

Epilepsy Detection Using Deep Neural Networks and Explainable Artificial Intelligence Techniques

by

Prajakta S. Rathod

A dissertation submitted to the faculty of Ahmedabad University in

partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Under the guidance of

Dr. Jayendra M. Bhalodiya

School of Engineering and Applied Science, Ahmedabad University,
Ahmedabad - 380009, Gujarat - India

January, 2026

© Prajakta S. Rathod; 2026
All Rights Reserved

Declaration by the Candidate

This is to certify that the dissertation titled, "**Epilepsy Detection Using Deep Neural Networks and Explainable Artificial Intelligence Techniques**" comprises my original work towards the degree of Doctor of Philosophy at Ahmedabad University and has been completed under the supervision of **Dr. Jayendra M. Bhalodiya**. The matter embodied in this dissertation has not been submitted elsewhere for the award of any other degree/diploma. I have not intentionally copied material from any other source, such as journals, books, magazines, websites, etc., and due acknowledgement has been made in the text of all material used from other sources.

Prajakta S. Rathod

Date: 30/01/26

Place: Ahmedabad

Certificate by the Supervisor

It is certified that, to the best of my knowledge, the above statements made by the student are correct.

Dr. Jayendra M. Bhalodiya

Date: 30/01/26

Place: Ahmedabad

Ahmedabad University

Certificate

This is to certify that the dissertation work titled "**Epilepsy Detection Using Deep Neural Networks and Explainable Artificial Intelligence Techniques**" has been completed by **Prajakta S. Rathod (AU1849007)** for the degree of Doctor of Philosophy at the School of Engineering and Applied Science of Ahmedabad University under my supervision.

Dr. Jayendra M. Bhalodiya: Supervisor

Date: 30/01/2026

The members of the Dissertation Advisory Committee were:

Prof. Mukesh Zaveri: External Expert

Dr. Mazad Zaveri: Internal Expert

Dr. Keyur Joshi: Internal Expert

Dr. Shefali Naik: Internal Expert

Dedication

This thesis is dedicated to my family, whose silent and countless sacrifices, patience, and selfless support made this journey possible. Their willingness to place my dreams above their own comforts, their understanding during long hours of work, and their quiet strength during challenging times gave me the courage to persevere. They stood beside me through long days and uncertain moments, often placing my aspirations above their own comfort and time. This achievement reflects their sacrifices, love, and unconditional support.

Abstract

Epilepsy is a neurological disorder in which a person experiences recurrent and unprovoked seizures. Around 50 million people have epilepsy globally. The number of epilepsy patients is increasing, as every year, 5 million new patients are confirmed to have epilepsy globally. Epilepsy is prevalent in low and middle-income countries where medical facilities are limitedly available. If epilepsy is not diagnosed at an early stage, then patients may have accidents or injuries during seizure incidents. Automated epilepsy detection systems play a significant role in the early diagnosis of epilepsy. The automated epilepsy detection system has several advantages over manual epilepsy detection. Automated systems can monitor longer Electroencephalogram (EEG) recordings without subjectivity, diagnose efficiently and accurately, and are more suitable for real-time seizure onset detection. Moreover, automated systems are useful in remote areas where trained medical experts are not available or available remotely. Such systems can be developed using advanced Machine Learning (ML) algorithms of Artificial Intelligence (AI) and Explainable Artificial Intelligence (XAI) methods. XAI methods play a significant role in explaining the predictions of AI-based epilepsy detection systems. The use of XAI methods helps these systems adapt in the medical domain for disease diagnosis.

This research aims to develop an automated and interpretable epilepsy detection system and boost the classification accuracy of the system with improved preprocessing techniques. Identify the most suitable among the widely used XAI methods and Deep Neural Networks (DNNs) for epilepsy detection. Five methods are implemented to achieve these research goals. In the first method, a Bidirectional Long Short-Term

Memory (Bi-LSTM) network is used to identify epileptic signals, and the XAI method, Layer-wise Relevance Propagation (LRP), is implemented in the Bi-LSTM network to explain the network predictions. The model classified epileptic and non-epileptic EEG signals with an accuracy of 87.25%. In the second method, EEG signals are converted into images by applying the Short-time Fourier Transform (STFT) to improve the classification accuracy. A DNN, suitable for image classification, i.e., the VGG16 network, is used to classify EEG signal images and achieves a classification accuracy of 99.89%. In the third method, the performance of widely used XAI methods, i.e., Shapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) and Guided Backpropagation, is compared to explain the network predictions. It is found that the XAI method SHAP is producing the best explanations for epileptic signal identification. Therefore, the SHAP method is selected in the fourth method. The performance of four DNNs - VGG16, Resnet18, Densenet121 and Squeeznet1_0 - is compared for epilepsy detection and explanation using the SHAP method. These four networks are reported to have produced promising results for epileptic signal image classification in the literature. Therefore, these four networks are used to classify EEG signals as seizure, interictal and non-seizure. It is found that Resnet18 achieved the highest classification accuracy of 99.7% and VGG16 produced the best explanations using the XAI method SHAP. DNNs are trained using transfer learning with a fine-tuning approach. This approach takes advantage of the previous learning of the model, and the model can be trained on a small amount of data and in less time.

All four methods are tested on the single-channel Bonn University dataset. The research is strengthened using a multichannel epilepsy dataset from Temple University

Hospital (TUH) in the fifth method. EEG signals from the TUH dataset are converted into images by applying STFT, and their visual quality is improved by applying filters. Performance of VGG16, Resnet18 and Densenet121 is compared on EEG signal images for epilepsy detection and explanation using the SHAP method. It is found that VGG16 achieved the best classification accuracy of 97.55% using the test dataset and 97.1% ± 0.2522 using the validation dataset. Moreover, it is observed that Densenet121 produced the best results with the SHAP method. Therefore, the SHAP method is used with Densenet121 to identify the top contributing channels in the identification of epileptic patients. These channels are attached to the specific head position according to the international 10-20 system. For epilepsy patients, the top contributing channels contain seizure patterns and are attached to the brain regions from where epileptic signals originate. Thus, these channels can help neurologists identify epileptogenic brain regions and plan appropriate treatment.

The XAI methods LRP and SHAP provided explanations for the deep models' predictions and made them interpretable. This helps increase trust in the AI-based models' decisions and adoption in the clinical decision-making process. Such a system could be beneficial for several purposes, such as initial monitoring of EEG recordings and cross-verification of epilepsy diagnosis by neurologists. Moreover, EEG technologists can use such a system to analyse EEG signals while preparing reports.

Acknowledgements

Completing this PhD has been one of the most challenging yet rewarding experiences of my life, and it would not have been possible without the support of many individuals to whom I am deeply grateful.

