f) \qquad (2)$$

$$i_t = \sigma(W_i.[S_{t-1}, X_t] + b_i) \qquad (3)$$

$$\tilde{c}_t = tanh(W_c.[S_{t-1}, X_t] + b_c) \qquad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{c}_t \qquad (5)$$

$$O_t = \sigma(W_o.[S_{t-1}, X_t] + b_o) \qquad (6)$$

$$S_t = O_t * tanh(C_t) \qquad (7)$$

C. Based on CNN-LSTM Approach

Explaining research chronological, including research design, research procedure (in the form of algorithms, Pseudocode or other), how to test and data acquisition [63]–[65]. The description of the course of research should be supported references, so the explanation can be accepted scientifically [66], [67]. Fig. 1 and Fig. 2 and Table I are presented center, as shown below and cited in the manuscript [63],[68]–[73]. Fig. 2(a) indicated that as 0.3≤α≤0.4, the wind turbine with the rotor velocity control mode can extract more electrical energy than that with the power control mode. Fig. 2(b) shown the smoothing function reaches to the smallest value as α=0.4.

On this architecture, we reference a paper by M. Golmohammadi et. al. They are using a state-of-the-art hybrid architecture that integrates both CNN with RNN (in this paper they are trying to compare either LSTM or GRU). The model is pretty big which integrates 2D CNN 2 layers, 1D CNN 1 layer, and an LSTM layer or GRU [74] layer. They also use a linear frequency cepstral coefficient-based approach as its feature extractions. This model using the input model of an image. The result is pretty good which better on CNN-LSTM than CNN-GRU. Essentially, this hybrid model works just like CNN and LSTM basic before but using CNN great feature extractor combine it with LSTM sequence classifier.

A reference paper [14] explores text-based emotion detection using CNN and BiLSTM, comparing Word2Vec [75] and GloVe [76] embeddings. Two scenarios are tested: classifying text into 'emotion' or 'no-emotion,' and identifying five specific emotions. Word2Vec-CNN-BiLSTM outperforms GloVe-CNN-BiLSTM in all metrices across different transportation datasets. The study concludes that Word2Vec-CNN-BiLSTM improves emotion detection performance in Text.

CNN-Bi-LSTM also use in Yahong Ma [77] research for extraction of spatial feature and capturing spatial feature. An integrated attention mechanism is utilized to allocate specific weights of various electrode channel. Using CHB-MIT dataset EEG for epilepsy detection with 3 categories (Normal, pre and mid seizure) this model reach 94.83% average accuracy, and using UCI dataset this model attain average accuracy 77.62%.

An additional research paper [78] presents another model for seizure prediction by enhancing Long Short-Term Memory networks (LSTM) with Batch Normalization (BNLSTM) and incorporating Channel and Spatial Attention (CASA) mechanisms. This model is designed to effectively capture temporal features in EEG data through its BNLSTM component, while simultaneously addressing spatial information using the CASA component.

In addition to the aforementioned research, the advancement of using CNN-LSTM methodologies has been evident in Epileptic Seizure detection. One notable approach involves a Convolutional Neural Network (CNN) with a Long Short-Term Memory (LSTM) network, as presented by Wang et al. [79]. This dual stream spgrutiotemporal hybrid network achieves remarkable results, boasting an accuracy, specificity, sensitivity, and ROC of 98%, 97.4%, 98.3%, and 96.8%, respectively, in three-class classification. For binary classification, the method outperforms other techniques with perfect accuracy scores. Another significant contribution comes from Srinivasan et al. [80], who employ a three-dimensional deep convolution auto-encoder (3D-DCAE) and a hybrid convolutional auto-encoder (LHCAE) to classify adult epilepsy. The LHCAE method attains impressive metrics with 99.08% accuracy, 99.21% sensitivity, 99.11% specificity, 99.09% precision, and an F1-score of 99.16%, showcasing its efficacy in distinguishing between interictal and ictal states.

Other work of Cao et al. [81] also propose to use squeeze-and-excitation networks (SENet) [82] and LSTM and implement adversarial learning-driven [83] domain-invariant deep feature representation method. This hybrid deep network (HDN) leverages adversarial learning to enhance the classification accuracy of seizure types by 5%. Similarly, Saqib et al. [84] introduce a regularization strategy for CNN-LSTM in EEG seizure detection, employing multi-task learning and achieving a notable improvement in the F1 score. Lastly, Mir et al. [85] contribute a deep learning model, combining a Deep Convolutional Autoencoder (DCAE) with Bidirectional Long Short-Term Memory (Bi-LSTM), achieving outstanding results with an accuracy of 99.8%, classification accuracy of 99.7%, sensitivity of 99.8%, specificity and precision of 99.9%, and an F1 score of 99.6%.

D. Matching Method with No Deep Learning

There are many types of research by many researchers to solve this problem even from 1999 on our reference paper. Because of the early development of modern modeling techniques like the deep learning method, there are some non-deep learning models and also some matching methods like a paper that consists of how to preprocessing or feature extracted from EEG data.

a) Time-Frequency Analysis

From paper [86] indicates that doing epileptic seizure detection is not necessarily using deep learning methods. Paper [87] also uses this method and wins the challenge from the first dataset on Kaggle. These papers inspired us to conduct another approach to a brainwave by using frequency-domain instead of just using raw time-domain signal data. The raw data, which is then preprocessed, is converted using time-frequency conversion as FFT (Fast Fourier Transform) or STFT (Short-Time Fourier Transform). These techniques then analyzed as in [44] to make further predictions by converting this frequency domain data into spectrogram data which they used PSD (Power Spectral Density) format. After being converted, frequency domain data become an input for their respective neural network or ANN. Paper [88] also uses frequency domain for its feature extraction (DFT and DCT), although it is not for epileptic seizure detection but the same

EEG brainwave. Paper [89] also uses frequency-domain conversion plus the wavelet-domain methods as well but using SVM for its classifier.

The challenge winner, Michael Hills [45] already won this challenge in 2014 by using FFT for its conversion, temporal and spectral correlation, and eigenvalues in both time and frequency domain for its feature selection. He was using FFT in low frequency at range 1 until 47 Hz. Afterward, these features were trained by a random forest classifier.

On a different problem, In [90] Zaeni et. al. develops an application for reading practice that records a user's electroencephalograph (EEG) signals to measure concentration levels. Alpha and Beta brainwaves are analyzed to estimate the user's reading comprehension scores using an Artificial Neural Network. The model, which uses four EEG inputs (low and high alpha and beta power), achieves a reasonable accuracy of 73.81% in estimating the scores. While in [91], Rochmah et. al. develops an approach on detecting driver drowsiness using electroencephalogram (EEG) readings using simple artificial neural networks (ANNs) with Backpropagation. The model uses inputs like eSense attention, theta waves, and alpha waves to assess drowsiness. With an architecture featuring multiple layers of neurons and specific activation functions, the model achieves a Mean Absolute Percentage Error (MAPE) of 0.02% and an accuracy rate of 90% in drowsiness detection.

Yazid et. al. [92] use Discrete Wavelet Transform frequency analysis combine with Local Binary Pattern Transition Histogram and Local Binary Pattern Mean Absolute Deviation for enabling efficient feature extraction suitable for machine learning classification, specifically Suport Vector Machine (SVM) and k-Nearest Neighbor (KNN). The model applied on Bonn Epilepsy EEG dataset for classifying ictal against non-ictal and achieve 99.6% accuracy.

b) Channel Selection

Channel Selection is one of preprocessing techniques to manipulate the number of channels because EEG brainwaves came on a big number of channels like 30,40, and even 70 like the first dataset. To make smaller channels, we need to used channel selection. Many papers like [24]–[26] are doing channel selections by various methods. There is a genetic algorithm approach, utilizing wavelet transform, real-time EEG analysis for event detection (REACT), NSGA-II, NSGA-III, and many things else. Those are hard algorithms to apply for a short time. So this paper [93] is using an easier method which is by taking 1 channel with the lowest standard deviation (SD) and the other 4 channels with the highest mutual information (MI) with the first channel. There is another method like using simple SVM [94] to pick the channels by Zhang et al.

III. PROPOSED WORKS

In this chapter, we present our contributions, which consist of modifying and tuning some of the existing models that we adopt. The complete system and its various sub-processes are illustrated in Fig. 4. We describe the rationale and the details of our three proposed modifications and show

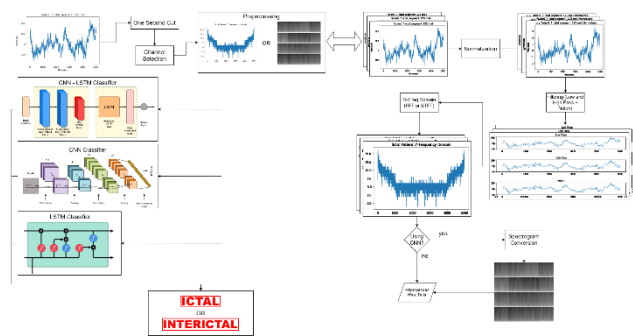how they improve the performance and robustness of the original models.

Fig. 4. Complete system design (left: deep learning model, right: data pre-processing)

The Comparative Analysis of Methodologies encompasses a thorough examination of selected deep learning architectures—CNN, LSTM, and CNN-LSTM—consider specific characteristics. CNNs specialize in grid-like data but face limitations with sequential information. LSTMs excel in capturing long-term dependencies in sequential data, mitigating gradient challenges. CNN-LSTM fusion combines spatial hierarchy capture with sequential memory, beneficial for tasks needing both, but introduces complexity, requiring careful tuning and possibly longer training times. In preparing the data, it is tailored to the requirements of each architecture, aligning with their respective characteristics.

In this study, we prepared three types of data to support our developed architectures. First is the FFT decomposed data from the time domain to frequency domain, sized at 16×400×1. Second, we have grayscale spectrogram data (400×400×1), derived through STFT. Lastly, RGB spectrogram data of the same resolution (400×400×3) is also prepared using STFT. These datasets are integral to providing diverse insights for our analysis.

In the training phase, the dataset is partitioned into three segments: 15% is allocated for testing, 20% of the remaining 85% is set aside for validation, and the rest is used for training. The validation data is not employed in the training process but is reserved for evaluating the model's performance. The training process ceases if the model's performance on the validation data deteriorates or if overfitting occurs. The performance measurement of each created model will be assessed by determining values in the confusion matrix to derive recall, precision, accuracy, and F1 score.

A. CNN-LSTM Model

This research explores two designed CNN-LSTM architectures. The first is an 8-layer CNN-LSTM with time distribution, taking input from FFT decomposition input size 16×400×1. The second is an 11-layer CNN-LSTM without time distribution, utilizing a grayscale spectrogram input of size 16×100×100. Second model draw inspiration from a referenced paper on seizure detection with Gated Recurrent Network (GRN). These architectures are tailored to the specific requirements of the dataset, showcasing their adaptability and potential effectiveness in seizure detection tasks.

Both models can be evaluated as follows. The first is an 8-layer model utilizing FFT decomposed sequence data as input. While sequences are not typically ideal for CNN models, this challenge was addressed with a time-distributed layer. Conversely, the 11-layer model employs grayscale inputs, aligning better with the CNN's strong capabilities in image feature extraction. Meanwhile, the LSTM classifier is expected to more effectively recognize the sequence patterns of EEG signal waveforms."

Our proposed model, a CNN-LSTM hybrid with a total of 8 layers, represents the culmination of our experimentation. It has consistently proven to be the most effective CNN-LSTM model in our trials. Table I provides an overview of the layers within this architecture. The CNN segment features a one-dimensional convolutional layer, a one-dimensional max-pooling layer, and a dropout layer [95]. In contrast, the LSTM portion incorporates 2 LSTM layers followed by 2 fully connected or dense layers.

What sets our model apart is its integration of a convolutional layer within a Time Distributed layer [96]. This innovative approach allows for the concurrent processing of inputs, treating them as multiple inputs for each one-dimensional convolutional layer. In practical terms, this means that instead of processing one input at a time, the model can handle multiple inputs simultaneously. In this context, "multiple inputs" refer to channels from the data. We employ FFT data with a shape of 16×400×1 as input, effectively treating each channel as a distinct input with 400 values.

During training, we utilize the Adam optimizer and employ categorical cross-entropy as the loss function for approximately 20 epochs. This meticulous training process ensures that our model excels in the analysis of sequential data, effectively capturing both spatial and temporal aspects.

The essence of our model aligns with the Time Distributed CNN with LSTM concept. This architecture seamlessly merges Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) to proficiently manage sequential data. Traditional CNNs excel at recognizing spatial patterns but often falter when confronted with temporal sequences. In contrast, LSTMs are adept at capturing long-term dependencies within sequential data. The Time Distributed layer envelops the CNNs, enabling the concurrent processing of input sequences. In practice, a sequence of frames is fed into the Time Distributed CNN, extracting both spatial and temporal features. These features are then transmitted to LSTM units, empowering the network to discern intricate patterns and relationships within the sequential data.

This hybrid approach proves particularly advantageous in domains such as video analysis and natural language processing, where a comprehensive understanding of both spatial and temporal facets of the data is indispensable.

The CNN-LSTM architecture's performance is detailed in the Results and Discussion section, offering a thorough evaluation of its effectiveness based on conducted experiments. Performance metrics like accuracy, precision,

and recall are likely discussed, providing insights into the model's strengths and potential areas for improvement.

Our proposed model, a CNN-LSTM hybrid with a total of 8 layers, represents the culmination of our experimentation. It has consistently proven to be the most effective CNN-LSTM model in our trials. Table I provides an overview of the layers within this architecture. The CNN segment features a one-dimensional convolutional layer, a one-dimensional max-pooling layer, and a dropout layer. In contrast, the LSTM portion incorporates 2 LSTM layers followed by 2 fully connected or dense layers.

TABLE I. CNN-LSTM 8 LAYERS

Layer	Name	Detail
1	Convolutional 1D	Filter = 16, Kernel = 4×4, Strides = 2×2
2	Dropout	Rate = 0.2
3	Max Pooling 1D	Pool Size = 2×2
4	Flatten	-
5	LSTM	Unit = 100, Return Sequence = True
6	LSTM	Unit = 80, Return Sequence = False
7	Dense	Unit = 100, Activation = ReLU
8	Dense	Unit = 2, Activation = Softmax

B. CNN Model

This study involves the design and evaluation of five CNN models with variations in parameters and input types. Model 1 is a 14-layer CNN with batch normalization, employing RGB spectrogram input and experimenting with large kernel and stride sizes, inspired by deep learning approaches in epilepsy detection with EEG. Model 2 is a modification of Model 1 with hyperparameter adjustments in three convolutional layers. Model 3 utilizes 11 layers, replacing batch normalization with more max pooling and altering hyperparameters in three convolutional layers, applied to greyscale spectrogram input. Model 4 is a 31-layer CNN incorporating batch normalization, 2D max pooling, additional convolutional layers, and L2 regularization at specified layers. Inspired by EEG applications for emotion recognition, it is tested for suitability in epilepsy detection. Model 5 is a 4-layer CNN with a simple architecture, using 2D convolution, pooling, flattening, and a dense output layer, tested against a reference for epilepsy seizure detection with EEG and CNN. Batch Normalization is implemented in first architecture for its crucial role in stabilizing input values, accelerating training, and preventing issues like vanishing gradients. It enables higher learning rates, acts as light regularization, streamlines model design, and contributes to faster convergence and enhanced performance.

Various CNN architectures were evaluated with spectrogram inputs. A comparison between 14-layer models (Model 1 and 2) examined the impact of hyperparameter changes, with Model 2's smaller kernel and stride outperforming slightly. The 11-layer CNN, using greyscale spectrogram and excluding batch normalization, showed slightly degraded performance for the first dataset but improved for the second. The CNN with 31 layers achieved the best performance, incorporating L2 regularization. This experiment highlighted the applicability of seizure detection in EEG for emotion detection. The 4-layer CNN, with a

simpler architecture, yielded reasonably good results, albeit lower than other architectures.

The model we proposed is CNN with a total of 31 layers. This is the best CNN model that we experiment with from the other 5 CNN models. This model was a reference to [11] where they are using it for emotion recognition while we are trying to complete the challenge of epileptic seizure detection. Table II shows the layers of our CNN Model, we use 6 convolutional layers, 6 max pooling layers, 6 batch normalization layers, 6 ReLU layers, and 2 dropout layers for CNN parts. It is similar to [11] structure but it has a different hyperparameter and the main difference is that we are using a 2D convolution layer and 2D max pooling layer instead of using a 3D convolution layer and 3D max pooling layer like in [11]. This model was constructed to employing more detailed feature extraction from the input.

We use spectrogram images for its input and in the experiment, we use both types of image (400×400×3 for RGB and 400×400×1 for grayscale). This model was trained around 30 epochs using Adam optimizer and categorical cross entropy as its loss function. There is also L2 regularization on layers 18, 22, and 29 because they have the biggest parameter to overcome overfitting.

TABLE II. CNN 31 LAYERS

Layer	Name	Detail
1	Convolutional 2D	Filter = 32, Kernel = 3×3, Strides = 1×1
2	Batch Normalization	-
3	ReLU	-
4	Max Pooling 2D	Pool Size = 2×2
5	Dropout	Rate = 0.2
6	Convolutional 2D	Filter = 64, Kernel = 3×3, Strides = 1×1
7	Batch Normalization	-
8	ReLU	-
9	Max Pooling 2D	Pool Size = 2×2
10, 14, 18, 22	Convolutioanal 2D	Filter = 128, Kernel = 3×3, Strides = 1×1
11, 15, 19, 23	Batch Normalization	-
12, 16, 20, 24	ReLU	-
13, 17, 21, 25	Max Pooling 2D	Pool Size = 2×2
26	Flatten	-
27	ReLU	-
28	Dropout	Rate = 0.2
29	Dense	Unit = 50, Activation = Linear
30	ReLU	-
31	Dense	Unit = 2, Activation = Softmax

We implement L2 regularization as a valuable technique in deep learning to prevent overfitting. This method adds a penalty based on the squared magnitude of weights to the loss function, promoting a balanced distribution of learning across features. The result is improved generalization to unseen data and increased model robustness, enhancing overall performance. The evaluation of the best-performing CNN model is presented in the Results and Discussion section.

C. LSTM Model

In this research, three LSTM architectures were employed. The first is a 7-layer LSTM, featuring a 256-unit LSTM in the initial layer, followed by two 256-unit LSTM layers, incorporating dropout layers for overfitting prevention. The hyperparameters were adjusted based on previous RNN-based epilepsy detection studies. The second model is a simpler 6-layer LSTM with Time Distributed, comprising a single 200-unit LSTM layer with dropout. The third model, also with Time Distributed, involves a time-distributed dense layer situated between two 100-unit LSTM layers, presenting a more intricate structure compared to the initial 7-layer LSTM model.

In this research, LSTM architectures employed FFT decomposed data, capitalizing on the model's inherent strengths. The first LSTM model, notable for its high unit count, delivered the highest accuracy. The second model, streamlined with a single LSTM layer and time-distributed elements, ranked second in performance efficiency. Another variant, though not surpassing the first two in accuracy, achieved the highest validation accuracy, demonstrating the effectiveness of diverse LSTM configurations in this context.