First and foremost, I would like to express my heartfelt gratitude to my supervisor, **Dr. Jayendra M. Bhalodiya,** for his invaluable guidance and unwavering support throughout my research journey. I am sincerely grateful for the opportunity to work under his guidance. His profound knowledge, insightful feedback, and patient mentorship played a pivotal role in shaping this research. It has been a privilege to learn under his guidance, through which I have gained invaluable insights and personal growth.

I am also grateful to the members of my dissertation advisory committee for their invaluable input and for generously sharing their time and knowledge. I am immensely thankful to **Dr. Shefali Naik** for her encouragement and continuous support during this research. I would like to express my heartfelt gratitude to **Dr. Mazad Zaveri**. His invaluable guidance and constructive feedback have greatly contributed to the quality and development of this research work. I am also grateful to **Prof. Mukesh Zaveri** and **Dr. Keyur Joshi** for their thoughtful critique and valuable suggestions that have significantly enriched the quality of this work.

I would like to express my gratitude to **Prof. Sanjay Chaudhary** for his invaluable guidance and support during the PhD coursework. I am also thankful to **Prof. Mehul Raval** for his valuable input at the beginning of this research journey, which was very significant in shaping the direction and depth of this research. I am also

thankful to other faculty members and supporting staff at the School of Engineering and Applied Science, Ahmedabad University, for their kind cooperation throughout my research journey.

I would like to extend my sincere thanks to the **library staff** for their courteous assistance, prompt responses, and the well-maintained facilities that significantly supported my research work. A special thanks goes to my colleagues, friends, fellow researchers, and lab mates whose camaraderie, discussions, and support made this journey enjoyable and intellectually stimulating.

There are no words strong enough to fully capture the depth of gratitude and love I feel for my family, without whom this thesis and this entire journey would not have been possible. I would like to express my deepest gratitude to my husband, **Mr. Sunil Rathod,** for his unwavering support, endless patience, and understanding that have been my greatest strength throughout this journey. A special thank you goes to my son, **Darsheel Rathod,** and daughter, **Muskaan Rathod,** for their patience, understanding, and unconditional support. This journey, with all its late nights and tired mornings, was never meant to take me away from you. And yet, in many ways, it was done because of you, and with you always in my heart. I would like to express my heartfelt gratitude to my **Parents** for always believing in me. Your sacrifices, constant support, and love have been the foundation of everything I have achieved. I am also thankful to my **other family members** for their patience and kind support. This PhD was never mine alone. It was carried by all of you. I hope this makes you proud, because it is as much yours as it is mine.

Prajakta Rathod

Contents

Abstract ..... **i**

Contents ..... **vi**

List of Figures ..... **x**

List of Tables ..... **xv**

Abbreviations ..... **xvii**

1. Background and Motivation **1**

1.1 Causes of epilepsy 2

1.2 Symptoms of epilepsy 4

1.3 Electroencephalogram (EEG) 5

1.3.1 Epileptic EEG 5

1.3.2 Non-epileptic EEG 5

1.4 International 10-20 system 7

1.4.1 Key features of the International 10-20 system 7

1.5 Need for an automated epilepsy detection system 9

1.6 Explainable Artificial Intelligence (XAI) methods 11

1.7 Role of XAI in epilepsy detection 13

1.8 Thesis contribution 14

1.9 Thesis outline 17

2. Review of Existing Approaches in Epilepsy Detection **21**

2.1 Literature survey 21

2.1.1 Survey of literature on EEG signal pre-processing methods 21

2.1.2 Survey of literature on ML-based approaches for epilepsy detection 25

2.1.3 Survey of literature on DL-based approaches for epilepsy detection 26

2.1.4 Survey of literature on transfer learning-based approaches for epilepsy detection 29

2.1.5 Survey of literature on XAI techniques for explaining epilepsy prediction 30
2.1.6 Survey of literature on single and multi-channel datasets used for epilepsy detection studies 33
2.2 Research problem 36
2.2.1 Primary research question 38
2.2.2 Secondary research question 38
3. Methodological Framework for Epilepsy Detection and Explainability Strategy 41
3.1 Dataset 41
3.1.1 Bonn University dataset 42
3.1.2 Temple University Hospital (TUH) dataset 42
3.2 Methods 44
3.2.1 Implementation of LRP in Bi-LSTM network 44
3.2.1.1 Data preprocessing 44
3.2.1.2 Proposed method 44
3.2.1.2.1 Implementation of LRP method in Bi-LSTM network 45
3.2.1.2.2 Relevance vector generation 47
3.2.2 Implementation of LRP in VGG16 network 48
3.2.2.1 Data preprocessing 48
3.2.2.2 Proposed method 50
3.2.2.3 Layer-wise Relevance Propagation (LRP) 52
3.2.3 Comparison of the XAI methods SHAP, Guided Backpropagation and LIME 56
3.2.3.1 Data preprocessing 56
3.2.3.2 Proposed method 57
3.2.3.3 Working of XAI methods 58
3.2.3.3.1 Guided Backpropagation 59
3.2.3.3.2 LIME 60
3.2.4 SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0 61
3.2.4.1 Data preprocessing 61

3.2.4.2 Proposed method....63
3.2.4.3 Validation methods....65
3.2.5 SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channel....65
3.2.5.1 Data pre-processing....65
3.2.5.1.1 Band-pass filtering....65
3.2.5.1.2 Signal to image conversion using STFT....66
3.2.5.1.3 Channels concatenation....67
3.2.5.1.4 Image filtration....68
3.2.5.1.4.1 Sharpening....69
3.2.5.1.4.2 Contrast enhancement....70
3.2.5.2 Proposed method....71
3.2.5.2.1 Classification....72
3.2.5.2.2 Explanation....74
3.2.5.2.2.1 Working of the SHAP method....74
3.2.5.2.3 Clinical relevance of technical findings....76
3.2.5.2.3.1 Dataset preparation....77
3.2.5.2.3.2 Identification of top contributing channels....77
3.2.5.3 Validation methods....79
3.2.5.3.1 K-fold cross-validation....79
3.2.5.3.2 Heatmap evaluation using AOPC values....80
4. Epilepsy Detection Outcomes and Explainability Insights....86
4.1 LRP in Bi-LSTM network....86
4.2 LRP in VGG16 network....90
4.3 Comparison of the XAI methods SHAP, Guided Backpropagation and LIME....92
4.4 SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0....96
4.5 SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channels....101
5. Evaluation of Explainable Epilepsy Detection and Clinical Implications....109
5.1 Justification for the datasets....109

5.1.1 Bonn University dataset 109
5.1.2 TUH dataset 109
5.2 EEG patterns during epileptic episodes 110
5.3 Justification for selecting EEG signal segments of 2 sec. and 4 sec. duration 111
5.4 Discussion of the results 112
5.4.1 LRP in the Bi-LSTM network 112
5.4.2 LRP in the VGG16 network 114
5.4.3 Comparison of the XAI methods SHAP, Guided Backpropagation and LIME 116
5.4.4 SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0 117
5.4.5 SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channels 120
5.5 Limitations 133
6. Conclusion 135
6.1 Future enhancement 138
List of Publications 140
Bibliography 142