Table III shows the layers of our LSTM Model. The model we proposed is an LSTM network with a total of 7 layers. This is the best LSTM model that we experiment with from the other 3 LSTM models. This model was a reference to [21] where they are using 3 LSTM and 4 LSTM for trial. We use the 3 LSTM with additional dropout with different rates. The difference between our model and Ibrahim et. al is the input which uses wavelet extraction, and we use FFT for its input with a shape of 16×400. This model is trained in 20 epochs with Adam optimizer and categorical cross-entropy for its loss function. The performance evaluation of these LSTM models is also presented in the Results and Discussion section.

TABLE III. LSTM LAYERS

Layer	Name	Detail
1	LSTM	Unit = 256, Return Sequence = True
2	Dropout	Rate = 0.2
3	LSTM	Unit = 256, Return Sequence = True
4	LSTM	Unit = 256, Return Sequence = True
5	Dropout	Rate = 0.25
6	Flatten	-
7	Dense	Unit = 2, Activation = Softmax

This study advance epileptic seizure detection using EEG data through optimized CNN, LSTM, and CNN-LSTM models. Our key contributions include a CNN-LSTM hybrid model excelling in processing both spatial and temporal data, and a 31-layer CNN demonstrating superior feature extraction. The enhanced accuracy and precision in seizure detection highlight the potential of these models in medical diagnostics, offering valuable insights for future neurological research and healthcare applications.

IV. DATASETS

The data for this study is drawn from two distinct sources. The primary dataset is from Kaggle.com, titled "UPenn and Mayo Clinic's Seizure Detection Challenge," which motivated our research. This dataset is divided into three types: ictal, interictal, and test data. It includes recordings from both humans and dogs. For human subjects, the number of recording channels varies from 16 to 72, and the sampling frequency ranges between 500 and 5000 Hz. In the case of dogs, each recording has 16 channels with a sampling frequency of 400 Hz. Each piece of data corresponds to one second, meaning that each data point is labeled on a per-second basis. We opted to include canine data because it offers benefits for epileptic seizure detection, being compatible with human medical equipment and showing similarities in clinical and neurophysiological traits to human epilepsy [97]–[100] paper described this similarity or helpfulness of canine data to human epilepsy. Specifically, paper [57] notes that treatment options for canine epilepsy resemble those for humans, justifying our inclusion of canine data. As demonstrated in Fig. 5, we present an example of ictal EEG data from a single channel, sampled at a frequency of 500 Hz.
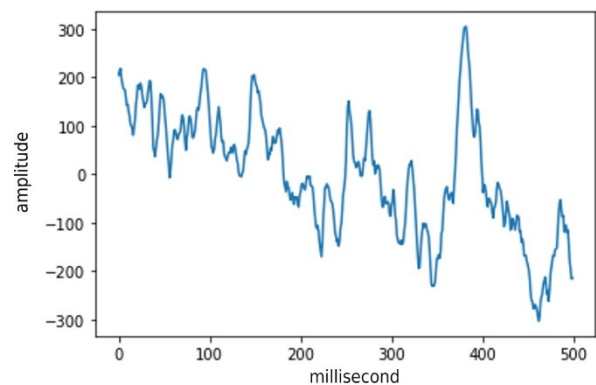

Fig. 5. Ictal EEG 1 channel data with frequency sampling 500 Hz

We utilized a second dataset, also sourced from a Kaggle.com challenge, which is accessible through an agreement with Levin Kuhlmann Ph.D., as cited in paper [101]. This dataset is named "Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge." It contains data from three human patients and spans more than 72 days of recordings. Unlike the first dataset, which has one-second data points, this dataset provides data in 10-minute blocks, each labeled accordingly. To align the formats of the two datasets, we divided each 10-minute block into approximately 580-600 individual seconds, each carrying its own label. For instance, a 10-minute seizure data block is split into 600 one-second segments, all labeled as seizure. For testing, we restricted our focus to a subset of the public dataset, consisting of 564 out of 1908 entries, along with an additional 20 labeled as 'ignored' (meaning these data points were not considered valid for testing).

The data in both datasets are formatted to represent brainwave readings. In the first dataset, the format is N×M, where N stands for the number of channels and M for the sample frequency or values per channel. For instance, a format of 72×5000 indicates 72 channels with 5000 values each. In the second dataset, the format is T×N×M, with T denoting time in seconds. An example format would be 600×16×400. Both datasets are sourced from reputable hospitals, ensuring high validity, and exhibit a relatively large volume of data. Additionally, they share common characteristics such as the number of channels and sampling

rate, although some specifications may differ. This choice is underpinned by the need for robust and comparable datasets. According from choosemuse.com, cited in paper [102], human brainwaves can be broken down into five categories that correspond to different mental states, (Fig. 6 is an illustration of a decomposed human brainwaves):

- Alpha wave (8-13 Hz) indicates a state of physical and mental relaxation.

- Beta wave (13-32 Hz) signifies being awake, alert, and involved in thinking and excitement.

- Gamma wave (32-100 Hz) is associated with heightened perception, learning, and problem-solving.

- Theta wave (4-8 Hz) relates to creativity, insight, dreams, and deep states of meditation.

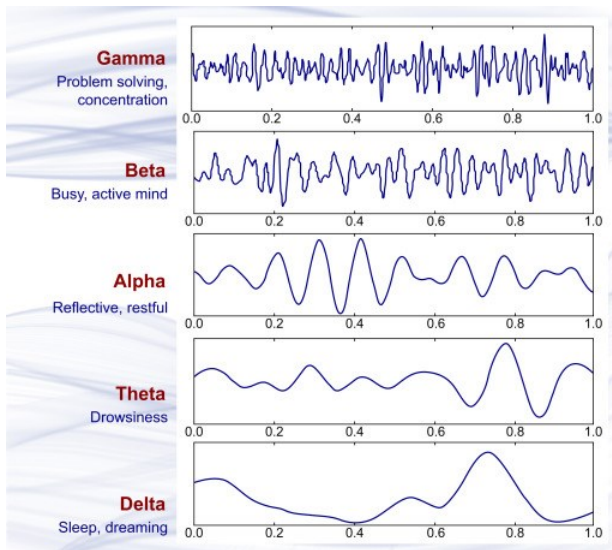- Delta wave (0.5-4 Hz) corresponds to deep, dreamless sleep and bodily repair.

Fig. 6. Decomposed human brainwaves [103]

V. PREPROCESSING

As indicated in Fig. 4, the right side illustrates the data preprocessing steps, which include segmenting the data and selecting specific channels. Below, we outline these preprocessing techniques:

A. One Second Segmentation

The second dataset comes in a 10-minute data format, equating to approximately 580-600 seconds per data point (since some are not exactly 600 seconds). To standardize the data format, we segment the data into one-second blocks. For instance, data from the second dataset often has a shape like 16×240,000. If the sampling frequency is 400, this data is transformed into a shape of 600×16×400.

B. Channel Selection

Given that the first dataset has varying numbers of channels, channel selection becomes necessary. As previously mentioned and supported by paper [51], we employed a straightforward method for channel selection. Unlike the approach in paper [51], we chose 16 channels

when more than 16 were available. This selection process is carried out in three steps:

1. For each channel on array data NxM, do standard deviation like equation (8).

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \tag{8}$$

2. Take 1 channel (let this channel be called A) with the smallest SD and do mutual information with pair of A and each other channels.

$$MI(U,V) = \sum_{i=1}^{|U|} \sum_{j=1}^{|V|} \frac{|U_i \cap V_j|}{N} \log \frac{N|U_i \cap V_j|}{|U_i||V_j|} \tag{9}$$

3. Take the other 15 channels with the highest MI value. So, we had A and 15 other channels.

C. Normalization

Brainwaves tend to have less structure amplitude (sometimes there is a high amplitude around 1000, and there is also data with nearly 0). Because of this, normalization was applied to reduce the difference between high value and low value. The output of this method is between 0 and 1. We use L2 normalization from library scikit-learn to do this.

$$||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \tag{10}$$

D. Filtering

In EEG data, some frequencies, like those from muscle or eye movements, are irrelevant for seizure detection. To address this, we apply three filters:

1. Butterworth low-pass filter (80 Hz cutoff) can be seen in Eq. (11) - Chosen based on the EEG signal characteristics, this filter limits frequencies above the Gamma wave range (up to 100 Hz), which are less impactful for seizure detection and often linked to muscle or eye movements.

$$H(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{fc}\right)^{2n}}} \tag{11}$$

2. Butterworth high-pass filter (0.5 Hz cutoff) can be seen in Eq. (12) - This threshold corresponds to the Delta state (deep sleep), filtering out frequencies that don't typically contain seizure-related information.

$$(f) = \frac{1}{\sqrt{1 + \left(\frac{fc}{f}\right)^{2n}}} \tag{12}$$

3. Notch filter (50 Hz) can be seen in Eq. (12) - Used to remove electrical grid noise, commonly interfering in EEG recordings

$$H(f) = \frac{1}{\sqrt{1 + \left(\frac{f - f_0}{fc}\right)^{2n}}} \qquad (13)$$

Low pass is used to eliminate high-frequency data, while the high pass is used to eliminate low-frequency data. The notch filter is used to cut 50 Hz frequency to remove the influence of 50 Hz power-line noise produce from an electrical device(s) where the recording session took place [1].