List of Figures

Fig. 1.1 Epileptic EEG signals (Shoeb 2010) 6
Fig. 1.2 Non-epileptic EEG signals (Shoeb 2010) 7
Fig. 1.3 International 10-20 system (Li et al. 2024) 8
Fig. 3.1 System architecture (Prajakta Rathod et al. 2022) 45
Fig. 3.2 Seizure signal and its spectrogram. Fig. a displays the seizure signal, and Fig. b displays its spectrogram (Prajakta Rathod et al. 2024) 49
Fig. 3.3 Non-seizure signal and its spectrogram. Fig. a shows a non-seizure signal, and Fig. b shows its spectrogram (Prajakta Rathod et al. 2024) 49
Fig. 3.4 Architecture of the proposed system (Prajakta Rathod et al. 2024) 51
Fig. 3.5 Working of the LRP method implemented in the VGG16 network (Prajakta Rathod et al. 2024) 52
Fig. 3.6 Seizure signal and its spectrogram. Fig. a shows a seizure signal of 23.6 sec. duration. It has 4097 data points over time. Fig. b shows the spectrogram of Fig. a (Rathod et al. 2025b) 56
Fig. 3.7 Non-seizure signal and its spectrogram. Fig. a shows a non-seizure signal of 23.6 sec. duration. It has 4097 data points over time. Fig. b shows the spectrogram of Fig. a (Rathod et al. 2025b) 57
Fig. 3.8 Block diagram of the system. EEG signals are segmented into 2 sec. duration pieces and converted into images using STFT. Classification is performed using the VGG16 model. The performance of three XAI methods - SHAP, Guided Backpropagation, and LIME – is compared to explain the classification results (Prajakta Rathod et al. 2025b) 57
Fig. 3.9 a, b and c represent interictal, non-seizure and seizure signals of 5 sec. duration and d, e and f represent their spectrograms, respectively. Colours in the spectrogram image represent the amplitude of a signal. The yellow colour in the image shows a higher amplitude, and the blue colour shows a lower amplitude (Prajakta Rathod et al. 2025a) 62

Fig. 3.10 a, b and c represent spectrogram images for interictal, non-seizure and epileptic seizure signals of 2 sec. duration, respectively, before normalisation, and d, e and f represent the corresponding normalised versions of these images (Prajakta Rathod et al. 2025a).....63
Fig. 3.11 Flow diagram of the system. EEG signals are preprocessed using STFT to obtain spectrogram images. Classification is performed using four CNN models, and the explainable AI method SHAP is used for explaining the classification results (Rathod et al. 2025a).....64
Fig. 3.12 (a) represents an epileptic signal of 4 sec. duration with 21 channels used in the study. (b) represents its spectrogram. The channels in the spectrogram are in the same sequence as that of Fig. 3.12 (a). Colours in the spectrogram image represent the power of a signal at different frequencies in different channels. The red colour in the spectrogram image shows a higher power, and the blue colour shows a lower power (Prajakta Rathod et al. 2025) 68
Fig. 3.13 (a) represents a non-epileptic signal of 4 sec. duration with 21 channels used in the study. (b) represents its spectrogram. The channels in the spectrogram are in the same sequence as that of Fig. 3.13 (a). Colours in the spectrogram image represent the power of a signal at different frequencies in different channels. The red colour in the spectrogram image shows a higher power, and the blue colour shows a lower power (Prajakta Rathod et al. 2025) 69
Fig. 3.14 (a) and (d) show epileptic and non-epileptic spectrogram images, respectively. Fig. (b) and (e) show epileptic and non-epileptic spectrogram images after sharpening. Fig. (c) and (f) show epileptic and non-epileptic spectrogram images after applying contrast enhancement (Prajakta Rathod et al. 2025).....72
Fig. 3.15 Flow diagram of the proposed system. EEG signals are pre-processed using STFT to obtain spectrogram images, which are then filtered to enhance their visual quality. Three CNN models - Densenet121, Resnet18 and VGG16- are used for classification, and the explainable AI method SHAP is used for explaining the classification results. The heatmap generated using the SHAP method is evaluated using the AOPC value. Top contributing channels in the classification of epileptic and non-epileptic patients are identified using the relevance matrix generated by the SHAP method (Prajakta Rathod et al. 2025) 73

Fig. 3.16 Architecture of the VGG16 network. Network layers are trained on image datasets of epileptic and non-epileptic EEG signal images. The output layer is changed to identify two classes (Prajakta Rathod et al. 2025).....74
Fig. 3.17 Identification of the top contributing channels using the relevance matrix generated by the SHAP method. In step 1, all positive relevance values from a particular channel are considered. In step 2, all positive relevance values of a particular channel of all images of one patient are added. In step 3, these values are sorted in descending order, and the median value is obtained. Channels with relevance value above the median are considered as the top contributing channels (Prajakta Rathod et al. 2025) 78
Fig. 4.1 Sample epileptic signals from the dataset. Signal data over time is plotted along the X-axis, and signal amplitude is shown on the Y-axis. The duration of the signal is 5.76 sec. i.e., 1000 data points (Prajakta Rathod et al. 2022).....87
Fig. 4.2 LRP results for sample epileptic signals in Fig. 4.1. Signals are marked in blue colour, and LRP results are marked in red colour, indicating part of the signal classifying it in the epileptic class (Prajakta Rathod et al. 2022) 87
Fig. 4.3 Sample non-epileptic signals from the dataset. Signal data over time is plotted along the X-axis, and signal amplitude is shown on the Y-axis. The duration of the signal is 5.76 sec. i.e., 1000 data points (Prajakta Rathod et al. 2022).....88
Fig. 4.4 LRP results for sample non-epileptic signals in Fig. 4.3. Signals are marked in blue colour, and LRP results are marked in red colour, indicating part of the signal classifying it in the non-epileptic class (Prajakta Rathod et al. 2022).....88
Fig. 4.5 Confusion matrix generated by the VGG16 network for the classification of epileptic seizure and non-seizure class images (Prajakta Rathod et al. 2024).....90
Fig. 4.6 LRP results for epileptic seizure class images. Fig. a, b and c show seizure signal images, and Fig. d, e and f show LRP results for images a, b and c, respectively (Prajakta Rathod et al. 2024) 91
Fig. 4.7 LRP results for non-seizure class images. Fig. a, b and c show non-seizure signal images, and Fig. d, e and f show LRP results for images a, b and c, respectively (Prajakta Rathod et al. 2024) 92
Fig. 4.8 Explanation of classification results produced by three XAI methods for seizure class images. Five sample images from the seizure class are shown. XAI