E. Frequency Domain Conversion

In accordance with the characteristics of EEG signals, the extraction of information within the signal becomes more facile to analyze and segregate when decomposition into the frequency domain is applied. Specifically, we used Fourier transform methods—either FFT (Fast Fourier Transform) or STFT (Short-Time Fourier Transform)—for this transformation. We generated spectrogram data using these methods, employing the Matplotlib library, which itself uses STFT for the transformation. For handling multiple channels, we merged multiple spectrograms into a single large image, as shown in Fig. 7. For the purpose of this study, we also experimented with RGB-type images. While traditional spectrograms are in grayscale, we aimed to assess whether RGB spectrograms could enhance our model's performance. The large image consists of 16 individual spectrograms, arranged in a 4×4 grid with four channels horizontally and four rows vertically, and no gaps between them.
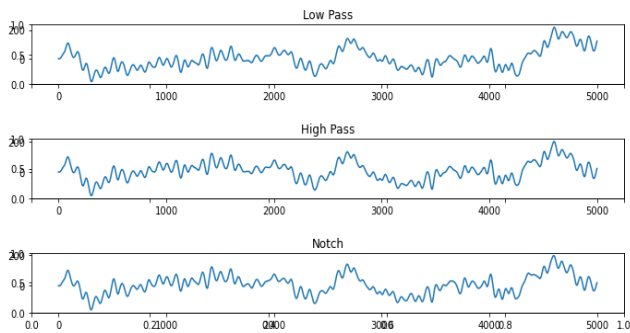

Fig. 7. Low pass, high pass, and notch filtering

FFT is an algorithm to compute DFT (Discrete Fourier Transform), where DFT itself was produced by doing decomposition from a sequence of values with frequency as its parameter. DFT equation could be seen at the equation below. DFT is a slow process, therefore FFT was born to reduce its complexity.

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N} kn} \qquad (14)$$

FFT like [104] is using a butterfly equation that will transform DFT matrices into sparse factors. Therefore from O(N2) down to O(N log N). FFT algorithm will use 2N point DFT like Fig. 8. Each butterfly will divide DFT into Log2N parts while each point is a complex number p and q. On the other side, STFT will have multiplication with window function (commonly Hann window) to non-zero values

because FFT sometimes will produce zero values on the sparse matrices which are not efficient.

$$STFT\{x(t)\}(\tau, \omega) \equiv X(\tau, \omega) \qquad (15)$$

$$X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)\omega(t - \tau)e^{-i\omega t}dt \qquad (16)$$

Where ω(t-τ) is a Hann window function as default. This STFT will give an output with shape determined by hop size, frame size, and sample rate. After STFT was computed, then it will be represented by calculating each magnitude from STFT output as picture or spectrogram data.
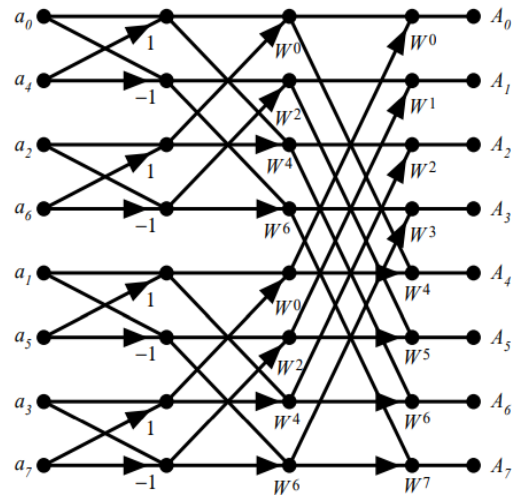

Fig. 8. Butterfly 8-point FFT

CNN architecture is designed for effective grid data processing, crucial for identifying spatial patterns and feature hierarchies in images. In this context, we decomposed EEG signals into the frequency domain, forming them into spectrogram images. RGB spectrograms depict frequency amplitude variations through different colors in the R (red), G (green), and B (blue) channels, while grayscale spectrograms use a single channel where grey levels represent frequency amplitude. In our experiment, we tested both spectrogram types to evaluate their performance. The results showed that, with a 31-layer CNN architecture, grayscale spectrograms yielded better outcomes. However, RGB spectrograms have a drawback in potentially varying performance due to the subjective choice of color representation.

F. Spectrogram Generation for CNN Input

This section discusses the data preparation required for models employing Convolutional Neural Networks (CNN). The data, illustrated as images, is bifurcated into two segments due to the diverse frequency samples present in the first dataset. Fig. 9 depicts this scenario: the left spectrogram is derived entirely from the second dataset and partially from the first dataset, while the right spectrogram is extracted from a portion of the second dataset.

The data utilized here aligns with the Fast Fourier Transform (FFT) data format, specifically 16×400. However, upon its conversion into a spectrogram, a distinct method is employed. Each channel is transformed into a square of

dimensions 100×100. These squares are then systematically arranged in rows to the right and columns downward, as illustrated in Fig. 9. This arrangement results in a structure with four channels to the right and four channels downward, adopting a data format of 400×400×1 for each data entry.
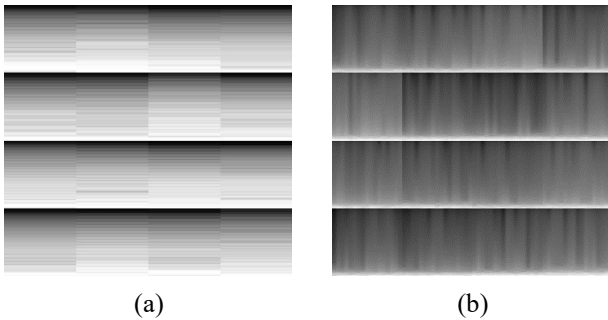

Fig. 9. (a) Spectrogram of 400 value data and (b) Spectrogram of 4000 value data

Moreover, there's an alternative segmentation approach. The data can be divided into segments with dimensions of 16×100×100×1. In this configuration, images are not aligned in rows and columns; instead, each channel is processed individually, allowing for a more granular analysis.

VI. RESULTS AND DISCUSSION

In this section, we display our testing outcomes along with accompanying figures. We also compare the different architectures we have tested. The metrics used for evaluation include accuracy, precision, recall, and F1 Score (Formula for our metrices can be seen in equations (16) to (19).

$$precision = \frac{TP}{TP+FP} \qquad (17)$$

$$recall = \frac{TP}{TP+FN} \qquad (18)$$

$$accuracy = \frac{TP+TN}{TP+FN+TN+FP} \qquad (19)$$

$$F1\ Score = \frac{2\ x\ precision\ x\ recall}{precision+recall} \qquad (20)$$

All metrics are calculated using weighted averages. We have specific testing scenarios, as not all datasets are treated the same:

- First Dataset

In this dataset, we only aim to identify whether a given data point indicates a seizure (ictal) or not (interictal). Unlike the original challenge, we do not consider an "early" label, which is assigned if a seizure is detected in the first 15 seconds. This label is unique to the first dataset and is not included in our testing.

- Second Dataset

In this dataset, the final label spans about 600 seconds. Post-processing is done on the model's output to create this final label, as each data point in the second dataset actually represents 10 minutes. We experimented with different methods to arrive at this final label:

1. A data point is labeled ictal if 300 out of 600 seconds are detected as ictal; otherwise, it is labeled interictal.

2. A data point is labeled interictal if 30 out of 60 seconds are detected as ictal; otherwise, it is labeled interictal.

3. A data point is labeled ictal if a minimum threshold of 10, 20, 30, or 60 continuous seconds within the 600 seconds are ictal; otherwise, it is labeled interictal.

4. A data point is labeled ictal if a minimum threshold of 5, 10, 20, or 30 continuous seconds within 60 seconds are ictal; otherwise, it is labeled interictal.

For example, if a 5-second data point has outputs of 1, 0, 1, 1, and 0, then according to the first and second methods, it will be labeled as ictal. But if we use a 3- second threshold, then according to the third and fourth methods, it will be labeled as interictal, since there are no 3 consecutive seconds with ictal labels.

The first dataset in our study focused on differentiating seizure and non-seizure states, omitting an "early" seizure label for simplicity. The second dataset involved a complex analysis over a 600-second period, using various threshold methods for labeling.

- K-Fold Cross-Validation

To prevent our model from overfitting to a specific training set, we utilize K-Fold cross-validation with K set to 4. This technique is only applied to models that perform well in terms of accuracy and other metrics.

A. CNN-LSTM Results

This model demonstrates strong accuracy in both the training and validation sets, achieving approximately 85% in both cases, as shown in accuracy history chart in Fig. 10. It also performs well in K-Fold testing, with accuracies of 80.97, 80.24, 81.38, and 80.75. Due to its moderate number of parameters, the model is efficiently run even on low-powered GPUs.

We developed two variations of this architecture: Model 1, which uses all available data, and Model 2, which uses only human data. According to the results in Table IV and Table V, Model 1 outperforms Model 2. It should be noted that for the second dataset, we had to rely on the third and fourth methods for label averaging, as the first and second methods proved ineffective. Even when considering human-only data from the first dataset, Model 1 still shows better performance than Model 2. For Model 1, we found that a 10-second threshold out of 60 seconds yielded the best results on the second dataset.

TABLE IV. RESULT OF CNN-LSTM MODEL

Metrics	Acc (Model 1)	Acc (Model 2)
Test Set (Split)	85.00	83.00
1ST Dataset (Human)	**90.54**	68.92
1ST Dataset (Both)	88.71	71.21
2ND Dataset (1)	56.51	36.47
2ND Dataset (2)	61.64	50.68
2ND Dataset (3)	68.41*	58.17*
2ND Dataset (4)	81.33*	73.24*

* Tested with k-fold cross validation k=4 then averaged

TABLE V. RESULT OF CNN-LSTM MODEL ON 1'ST DATASET

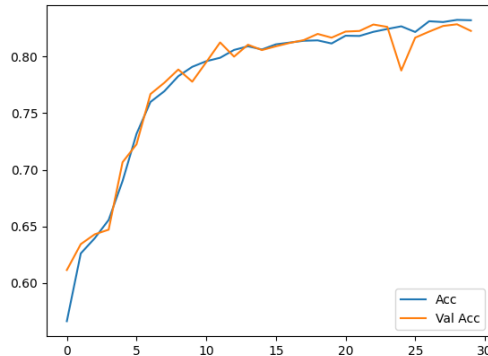| Metrics	Model 1 (Split Set)	Model 1 (Test Set)	Model 2 (Split Set)	Model 2 (Test Set)
Accuracy	85.00	89.00	83.00	71.00
Precision	87.00	87.00	83.00	87.00
Recall	86.00	86.00	82.00	71.00
F1 score	**86.00**	**87.00**	83.00	78.00

Fig. 10. Accuracy CNN-LSTM Model

Comparative analysis of models:

- High accuracy in both training and validation (~85%).

- Efficient performance with a moderate number of parameters, even on low-powered GPUs.

- Model Comparison: Model 1 (all data) outperforms Model 2 (human data only), especially for the second dataset using the third and fourth threshold methods.

In-depth analysis reveals that Model 1 outperforms, particularly for the second dataset, due to its capability in handling a broader range of data variations, thus providing a more robust seizure detection compared to Model 2, which is limited to human data.

B. CNN Results

This model is designed with multiple layers to capture increasingly intricate features from the input data. We generated five different models through this experiment. Models 1 and 2 utilize both and only human data, respectively, and are based on RGB spectrogram images. On the other hand, Models 3 and 4 use both and only human data, respectively, but are built on grayscale spectrogram images. The fifth model incorporates L2 regularization and is derived from the third model, using grayscale images. The model shows impressive accuracy on both training and validation sets, approximately 95% and 89% respectively, as evidenced in Fig. 11. The first model was trained for 20 epochs, and the third model for 30 epochs. Their K-Fold test accuracies are 84.38, 80.35, 86.02, and 84.02. According to the results in Table VI and Table VII, Model 3 delivers the best performance metrics, particularly with a 30-second threshold on the fourth method for the second dataset.

It's worth noting that although there was a high score of 91.10 on the second dataset with model 4 (with an F1 score of 86.39 when weighted), this is not considered reliable because the model labeled all data as interictal due to class imbalance. As depicted in the second image of Fig. 11, the models produced inconsistent scores in the validation sets. L2

regularization did not improve the performance of the fifth model, reducing it by approximately 15%. In terms of other metrics, the models based on 31 layers of CNN perform well on the first dataset, due to its one-second label data. However, their performance is subpar on the second dataset, which is more suited for seizure prediction rather than seizure detection.

TABLE VI. RESULT OF CNN MODEL

Metrics	Model 1 (Acc)	Model 2 (Acc)	Model 3 (Acc)	Model 4 (Acc)
Test Set (Split)	89	89	92	67
1st Dataset (Human)	92	91	88	82
1st Dataset (Both)	93	85	91	77
2nd Dataset (1)	12.32	12.5	12.15	91
2nd Dataset (2)	12.32	25	69	91
2nd Dataset (3)	13.25*	81.13*	81.25*	91.00*
2nd Dataset (4)	51.00*	74.75*	55.75*	91.00*

* Tested with k-fold cross validation $k = 4$ then averaged

TABLE VII. RESULT OF LSTM MODEL

Metrics	Acc(Model 1)	Acc(Model 2)
Test Set (Split)	90	74
1st Dataset (Human)	89.1	89.45
1st Dataset (Both)	86.16	66.4
2nd Dataset (1)	40	42.98
2nd Dataset (2)	52.57	74.49
2nd Dataset (3)	73.79*	75.47*
2nd Dataset (4)	83.34*	89.17*

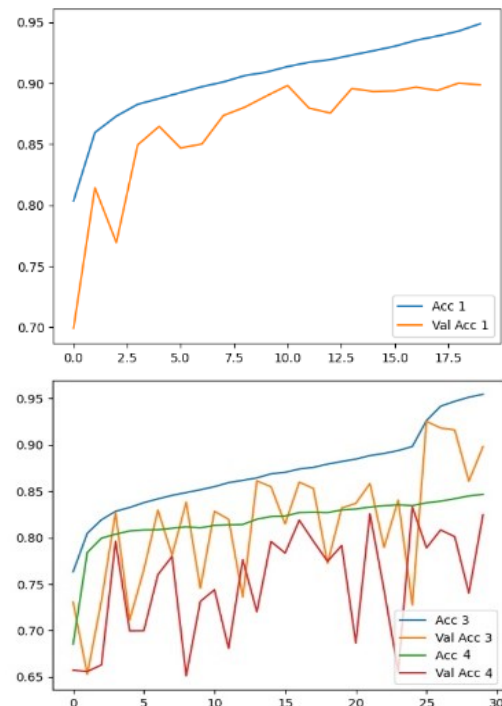* Tested with k-fold cross validation k=4 then averaged

Fig. 11. Accuracy CNN 1st Model and Comparison 3rd and 4th Models

Comparitive analysis of models :

- Multiple layers designed to capture intricate features.

- Five model variations, with Models 3 and 5 (grayscale spectrograms) showing better performance than Models 1 and 2 (RGB spectrograms).

- Model Comparison: Model 3 (both data types, grayscale) has the best performance, especially with a 30-second threshold. However, Model 4's high score on the second dataset is unreliable due to class imbalance.

Grayscale spectrogram-based Models 3 and 5 outperform RGB-based Models 1 and 2, emphasizing the importance of spectrogram type in feature extraction. Model 3 exhibits notable effectiveness with a 30-second threshold, showcasing robustness against class imbalance

C. LSTM Results

This model employs multiple LSTM layers, making it prone to overfitting. It achieves an impressive training accuracy of 97%, but the validation accuracy is less satisfactory. The multiple LSTM layers allow the model to extract and remember intricate features, which leads to overfitting after about 20 epochs. We created two versions of this model: the first version uses a mix of both human and canine data, while the second version uses only human data.

As illustrated in Fig. 12, the model performs well on the training set but falls short on the validation set. It utilizes FFT data for its input. Its K-Fold validation scores are 76.95, 75.97, 76.60, and 75.87. Despite the mediocre validation scores, the model shows promising results on the test sets for both dataset 1 and dataset 2, as can be seen in Table VIII. For the second model, its performance on the test set is not as impressive. As shown in Table VIII and Table IX, the F1 score.
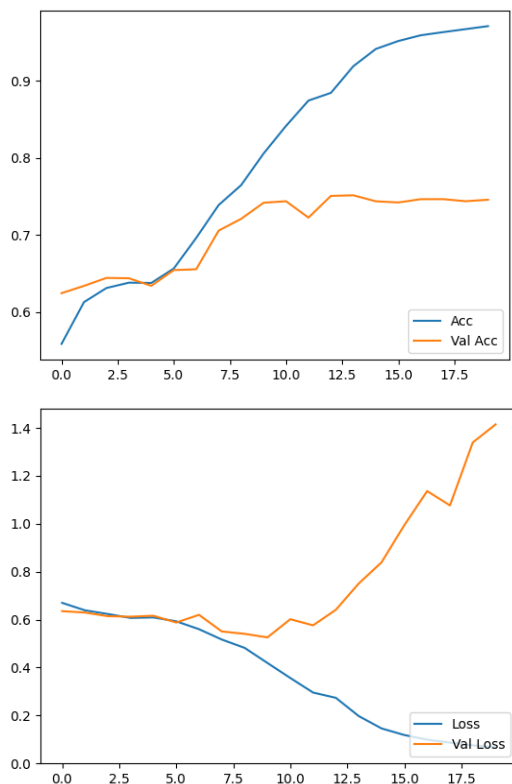

Fig. 12. LSTM Model 1: accuracy, validation accuracy and loss, validation loss

For the model that uses only human data is 10% lower than that of the first model. The first model demonstrates greater stability, leading us to select it as the best LSTM model. Overall, the LSTM models outperform the CNN models on the test set for the second dataset. This is likely because LSTM networks are better suited for handling continuous data, which is the nature of the second dataset. The superior performance of Model 1, as shown in Table VIII, is largely due to its extensive training on both human and canine EEG data, enhancing its effectiveness in epilepsy cases. Table IX further illustrates this, highlighting Model 1's robust pattern recognition and broader adaptability, in contrast to Model 2's limited scope due to training solely on human data. Comparative analysis of models:

- High training accuracy (97%) but lower validation accuracy due to overfitting.

- Better performance on test sets for both datasets, particularly Model 1 (mixed data).

- Model Comparison: Model 1 (mixed data) is more stable and performs better than Model 2 (human data only).

High training accuracy (97%) but lower validation accuracy indicates overfitting. Better test set performance, especially for Model 1 (mixed data). Model 1 proves more stable and outperforms Model 2 (human data only) in the comparison.

TABLE VIII. RESULT OF LSTM MODEL ON 1'ST DATASET

Metrics	Model 1 (Split Set)	Model 2 (Split Test)
Acc (Split)	89	89
Prec (Split)	90	89
Rec (Split)	89	89
F1 (Split)	89	88
Acc (Test)	93	85
Prec (Test)	89	87
Rec (Test)	93	86
F1 (Test)	90	86

D. Results Comparison

In this study, various models were tested, but due to space limitations, not all are included in this paper. The results are summarized in Fig. 11, featuring 2 CNN-LSTM models, 5 CNN models, and 3 LSTM models. The CNN-LSTM category includes a second model with 11 layers that uses an image spectrogram as input, comprising 3 convolutional layers, 2 max pooling layers, and 2 LSTM layers. In the CNN category, variations include models with 14 layers that differ in hyperparameters, a simplified 11-layer model that replaces batch normalization with max pooling, and others with 31 and 4 layers. In the LSTM category, we present 6 and 7-layer models, where the latter includes two LSTM layers. Based on the results in Figures, the best-performing model is a CNN with 31 layers that uses a grayscale spectrogram. We've highlighted the highest and satisfactory scores in green and orange, respectively. These scores are noteworthy because the models genuinely attempt to predict outcomes rather than defaulting to interictal labels. We conclude that the comparison results are:

- Detailed comparison in Fig. 13, highlighting strengths and weaknesses of each model.