methods used are SHAP, Guided Backpropagation and LIME (Prajakta Rathod et al. 2025b) 94
Fig. 4.9 Explanation of classification results produced by three XAI methods for non-seizure class images. Five sample images from the non-seizure class are shown. XAI methods used are SHAP, Guided Backpropagation and LIME (Prajakta Rathod et al. 2025b) 95
Fig. 4.10 Sensitivity graph for four CNN models. The X-axis represents four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0 and three classes: interictal, non-seizure and seizure. The Y-axis represents the sensitivity value (Prajakta Rathod et al. 2025a) 98
Fig. 4.11 Specificity graph for four CNN models. The X-axis represents four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0 and three classes: interictal, non-seizure and seizure. The Y-axis represents the specificity value (Prajakta Rathod et al. 2025a) 99
Fig. 4.12 Accuracy graph for four CNN models. The X-axis represents four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0. The Y-axis represents the accuracy value (Prajakta Rathod et al. 2025a) 99
Fig. 4.13 Confusion matrix for four CNN models, Resnet18, Densnet121, VGG16 and Squeezenet1_0 obtained for three classes: interictal, non-seizure and seizure (Prajakta Rathod et al. 2025a) 100
Fig. 4.14 SHAP method results for four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0. Two sample images from each of the three classes, interictal, non-seizure and seizure, are shown. The SHAP method marks positively contributing features in pink shade and negatively contributing features in blue shade for classification in a particular class (Prajakta Rathod et al. 2025a) 101
Fig. 4.15 a, b and c show the accuracy, sensitivity and specificity graphs, respectively, for Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025) 103
Fig. 4.16 Confusion matrix for three CNN models, Densenet121, Resnet18 and VGG16, obtained for two classes, epileptic and non-epileptic. (Prajakta Rathod et al. 2025) 103
Fig. 4.17 SHAP method results for epileptic signal images using three CNN models: Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025) 104

Fig. 4.18 SHAP method results for non-epileptic signal images using three CNN models: Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025).....105
Fig. 4.19 AOPC curves obtained for all three networks: Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025).....106
Fig. 5.1 Median plot of the sum of relevance values of all channels as described in Section 3.2.5.2.3.2 (Prajakta Rathod et al. 2025) 123
Fig. 5.2 ROC curve for Densenet121, Resnet18 and VGG16 126

List of Tables

Table 3.1 Bonn University Dataset Description 42
Table 3.2 Data from the TUH Dataset Used in the Research Study (Prajakta Rathod et al. 2025) 43
Table 3.3 Dataset Description (Prajakta Rathod et al. 2022).....44
Table 3.4 Data Used in Research Study (Prajakta Rathod et al. 2024).....48
Table 3.5 Data Used in Research Study (Rathod et al. 2025b).....56
Table 3.6 Data Used in Research Study (Prajakta Rathod et al. 2025a).....61
Table 4.1 Classification Results for VGG16 Network (Prajakta Rathod et al. 2024) .90
Table 4.2 Classification Results for VGG16 Network (Prajakta Rathod et al. 2025b)93
Table 4.3 5-Fold Cross-Validation Results of Four Pre-Trained CNN Models Using Validation Dataset (Prajakta Rathod et al. 2025a) 96
Table 4.4 Results of Four CNN Models on Test Dataset (Prajakta Rathod et al. 2025a) 97
Table 4.5 Classification Results of Four Pre-Trained CNN Models (Prajakta Rathod et al. 2025a) 97
Table 4.6 Performance of Three CNN Models on the Validation Dataset Using 5-Fold Cross-Validation and on the Test Dataset (Prajakta Rathod et al. 2025) 102
Table 4.7 K-Fold Cross-Validation Results of VGG16 Network on the Validation Dataset for k=5 and k=10 (Prajakta Rathod et al. 2025) 102
Table 4.8 Classification Results of Three CNN Models, Densenet121, Resnet18 and VGG16, for Epileptic and Non-epileptic Signal Images (Prajakta Rathod et al. 2025) 102
Table 4.9 Evaluation of the Heatmap of Three CNN Models Using AOPC Value (Prajakta Rathod et al. 2025) 106
Table 4.10 Top Contributing EEG Channels in Epileptic and Non-epileptic Patients' Classification (Prajakta Rathod et al. 2025) 106

Table 5.1 Comparison with Existing Systems Using DL Methods (Prajakta Rathod et al. 2022) 113
Table 5.2 Comparison with Existing Systems Using the Same Dataset (Prajakta Rathod et al. 2024) 116
Table 5.3 Comparison with Existing Systems Using the Same Dataset (Prajakta Rathod et al. 2025b) 117
Table 5.4 Comparison with Existing Systems Using the Same Dataset in Terms of Classification Accuracy (Prajakta Rathod et al. 2025a) 119
Table 5.5 Comparison with Existing Systems Using Transfer Learning Approach (Prajakta Rathod et al. 2025a)..... 120
Table 5.6 Area Under the ROC Curve for Densenet121, Resnet18 and VGG16 126
Table 5.7 Comparison with Existing Systems Using the TUEP Dataset in Terms of Classification Accuracy (Prajakta Rathod et al. 2025) 127

Abbreviations

Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AOPC	Area Over Perturbation Curve
AUROC	Area Under ROC Curve
Bi-LSTM	Bidirectional Long Short-Term Memory
CHB-MIT	Children's Hospital Boston - Massachusetts Institute of Technology
CNN	Convolutional Neural Network
CNS	Central Nervous System
CVANN	Complex-valued Neural Network
CWT	Continuous Wavelet Transform
DCAE	Deep Convolutional Autoencoder
DCNN	Deep Convolutional Neural Network
DeepLIFT	Deep Learning Important Features
DL	Deep Learning
DNA	Deoxyribonucleic Acid
DNN	Deep Neural Network
DT	Decision Tree
DWT	Discrete Wavelet Transform
EDF	European Data Format
EEG	Electroencephalogram
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
FPR	False Positive Rate
GB	Gradient Boosting
GCNN	Graph Convolutional Neural Network
HNN	Hierarchical Neural Network
KNN	K-Nearest Neighbours

LIME	Local Interpretable Model-agnostic Explanations
LR	Logistic Regression
LRP	Layer-wise Relevance Propagation
LSTM	Long Short-Term Memory
MAS	Mean Amplitude Spectrum
ML	Machine Learning
MLP	Multi-Layer Perceptron
NB	Naive Bayes
NLP	Natural Language Processing
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
PSD	Power Spectral Density
PSDED	Power Spectrum Density Energy Diagrams
ReLU	Rectified Linear Unit
REM	Rapid Eye Movement
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
RPS	Ramanujan Periodic Subspace
SHAP	SHapley Additive exPlanations
STFT	Short-time Fourier Transform
SVM	Support Vector Machines
TPR	True Positive Rate
TUH	Temple University Hospital
XAI	Explainable Artificial Intelligence

1. Background and Motivation

Epilepsy is a pervasive neurological disorder worldwide (Fisher et al. 2014). Approximately 50 million individuals worldwide are affected by epilepsy. It is approximately 1% of the global population. Every year, 5 million new epilepsy patients are diagnosed globally. It is primarily found in low and middle-income countries with minimal healthcare facilities (Espinosa-Jovel et al. 2018b). Additionally, it is most commonly observed in individuals younger than two years and older than 65 years (Wojewodka et al. 2021). However, it can develop at any age. Medication can help around 70% of epilepsy patients control their seizure attacks (Lee 2014). However, approximately 30% of epilepsy patients may require surgery to control their seizure attacks (Munakomi and Das 2025). Suppose epilepsy patients do not receive proper treatment or medication on time, it may cause accidents, injuries, or other health-related problems, and sometimes it may result in sudden death as well. It may increase the mortality rate for epilepsy patients. Generally, epilepsy patients have a higher mortality rate than the normal population. The prevalence of epilepsy is not the same globally. It is different in different regions. Rural areas often have a higher prevalence than urban areas. Moreover, low and middle-income countries report higher cases of epilepsy than high-income countries. Specifically, these regions accounted for 82.1% of incident cases, 80.4% of prevalent cases and 84.7% of epilepsy-related mortality (“Global, Regional, and National Burden of Epilepsy, 1990–2021” 2025). In low and middle-

income countries, epilepsy may not be diagnosed on time or treated properly due to the lack of availability of medical facilities. It may result in a higher prevalence of epilepsy in these countries. Thus, epilepsy is becoming a global health issue, and timely diagnosis and treatment are required to control the prevalence of this disorder. The automated epilepsy detection system has many advantages over manual epilepsy detection systems.

1.1 Causes of epilepsy

Epilepsy is a chronic neurological condition involving abnormal brain function, which manifests as spontaneous and recurrent seizures. In this disorder, the brain's electrical activity suddenly becomes abnormal, impacting a patient's movement, behaviour, sensation, or consciousness. There can be many causes for developing epilepsy, which may vary from individual to individual. The main causes of epilepsy include brain injury, genetic factors, tumours, infections, or other health conditions (Balestrini et al. 2021). However, in many cases, the leading cause of epilepsy is unknown. Genetic mutations, i.e., changes in the Deoxyribonucleic Acid (DNA) sequence of an organism, can result in epileptic seizures. Some genetic disorders, like Dravet syndrome, can cause epileptic seizures (Balestrini et al. 2021). Epilepsy caused by genetic factors may run in families. Epilepsy can also be caused by brain tissue damage during brain injury (Balestrini et al. 2021). A condition of the occurrence of a stroke in which blood flow to the brain is interrupted may cause epilepsy (Zöllner et al. 2021). The abnormal development of brain cells can cause epilepsy (Bozzi et al. 2012). Lack of oxygen at the time of birth can also result in the occurrence of epileptic seizures at a later stage in life (Gailus et al. 2021). Infection affecting the brain parts or their

surrounding areas is also one of the causes of epilepsy (Balestrini et al. 2021). Brain tumours can also result in epileptic seizure attacks (Balestrini et al. 2021). Neurodegenerative diseases, which involve the degeneration of brain cells or other forms of dementia, also increase the risk of epilepsy (Balestrini et al. 2021). Certain metabolic or biochemical conditions, such as low blood sugar and sodium or calcium level imbalances, can also cause epileptic seizures (Bartolini et al. 2023) (Nardone et al. 2016). Some drugs that affect the Central Nervous System (CNS) can provoke epileptic seizures (Balestrini et al. 2021). Epileptic seizures can also be triggered by the use of certain recreational drugs (Wolfe et al. 2019). In rare cases, high fever in children may result in epilepsy at a later age (Park and Yoo 2025). Multiple sclerosis is a disease in which the protective covering of nerves breaks down. It can cause lesions on the brain and spinal cord, which may result in seizures (Antal et al. 2022). In some individuals, lack of sleep may result in an increased likelihood of the occurrence of seizures (Dell'Aquila and Soti 2022). In some individuals, certain visual patterns, flashing lights, or video games can trigger seizures (Ricci et al. 1998). Sometimes, emotional or physical stress can trigger seizures (Jhaveri et al. 2023). These are some of the leading causes of epileptic seizures in individuals.

The most common causes of epilepsy vary with age and geography. In younger patients, CNS infections are the most common cause of new-onset seizures, followed by metabolic causes in descending order. In the middle-aged group, stroke is the most common cause, followed by CNS infections and metabolic causes, in descending order. In elderly patients, stroke is the leading cause of new-onset seizures (Bhatia et al. 2022). In high-income countries, stroke is the predominant cause of epilepsy in older adults

(Galovic et al. 2021), whereas in low- and middle-income countries, infections and birth-related injuries play a more significant role (Espinosa-Jovel et al. 2018a).

1.2 Symptoms of epilepsy

The most common symptom of epilepsy is seizures. Focal Seizures originate in one specific area of the brain. In focal aware seizures, a person has awareness of their surroundings. Its symptoms include uncontrolled jerking movements of the hand or face, unusual smells, tastes, or visual distortions; the person may have a feeling of fear or anxiety, nausea, or dizziness (Stafstrom and Carmant 2015). In focal impaired awareness seizures, the person may lose awareness, stare blankly for some time, engage in lip-smacking or hand-wringing, may have confusion or disorientation, and not respond to others during the seizure attack (Stafstrom and Carmant 2015). Generalised seizures simultaneously impact both cerebral hemispheres, and the patient may lose consciousness, his body may get stiff, he may have jerking or rhythmic shaking, he may lose bladder control, or he may bite his tongue (Stafstrom and Carmant 2015). Moreover, a person may feel confused or tired after a seizure. Absence seizures are mostly found in children, in which they may stare for a short time, lose awareness, or may have blinking or lip-smacking (Stafstrom and Carmant 2015). In the case of atonic seizures, there is a sudden loss of muscle tone, due to which a person may collapse or drop the head or limbs suddenly (Stafstrom and Carmant 2015). In myoclonic seizures, there is jerking of the arms or legs (Stafstrom and Carmant 2015). In tonic seizures, the muscle gets stiff and rigid (Stafstrom and Carmant 2015). This thesis does not address one particular type of seizure; rather, it focuses on patient-specific signal analysis, which is elaborated in Chapters 3, 4, and 5.

1.3 Electroencephalogram (EEG)

The brain's electrical activity is recorded through a medical procedure called an EEG (Rayi and Murr 2025). During this test, small electrodes are attached to the scalp. These electrodes detect electrical impulses through which brain cells communicate with each other. This EEG test is carried out in diagnosing epilepsy, seizures, sleep disorders and head injuries, and detects the location of brain tumours (Rayi and Murr 2025). It is also used to assess the level of brain activity in case of coma. The patterns displayed during EEG tests vary depending on the condition of a person. It differs in different states, like awake, sleeping, thinking, relaxed, performing mental tasks like solving a puzzle or mathematical sum, or focused (Soufineyestani et al. 2020).