- Best Performance: 31-layer CNN using grayscale spectrogram.

- Noteworthy Points: Models genuinely attempt to predict outcomes rather than defaulting to interictal labels.

second dataset, ideal for seizure prediction, but were prone to overfitting. This differentiation highlights the specific strengths and appropriate applications of each model type within our study.

Our model has been refined with extra layers and units for better pattern detection, balancing this against computational efficiency. Additions like batch normalization and dropout, alongside the Adam optimizer, help reduce training time. However, more complex models do require longer testing. Adequate training data is key for improved performance but extends training duration. We've optimized resource use, especially memory and GPU, for better time efficiency.

E. Previous Work Comparison

Comparing with research that is similar to our reset, we analyzed three significant studies. The first researches in [105] reference for CNN, utilized a 3D CNN architecture with 31 and 48 layers, achieving accuracies up to 99.74% using the DEAP dataset for emotion recognition. The second study in [106]- reference for LSTM, applied RNN with discrete wavelet transform preprocessing and eigenvalue feature extraction, achieving 99% accuracy and comparing it against methods like Logistic Regression, SVM, KNN, RF, and Decision Tree. The third research in [107] reference for CNN-LSTM, explored a hybrid CNN-RNN architecture with regularization methods to counteract overfitting, using the Big Data TUG EEG dataset and reporting 30% accuracy with a false alarm rate of 6 times in 24 hours.

Our research optimized deep learning architecture for our dataset by customizing CNN, LSTM, and CNN-LSTM structures. Unlike previous models, we explored tailored variants for dataset-specific characteristics, aiming to identify the most effective model variant. The findings cover dataset implications, architectural adjustments, and performance metrics, along with a discussion of strengths, weaknesses, and key factors."

Previous research on the utilization of EEG signals has exhibited variations in detection objectives, datasets, selected features, and solution methodologies. Several studies are delineated in the related works section. For comparison, the following provides explanations of some studies that utilized the same dataset from UPenn-Mayo Clinic.

The study in [108], utilized UPenn-Mayo Clinic and CHB-MIT datasets with 100 epilepsy patients each. It focused on developing effective automatic seizure detection through individual and global approaches, achieving 99.17% accuracy with the individual approach and 92.69% with the global approach. This highlighted a trade-off between individualized accuracy and efficiency in training time.

Another study in [109] used datasets from Freiburg Hospital, Children's Hospital of Boston-MIT, and UPenn-Mayo Clinic. This research aimed to enhance memory efficiency and hardware compatibility using Integer-Net/CNN, successfully achieving a sevenfold memory efficiency improvement with a mere 2% accuracy reduction.

Lastly, the research in [110] study applied various machine learning methods such as SVM, XGB, KNN, and ensemble models on the UPenn-Mayo Clinic dataset. It
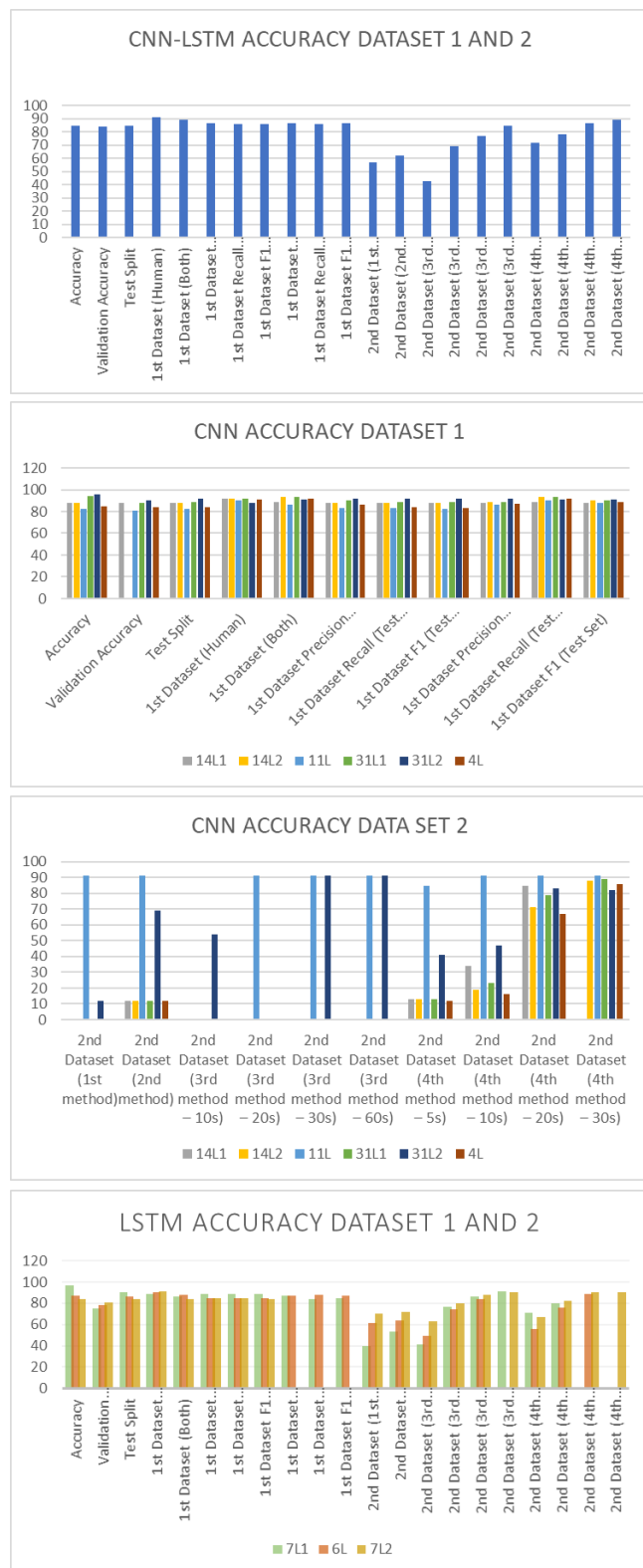
Fig. 13. Result of CNN-LSTM model on each dataset

In our enhanced comparative analysis, CNN models excelled in feature extraction, making them effective for the first dataset with shorter label data. However, they struggled with class imbalance. On the other hand, LSTM models showed superior performance with continuous data in the

utilized an extensive range of EEG signal features, achieving the highest AUC with the ensemble model, with the XGB as the best performing single model. The research exhibited high AUC and Specificity across patients, with Sensitivity mostly above 0.89, except for one outlier achieve only 0.74.

Each study contributes uniquely to epileptic seizure detection, advancing the field through innovative approaches in data utilization, model efficiency, and accuracy optimization.

In a study similar to our research, utilizing dataset in [111], more than 646 participants from 478 teams developed a total of 10,082 algorithms for epilepsy seizure prediction. These teams employed a variety of features, including spectral power, fractal dimensions, and statistical distributions, and utilized machine learning algorithms such as XGBoost, KNN, and SVM. The algorithms' performance was assessed based on AUC, sensitivity, and specificity. The highest AUC achieved was 0.80, while the lowest AUC among the top teams was 0.73, demonstrating a range of successes in different seizure prediction methodologies.

Our research excels in applying diverse deep learning models (CNN, LSTM, CNN-LSTM) with innovative use of FFT and STFT for feature extraction, achieving high accuracies (90.54% CNN-LSTM, 93% CNN, 89% LSTM). This not only broadens model efficacy understanding but also enhances EEG signal analysis, showcasing the effectiveness of combined methodologies in seizure detection.

Despite high accuracy, further exploration is needed on the accuracy-computational efficiency trade-off, particularly compared to the Integer-Net/CNN approach. Future studies should assess model performance across diverse datasets and features for enhanced generalizability

VII. CONCLUSION

In our comprehensive study, the CNN-LSTM and LSTM Model 1 demonstrate enhanced performance, attributed to the incorporation of both human and canine data, facilitating superior handling of data variations. Notably, the CNN with Grayscale Spectrogram surpasses RGB spectrogram models in the CNN 31-layer model, underscoring the importance of feature extraction for optimal performance. The 31-layer CNN excels in epileptic seizure detection for the first dataset, while the 7-layer LSTM, particularly with a 60-second threshold, outperforms alternative models for the second dataset. Model enhancement through layer additions strikes a balance between improved pattern detection and computational efficiency for practical use. Overall, our research identifies the sequentially selected best-performing models as the 31-layer CNN, 7-layer LSTM, and finally, the 8-layer CNN-LSTM. Reflecting on previous studies, we identify key opportunities for improvement in developing sophisticated data features to enhance EEG signal differentiation and accuracy. Exploring diverse, high-performance architectures and employing ensemble models emerge as promising avenues for advancing overall model effectiveness. Our research target is to contribute to the field of deep learning in the medical domain, particularly in epilepsy. We sincerely hope that the promising accuracy results from our research can contribute significantly to the

medical field, especially in epilepsy cases, for application in medical devices.

VIII. FUTURE WORKS

In our future work, we plan to leverage wavelet transforms, which provide variable time-frequency resolution, enabling effective detection of both rapidly and slowly evolving signal components. Scalograms, a method for visualizing wavelets, will be implemented in CNNs for enhanced signal analysis. Ensemble models, proven to be highly effective in numerous studies, will be utilized for their combined strengths and improved overall performance. The integration of multi-modal data sources, such as EEG, Electro Cardio Graphy (EKG), Photoplethysmography (PPG), and temperature measurements, will enrich data features and increase accuracy. Moreover, transfer learning strategies will be adopted to simplify and expedite the training process, tailoring the learning to our specific requirements.