1.3.1 Epileptic EEG

Epileptic EEG can show the following patterns.

- Abnormal brain waves: EEG of abnormal brain waves can have patterns like sharp waves, spikes, or spike-and-wave discharges (Emmady and Anilkumar 2025).
- Seizure activity: EEG recorded during a seizure attack shows abnormal brain activity, like spikes and sharp waves (Li et al. 2020).
- Interictal activity: EEG recorded during the interictal period, i.e., the period between two consecutive seizures, shows spikes or sharp waves (Emmady and Anilkumar 2025).

1.3.2 Non-epileptic EEG

A non-epileptic EEG can show the following patterns

- Normal brain activity: EEG recorded during normal brain activity can show a rhythmic pattern. These patterns can have alpha, beta, theta, and delta waves with different frequencies (Nayak and Anilkumar 2025).
- Psychogenic or other neurological conditions: EEG recorded during psychogenic or other neurological conditions, like sleep disorders, stroke, or brain injuries, can show abnormalities that are not related to epilepsy. These abnormalities do not show epileptic patterns like spikes or sharp waves (Huff et al. 2025).

Fig. 1.1 shows the EEG for epileptic seizure signals, and Fig. 1.2 shows the EEG for non-epileptic signals.

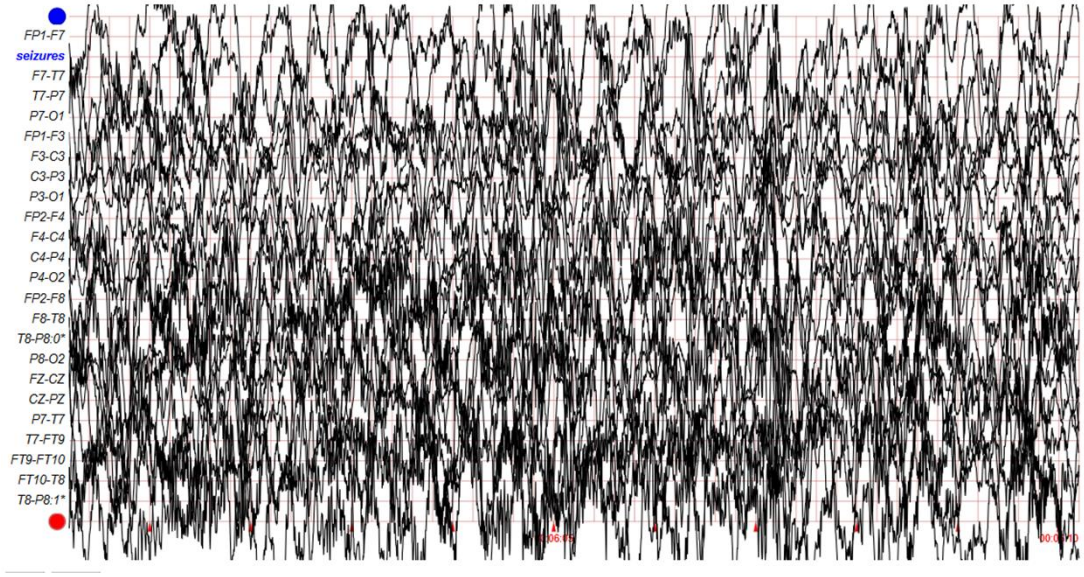


Fig. 1.1 Epileptic EEG signals (Shoeb 2010)

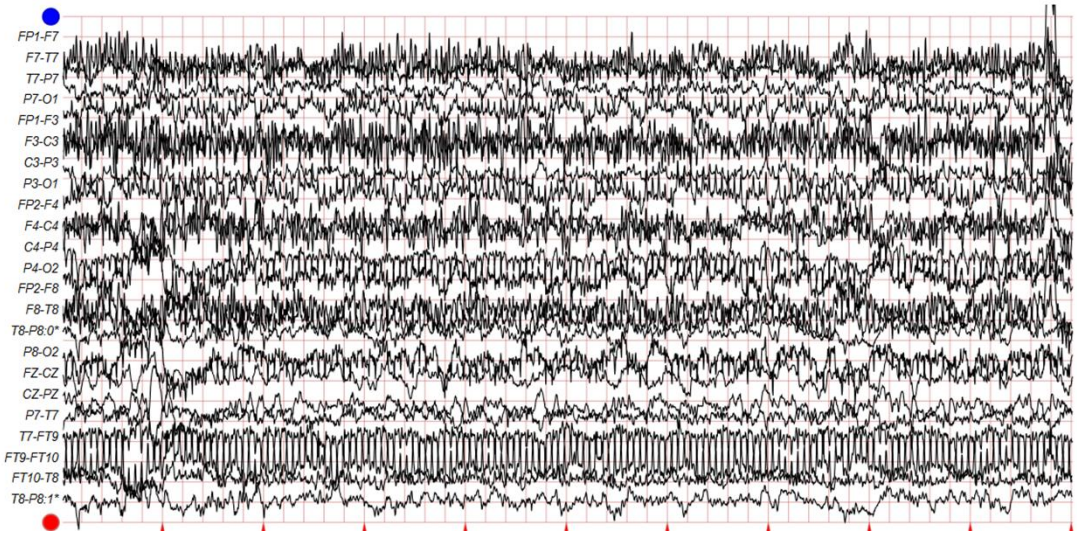


Fig. 1.2 Non-epileptic EEG signals (Shoeb 2010)

1.4 International 10-20 system

The International 10-20 system is a universally accepted method for identifying and positioning electrodes on the scalp during EEG recordings and other neurophysiological evaluations (Soufineyestani et al. 2020). This system ensures that the study outcomes of a subject can be effectively analysed and compared. Fig. 1.3 shows the electrode placement in the International 10-20 system.

1.4.1 Key features of the International 10-20 system

The International 10-20 system has two main reference points, nasion and the inion, as shown in Fig. 1.3. The nasion is a distinct anatomical landmark located at the junction between the forehead and the nasal bridge. The inion is the bony prominence at the back of the skull. The International 10-20 System is so named because the spacing between adjacent electrodes corresponds to either 10% or 20% of the total distance along the anterior–posterior or lateral axis of the skull (Soufineyestani et al. 2020).

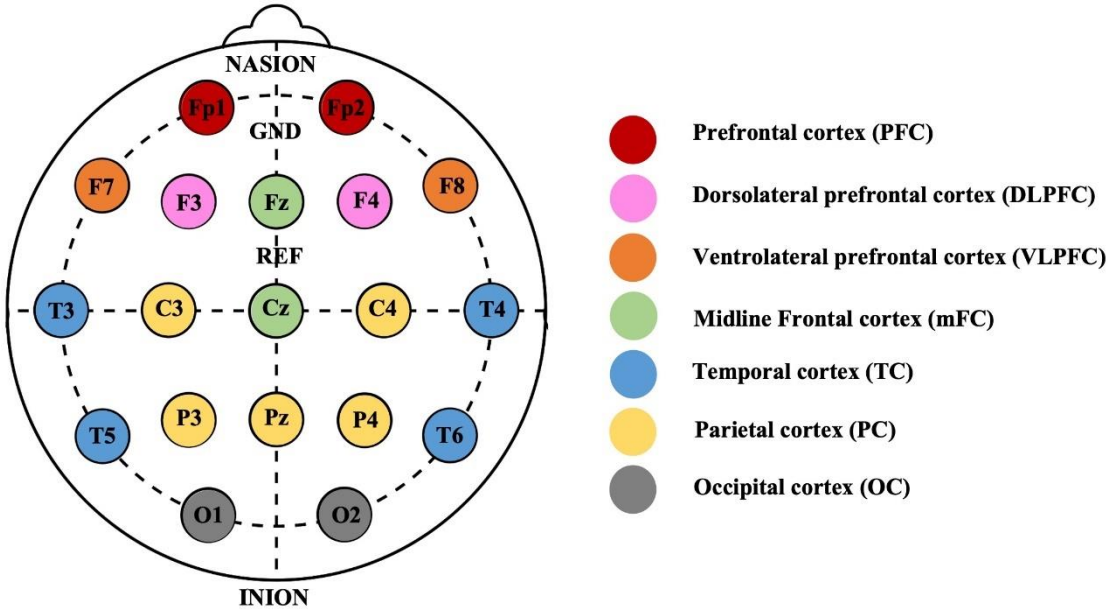


Fig. 1.3 International 10-20 system (Li et al. 2024)

The electrodes are attached to the scalp at specific positions. The brain is anatomically divided into distinct regions known as the prefrontal, frontal, temporal, parietal, occipital, and central lobes.

The names of the electrodes denote the brain region, like

- F - Frontal
- C - Central
- P - Parietal
- O - Occipital
- T - Temporal
- A - Auricular or earlobe reference

Electrodes marked with even numbers (such as 2, 4, 6, and 8) are positioned over the right hemisphere of the brain, while those marked with odd numbers (such as 1, 3, 5, and 7) are placed over the left hemisphere. This numerical labelling helps ensure

consistent and symmetric electrode positioning across both sides of the scalp for accurate neurophysiological measurements. Electrodes named Cz and Fz are 'ground' or 'common' reference points for all other electrodes. Electrodes A1 and A2 are used as contralateral referencing electrodes for other EEG electrodes.

The international 10-20 system is bilaterally symmetric. This arrangement places all the electrodes consistently and systematically. It helps to identify brain regions with abnormal electrical activity. The results generated using this system can be compared across different patients, labs, and research studies.

1.5 Need for an automated epilepsy detection system

For epilepsy diagnosis in primary and secondary care settings, manual epilepsy detection is considered the gold standard, in which neurologists and EEG technologists examine EEG recordings manually to identify epileptic seizure patterns such as spikes, sharp waves, or other abnormalities (Myers et al. 2025). In this case, the outcome is based on the clinician's expertise, experience, and subjective judgment. As these signals have enormous sequences, it may cause subjective errors for continuous manual tracking, whereas in automated epilepsy detection, the system automatically processes EEG data for a long time and identifies seizure patterns, which can be verified by the subject matter expert. These systems use Machine Learning (ML) algorithms and signal processing techniques to analyse EEG data and identify epileptic patterns.

Current Artificial Intelligence (AI) systems for epilepsy diagnosis provide many benefits in the analysis of EEG signals and seizure prediction, but they have several limitations that restrict their widespread clinical use. A major challenge lies in data-related issues. The development of robust AI models requires large, diverse, and well-

annotated datasets, but such datasets are limited. In addition, EEG recordings vary across hospitals (different equipment, sampling rates, electrode placements), making model transferability difficult. Manual annotation of EEGs by experts is time-consuming, and inter-rater variability affects the consistency of labelling. Furthermore, artefacts like muscle movements, eye blinks, or external noise present in the EEG can be misinterpreted as a seizure pattern by the models. Clinical integration presents another major limitation. Many AI models, especially those based on Deep Learning (DL), are considered "black-box" systems, offering little interpretability or transparency, which makes clinicians hesitant to rely on them. Moreover, sharing sensitive EEG and clinical data for training models raises privacy/security issues.

Although AI-based systems for epilepsy detection have certain limitations, they provide notable advantages compared to traditional manual detection methods. Manual epilepsy detection can have some drawbacks, such as the same EEG being interpreted differently by different clinicians. It depends on the expertise and fatigue level of the clinician. Automated epilepsy detection systems use modern AI techniques and generate consistent interpretations with high performance. However, the predictive performance of the system is highly influenced by the selected ML algorithms and the quality, quantity, and diversity of the training data used during model development. In manual epilepsy detection, analysing longer EEG recordings may take more time and cause a delay in epilepsy diagnosis. However, automated systems can analyse longer EEG data in less time, resulting in efficient and accurate diagnosis of the disease. Automated systems are useful in busy and demanding workplaces and remote areas with limited medical facilities and specialist access. EEG signals may contain subtle or complex epileptic patterns that can be missed in manual observation. However, such

patterns can be easily identified by automated systems with the help of advanced AI techniques. Moreover, automated systems are scalable and can be deployed in various settings. Manual systems are highly dependent on domain experts, whereas automated systems reduce dependency on domain experts for continuous monitoring. Manual systems may have high operational costs when long-term monitoring and repeated evaluation are needed. In the case of AI-based automated systems, only the initial cost for development and deployment is high, but it reduces the cost for labour-intensive analysis in the long run without affecting the quality of assessment and patient care. Real-time onset seizure detection is difficult with manual systems, as it needs constant human attention. Automated systems are beneficial in such situations where real-time monitoring of EEG data is possible for seizure onset detection. Thus, automated systems are well-suited for initial monitoring of EEG, real-time EEG monitoring, and analysing and interpreting long EEG recordings. Moreover, it can be used by experts to cross-verify disease predictions and confirm the diagnosis. Currently, some AI-based automated epilepsy detection systems are available in the market and used by medical practitioners and epilepsy patients (Emoryy 2025) (“Neuro Event Labs - Seizures. Clearly,” n.d.) (Psaizi, n.d.).