Enhancing model performance necessitates careful consideration of complexity, computational efficiency, resource requirements, and real-world implementation challenges. Various research advancements offer promising solutions to address these constraints.

REFERENCES

[1] S. Jaafar and M. Mohammadi, "Epileptic seizure detection using deep learning approach," *ArXiv*, vol. 3, no. 2, pp. 2–7, 2018, doi: 10.21928/uhdjst.v3n2y2019.pp41-50.

[2] M. Zhou *et al.*, "Epileptic seizure detection based on EEG signals and CNN," *Front Neuroinform*, vol. 12, pp. 1–14, 2018, doi: 10.3389/fninf.2018.00095.

[3] W. Yang, M. Joo, Y. Kim, S. H. Kim, and J. M. Chung, "Hybrid Machine Learning Scheme for Classification of BECTS and TLE Patients Using EEG Brain Signals," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3038948.

[4] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *Neurology in Practice*, vol. 76, no. 2, 2005, doi: 10.1136/jnnp.2005.069245.

[5] M. H. Purnomo, Y. Kristian, E. Setyati, U. Delfana Rosiani, and E. I. Setiawan, "Limitless possibilities of pervasive biomedical engineering: Directing the implementation of affective computing on automatic health monitoring system," *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 1-4, 2016, doi: 10.1109/ICITEED.2016.7863314.

[6] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, "Epilepsy Detection From EEG Using Complex Network Techniques: A Review," *IEEE Reviews in Biomedical Engineering*, vol. 16. 2023. doi: 10.1109/RBME.2021.3055956.

[7] H. Cui, A. Liu, X. Zhang, X. Chen, J. Liu and X. Chen, "EEG-Based Subject-Independent Emotion Recognition Using Gated Recurrent Unit and Minimum Class Confusion," in *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2740-2750, 2023.

[8] T. Daghriri, F. Rustam, W. Aljedaani, A. H. Bashiri, and I. Ashraf, "Electroencephalogram Signals for Detecting Confused Students in Online Education Platforms with Probability-Based Features," *Electronics (Basel)*, vol. 11, no. 18, p. 2855, 2022.

[9] V. A. S. M. Anala and G. Bhumireddy. *Comparison of Machine Learning algorithms on detecting the confusion of students while watching MOOCs*. Blekinge Institute of Technology, 2022.

[10] H. Zeng *et al.*, "A lightGBM-based EEG analysis method for driver mental states classification," *Comput. Intell. Neurosci.*, vol. 2019, 2019.

[11] H. Zeng *et al.*, "InstanceEasyTL: An improved transfer-learning method for EEG-based cross-subject fatigue detection," *Sensors*, vol. 20, no. 24, p. 7251, 2020.

[12] Z.-K. Gao, Y.-L. Li, Y.-X. Yang, and C. Ma, "A recurrence network-based convolutional neural network for fatigue driving detection from

EEG," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 11, 2019.

[13] R. Chai *et al.*, "Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 715–724, 2016.

[14] L. Yao, J. L. Baker, N. D. Schiff, K. P. Purpura, and M. Shoaran, "Predicting task performance from biomarkers of mental fatigue in global brain activity," *J. Neural Eng.*, vol. 18, no. 3, p. 036001, 2021.

[15] Y. Du, Y. Xu, X. Wang, L. Liu, and P. Ma, "EEG temporal–spatial transformer for person identification," *Sci. Rep.*, vol. 12, no. 1, p. 14378, 2022.

[16] J. Sun *et al.*, "A hybrid deep neural network for classification of schizophrenia using EEG Data," *Sci. Rep.*, vol. 11, no. 1, p. 4706, 2021.

[17] H. Sun *et al.*, "Automated tracking of level of consciousness and delirium in critical illness using deep learning," *NPJ Digit. Med.*, vol. 2, no. 1, p. 89, 2019.

[18] M. de Bardeci, C. T. Ip, and S. Olbrich, "Deep learning applied to electroencephalogram data in mental disorders: A systematic review," *Biol. Psychol.*, vol. 162, p. 108117, 2021.

[19] S. Das *et al.*, "Machine learning approaches for electroencephalography and magnetoencephalography analyses in autism spectrum disorder: A systematic review," *Prog. Neuropsychopharmacol Biol. Psychiatry*, vol. 123, p. 110705, 2023.

[20] K. Y. Halim, D. T. Nugrahadi, M. R. Faisal, R. Herteno, and I. Budiman, "Gender classification based on electrocardiogram signals using long short term memory and bidirectional long short term memory," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 606–618, 2023.

[21] M. F. Maulana, S. Sa'adah, and P. E. Yunanto, "Crude Oil Price Forecasting Using Long Short-Term Memory," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 7, no. 2, pp. 286–295, 2021.

[22] M. S. Islam *et al.*, "Machine learning-based music genre classification with pre-processed feature analysis," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 7, no. 3, pp. 491–502, 2021.

[23] A. Z. R. Adam and E. B. Setiawan, "Social Media Sentiment Analysis using Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU)," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 1, pp. 119–131, 2023.

[24] K. A. Tanjaya, M. F. Naufal, and H. Arwoko, "Pilates Pose Classification Using MediaPipe and Convolutional Neural Networks with Transfer Learning," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 2, pp. 212–222, 2023.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[26] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[30] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[31] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

[32] I. Ullah, M. Hussain, E. Qazi, and H. Aboalsamh, "An Automated System for Epilepsy Detection using EEG Brain Signals based on Deep Learning Approach Insight Centre for Data Analytics, National University of Ireland, Galway, Ireland Visual Computing Lab, Department of Computer Science, College of Com," *Expert Syst. Appl.*, no. 107, pp. 61–71, 2018.

[33] J. Cho and H. Hwang, "Spatio-temporal representation of an electoencephalogram for emotion recognition using a three-

[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[35] J. Kumar, R. Goomer, and A. K. Singh, "Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model for Cloud Datacenters," *Procedia Comput. Sci.*, vol. 125, pp. 676–682, 2018, doi: 10.1016/j.procs.2017.12.087.

[36] M. Golmohammadi *et al.*, "Gated recurrent networks for seizure detection," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–5, 2017.

[37] D. E. Cahyani, A. P. Wibawa, D. D. Prasetya, L. Gumilar, F. Akhbar, and E. R. Triyulinar, "Text-Based Emotion Detection using CNN-BiLSTM," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 1–5, 2022, doi: 10.1109/ICORIS56080.2022.10031370.

[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, 2015.

[39] K. Fukushima, "Visual feature extraction by a multilayered network of analog threshold elements," *IEEE Transactions on Systems Science and Cybernetics*, vol. 5, no. 4, pp. 322–333, 1969.

[40] C. Gómez, P. Arbeláez, M. Navarrete, C. Alvarado-Rojas, M. Le Van Quyen, and M. Valderrama, "Automatic seizure detection based on imaged-EEG signals through fully convolutional networks," *Sci. Rep.*, vol. 10, no. 1, p. 21833, 2020.

[41] A. Gramacki and J. Gramacki, "A deep learning framework for epileptic seizure detection based on neonatal EEG signals," *Sci. Rep.*, vol. 12, no. 1, p. 13010, 2022.

[42] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, pp. 278–282, 1995.

[43] J. Caffarini *et al.*, "Engineering nonlinear epileptic biomarkers using deep learning and Benford's law," *Sci. Rep.*, vol. 12, no. 1, p. 5397, 2022.

[44] A. Abdelhameed and M. Bayoumi, "A deep learning approach for automatic seizure detection in children with epilepsy," *Front Comput. Neurosci.*, vol. 15, p. 650050, 2021.

[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process Syst.*, vol. 25, 2012.

[46] N. Ilakiyaselvan, A. N. Khan, and A. Shahina, "Deep learning approach to detect seizure using reconstructed phase space images," *J. Biomed. Res.*, vol. 34, no. 3, p. 240, 2020.

[47] S. Butterworth, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, no. 6, pp. 536–541, 1930.

[48] F. Hassan, S. F. Hussain, and S. M. Qaisar, "Epileptic Seizure Detection Using a Hybrid 1D CNN-Machine Learning Approach from EEG Data," *J. Healthc. Eng.*, vol. 2022, 2022.

[49] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process Syst.*, vol. 28, 2015.

[50] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[51] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[52] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, 2015.

[54] I. Aliyu, Y. B. Lim, and C. G. Lim, "Epilepsy detection in EEG signal U sing recurrent neural network," *ACM International Conference Proceeding Series*, pp. 50–53, 2019, doi: 10.1145/3325773.3325785.

[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[56] A. Pranolo, Y. Mao, A. P. Wibawa, A. B. P. Utama, and F. A. Dwiyanto, "Robust LSTM With Tuned-PSO and Bifold-Attention Mechanism for Analyzing Multivariate Time-Series," *IEEE Access*, vol. 10, pp. 78423–78434, 2022, doi: 10.1109/ACCESS.2022.3193643.

[57] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[58] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[59] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, pp. 1942–1948, 1995.

[60] X. Hu and Q. Yuan, "Epileptic EEG identification based on deep Bi-LSTM network," in *2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT)*, pp. 63–66, 2019.

[61] M. Nasseri *et al.*, "Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning," *Sci. Rep.*, vol. 11, no. 1, p. 21935, 2021.

[62] A. Saputra, A. Wibawa, U. Pujianto, A. B. P. Utama, and A. Nafalski, "LSTM-based Multivariate Time-Series Analysis: A Case of Journal Visitors Forecasting," *ILKOM Jurnal Ilmiah*, vol. 14, pp. 57–62, 2022, doi: 10.33096/ilkom.v14i1.1106.57-62.

[63] T. Kimura, K. Takeshita, T. Toyono, M. Yokota, K. Nishimatsu, and T. Mori, "Network failure detection and diagnosis by analyzing syslog and SNS data: Applying big data analysis to network operations," *NTT Technical Review*, vol. 11, no. 11, 2013.

[64] A. D. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors*, vol. 19, no. 2, p. 326, 2019.

[65] F. Al-Turjman, H. Zahmatkesh, and L. Mostarda, "Quantifying uncertainty in internet of medical things and big-data services using intelligence and deep learning," *IEEE Access*, vol. 7, pp. 115749–115759, 2019.

[66] G. Nguyen *et al.*, "Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey," *Artif. Intell. Rev.*, vol. 52, pp. 77–124, 2019.

[67] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.

[68] S. Kumar and M. Singh, "Big data analytics for healthcare industry: impact, applications, and tools," *Big data mining and analytics*, vol. 2, no. 1, pp. 48–57, 2018.

[69] L.-M. Ang, K. P. Seng, G. K. Ijemaru, and A. M. Zungeru, "Deployment of IoV for smart cities: Applications, architecture, and challenges," *IEEE access*, vol. 7, pp. 6473–6492, 2018.

[70] B. P. L. Lau *et al.*, "A survey of data fusion in smart city applications," *Information Fusion*, vol. 52, pp. 357–374, 2019.

[71] Y. Wu *et al.*, "Large scale incremental learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 374–382, 2019.

[72] A. Mosavi, S. Shamshirband, E. Salwana, K. Chau, and J. H. M. Tah, "Prediction of multi-inputs bubble column reactor using a novel hybrid model of computational fluid dynamics and machine learning," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 482–492, 2019.

[73] V. Palanisamy and R. Thirunavukarasu, "Implications of big data analytics in developing healthcare frameworks–A review," *Journal of King Saud University-Computer and Information Sciences*, vol. 31, no. 4, pp. 415–425, 2019.

[74] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[75] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[76] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[77] Y. Ma *et al.*, "A Multi-Channel Feature Fusion CNN-Bi-LSTM Epilepsy EEG Classification and Prediction Model Based on Attention Mechanism," *IEEE Access*, vol. 11, pp. 62855–62864, 2023, doi: 10.1109/ACCESS.2023.3287927.

[78] M. Ma *et al.*, "Early Prediction of Epileptic Seizure Based on the BNLSTM-CASA Model," *IEEE Access*, vol. 9, pp. 79600–79610, 2021, doi: 10.1109/ACCESS.2021.3084635.

[79] X. Wang, Y. Wang, D. Liu, Y. Wang, and Z. Wang, "Automated recognition of epilepsy from EEG signals using a combining space–time algorithm of CNN-LSTM," *Sci. Rep.*, vol. 13, no. 1, p. 14876, 2023.

[80] S. Srinivasan, S. Dayalane, S. kumar Mathivanan, H. Rajadurai, P. Jayagopal, and G. T. Dalu, "Detection and classification of adult epilepsy using hybrid deep learning approach," *Sci. Rep.*, vol. 13, no. 1, p. 17574, 2023.

[81] X. Cao, B. Yao, B. Chen, W. Sun, and G. Tan, "Automatic seizure classification based on domain-invariant deep representation of EEG," *Front. Neurosci.*, vol. 15, p. 760987, 2021.

[82] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

[83] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[84] M. Saqib, Y. Zhu, M. Wang, and B. Beaulieu-Jones, "Regularization of deep neural networks for EEG seizure detection to mitigate overfitting," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 664–673, 2020.

[85] W. A. Mir, M. Anjum, and S. Shahab, "Deep-EEG: an optimized and robust framework and method for EEG-based diagnosis of epileptic seizure," *Diagnostics*, vol. 13, no. 4, p. 773, 2023.

[86] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time-frequency analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, 2009, doi: 10.1109/TITB.2009.2017939.

[87] M. Hills, "Seizure detection using FFT, temporal and spectral correlation coefficients, eigenvalues and Random Forest," *Github, San Francisco, CA, USA, Tech. Rep*, pp. 1–10, 2014.

[88] K. Tsuru and G. Pfurtscheller, "Brainwave Biometrics: A New Feature Extraction Approach with the Cepstral Analysis Method," *Biomedical Engineering*, vol. 50, no. 1, pp. 162–167, 2012, doi: 10.11239/jsmbe.50.162.

[89] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "EEG seizure detection and prediction algorithms: a survey," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, pp. 1–21, 2014, doi: 10.1186/1687-6180-2014-183.

[90] I. A. E. Zaeni, U. Pujianto, A. R. Taufani, M. Jiono, and P. S. T. Muhammad, "Concentration Level Detection Using EEG Signal on Reading Practice Application," in *2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, pp. 354–357, 2019. doi: 10.1109/ICEEIE47180.2019.8981453.

[91] A. Rochmah, S. Sendari, and I. A. E. Zaeni, "Sleepiness Detection For The Driver Using Single Channel EEG With Artificial Neural Network," in *2019 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, pp. 80–85, 2019. doi: 10.1109/ICAMIMIA47173.2019.9223371.

[92] M. Yazid *et al.*, "Simple Detection of Epilepsy from EEG Signal Using Local Binary Pattern Transition Histogram," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3126065.

[93] A. Bhattacharyya and R. B. Pachori, "A Multivariate Approach for Patient-Specific EEG Seizure Detection Using Empirical Wavelet Transform," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2003-2015, 2017, doi: 10.1109/TBME.2017.2650259.

[94] Y. Zhang, S. Yang, Y. Liu, Y. Zhang, B. Han, and F. Zhou, "Integration of 24 feature types to accurately detect and predict seizures using scalp EEG signals," *Sensors*, vol. 18, no. 5, 2018, doi: 10.3390/s18051372.

[95] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[96] S. Montaha, S. Azam, A. K. M. R. H. Rafid, M. Z. Hasan, A. Karim, and A. Islam, "TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study," *IEEE Access*, vol. 10, pp. 60039–60059, 2022, doi: 10.1109/ACCESS.2022.3179577.

[97] M. Berendt, H. Høgenhaven, A. Flagstad, and M. Dam, "Electroencephalography in dogs with epilepsy: Similarities between human and canine findings," *Acta Neurol Scand*, vol. 99, no. 5, pp. 276–283, 1999, doi: 10.1111/j.1600-0404.1999.tb00676.x.

[98] I. E. Leppik, E. N. Patterson, L. D. Coles, E. M. Craft, and J. C. Cloyd, "Canine status epilepticus: A translational platform for human therapeutic trials," *Epilepsia*, vol. 52, no. 8, pp. 31–34, 2011, doi: 10.1111/j.1528-1167.2011.03231.x.

[99] A. Catala, M. Grandgeorge, J. L. Schaff, H. Cousillas, M. Hausberger, and J. Cattet, "Dogs demonstrate the existence of an epileptic seizure odour in humans," *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, 2019, doi: 10.1038/s41598-019-40721-4.

[100] B. H. Brinkmann *et al.*, "Crowdsourcing reproducible seizure forecasting in human and canine epilepsy," *Brain*, vol. 139, no. 6, pp. 1713–1722, 2016, doi: 10.1093/brain/aww045.

[101] L. Kuhlmann *et al.*, "Epilepsyecosystem.org: Crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG," *Brain*, vol. 141, no. 9, pp. 2619–2630, 2018, doi: 10.1093/brain/awy210.

[102] R. Jaswal, "Brain Wave Classification and Feature Extraction of EEG Signal by Using FFT on Lab View," *International Research Journal of Engineering and Technology*, pp. 1208–1212, 2016.

[103] P. A. Abhang, B. W. Gawali, and S. C. Mehrotra. *Introduction to EEG- and Speech-Based Emotion Recognition*. Academic Press. 2016. doi: 10.1016/C2015-0-01959-1.

[104] P. Heckbert, "Fourier Transforms and the Fast Fourier Transform (FFT) Algorithm," *Notes Computer Graphics*, vol. 3, no. 2, pp. 15–463, 1995.

[105] J. Cho and H. Hwang, "Spatio-temporal representation of an electoencephalogram for emotion recognition using a three-dimensional convolutional neural network," *Sensors*, vol. 20, no. 12, pp. 1–18, 2020, doi: 10.3390/s20123491.

[106] I. Aliyu, Y. B. Lim, and C. G. Lim, "Epilepsy detection in EEG signal U sing recurrent neural network," *ACM International Conference Proceeding Series*, pp. 50–53, 2019, doi: 10.1145/3325773.3325785.

[107] M. Golmohammadi *et al.*, "Gated recurrent networks for seizure detection," *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1-5, 2017, doi: 10.1109/SPMB.2017.8257020.

[108] A. Ciurea, C.-P. Manoila, A.-M. Tautan, and B. Ionescu, "Low Latency Automated Epileptic Seizure Detection: Individualized vs. Global Approaches," in *2020 International Conference on e-Health and Bioengineering (EHB)*, pp. 1–4, 2020, doi: 10.1109/EHB50910.2020.9280267.

[109] N. D. Truong *et al.*, "Integer Convolutional Neural Network for Seizure Detection," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 8, no. 4, pp. 849–857, 2018, doi: 10.1109/JETCAS.2018.2842761.

[110] Y. Shen, "Machine Learning Based Epileptic Seizure Detection for Responsive Neurostimulator System Optimization," *J. Phys. Conf. Ser.*, vol. 1453, no. 1, p. 12089, 2020, doi: 10.1088/1742-6596/1453/1/012089.

[111] L. Kuhlmann, LizLopez, M. O'Connell, rudyno5, S. Wang, and W. Cukierski. *Melbourne University AES/MathWorks/NIH Seizure Prediction*. Kaggle, 2016.