1.6 Explainable Artificial Intelligence (XAI) methods

AI models surpass human capabilities and achieve remarkable performance in multiple domains, including healthcare, finance, retail and e-commerce, transportation and autonomous vehicles, manufacturing industry, agriculture, education, entertainment and media, Natural Language Processing (NLP) applications, cybersecurity, environment and sustainability, and scientific research (Rashid and

Kausik 2024). Day by day, the architecture of these AI models is improving and becoming more complicated and deeper. But at the same time, many of these models are black boxes with no transparency in their decisions. XAI techniques provide transparency into the internal reasoning of AI models, allowing users to comprehend how specific decisions are made. The XAI methods help make AI systems' output interpretable and explainable to humans (Yang 2022) (Linardatos et al. 2020) (Das and Rad 2020) (Pocevičiūtė et al. 2020) (Samek et al. 2021). XAI methods provide the following benefits when used with the AI models:

1. **Transparency:** XAI methods make the model's internal working and decision-making process more transparent.
2. **Interpretability:** XAI methods provide insight into why the model has arrived at a particular decision.
3. **Accountability:** XAI methods enable stakeholders to assess the model outputs' fairness, bias, and reliability.
4. **User-centric design:** Tailor the explanations to the knowledge level and needs of different users, like developers, end-users, etc.

XAI methods help developers identify flaws in AI systems and fix them. By uncovering and addressing algorithmic biases, these methods contribute to the development of fair and equitable AI systems. Transparency in the decisions of AI systems makes them trustworthy and increases the chance of the adoption of these systems in the industry. XAI methods can be applicable in many domains, like healthcare, finance, law enforcement, autonomous vehicles, and human resources (Yang 2022) (Linardatos et al. 2020) (Das and Rad 2020) (Pocevičiūtė et al. 2020) (Samek et al. 2021). Using XAI makes AI systems more robust and helps build user

confidence in AI systems' decisions, especially in critical domains like healthcare, finance, and law enforcement.

1.7 Role of XAI in epilepsy detection

XAI enhances the effectiveness of AI-driven epilepsy detection systems by providing several advantages, such as enhancing transparency, improving trustworthiness, and facilitating the acceptance and adoption of these systems in clinical practice. Clinicians and neuroconsultants may be doubtful about AI predictions and diagnoses for epilepsy disorders without understanding the reasons behind the predictions made by AI systems. Thus, to bridge this gap, XAI methods can provide visual explanations for identifying a person as an epileptic patient. These methods can highlight features like epileptic pattern, amplitude, and frequency of EEG waves that contribute to the identification of a person as an epileptic patient. It helps clinicians to cross-verify and trust the decisions of the AI model. It can provide better collaboration between AI developers and medical practitioners.

XAI methods can identify specific biomarkers, such as specific wave patterns in the EEG, which provide clinicians with meaningful data in the EEG. EEG signals may have different epileptic patterns depending on the type of epilepsy with which a patient is suffering. XAI methods can highlight patient-specific epileptic patterns in EEG, making the system more personalised. Moreover, XAI methods can reveal flaws in the AI systems if they identify artefacts or noise in the EEG as an epileptic pattern. It helps researchers and developers to correct and validate AI predictions. Moreover, XAI methods can act as an educational tool to understand the processing of the EEG signals for an epilepsy diagnosis. The interpretations provided by XAI methods can

help address ethical concerns and meet regulatory standards by justifying the use of AI in epilepsy diagnosis. XAI methods can explain the subtle preictal seizure pattern identification, which might be neglected otherwise. This can help early-stage prediction of seizure onset and can help provide the necessary action.

1.8 Thesis contribution

The primary objective of this research is to develop a clinically acceptable automated system for the accurate classification of epileptic and non-epileptic patients using Deep Neural Networks (DNNs). The study aims to integrate XAI methods to interpret the model's prediction and identify the most influential EEG channels contributing to the detection of epilepsy. By combining the predictive power of DNN architectures with the interpretability offered by XAI techniques such as SHapley Additive exPlanations (SHAP) and Layer-wise Relevance Propagation (LRP), the proposed system seeks to achieve both high diagnostic accuracy and clinical transparency, thereby supporting neurologists in decision-making and enhancing trust in AI-assisted epilepsy diagnosis.

Epilepsy is a chronic neurological disorder characterised by recurrent seizures, and accurate, timely diagnosis is critical for effective treatment and patient management. Manual inspection of EEG signals by neurologists is time-consuming, subjective, and prone to human error, which emphasises the urgent need for an automated and interpretable epilepsy detection system. Converting EEG signals into image representations helps to achieve higher model accuracy. This transformation facilitates the learning of complex patterns that may be difficult to detect in raw time-series form. Compared to traditional ML methods, DNNs eliminate the need for manual

feature extraction, allowing for improved generalisation across diverse subjects and recording conditions. Incorporating transfer learning allows pre-trained models on large-scale datasets to be fine-tuned on EEG images, effectively reducing computational costs while accelerating convergence. To ensure trust and clinical acceptance, the study integrates XAI techniques, such as LRP and SHAP, to visualise the salient features and channels that contribute most to model predictions, thereby enhancing interpretability and transparency. These insights can help neurologists validate automated decisions and identify biomarkers relevant to epileptic activity. The research utilises benchmark EEG datasets, namely the Bonn University dataset for controlled laboratory validation and the Temple University Hospital (TUH) dataset for real-world, large-scale testing, ensuring both reliability and robustness.

EEG has long been the most widely used tool for detecting abnormal EEG patterns, as it provides direct insights into the electrical activity of the brain through multiple scalp electrodes, each representing a specific cortical region. However, not all EEG channels contribute equally to the detection of epileptic activity. While certain electrodes capture background rhythms, others are more directly associated with epileptiform discharges and seizure onset zones, which are crucial for distinguishing epileptic patients from non-epileptic individuals. Identifying these channels is particularly important for localising the epileptogenic brain regions that drive seizure generation, thereby aiding both diagnosis and surgical planning in drug-resistant epilepsy.

Thus, identifying the top contributing EEG channels in determining epileptic and non-epileptic patients establishes a clinically meaningful link to epileptogenic brain regions, bridging the gap between computational modelling and practical neurological

application. Overall, this work seeks to bridge the gap between high-performance automated epilepsy detection and clinical interpretability, contributing to a new generation of trustworthy, explainable, and data-efficient diagnostic systems that can support clinicians in making faster and more reliable epilepsy diagnoses.

This thesis makes the following primary contributions:

I. Work done on the Bonn University EEG dataset, which is a single-channel dataset

1) An automated epileptic seizure detection using a Bidirectional Long Short-Term Memory (Bi-LSTM) network, and the model's decisions were interpreted using the LRP explainability method. The system reached a classification accuracy of 87.25%.
2) Automated epileptic seizure signal identification with the help of the VGG16 network using transfer learning with a fine-tuning approach and explanation of results using the XAI method, LRP. The model achieved a higher classification accuracy of 99.89%, which is improved by converting EEG signals into images.
3) Automated epileptic seizure signal identification with the help of the VGG16 network using transfer learning with a fine-tuning approach. Comparative analysis is performed for the three XAI methods - SHAP, guided backpropagation and Local Interpretable Model-agnostic Explanations (LIME).
4) Comparative analysis for epileptic seizure signal identification using four DNNs - Resnet18, Densenet121, VGG16 and Squeezenet1_0. The accuracy

achieved by Resnet18, Densenet121, VGG16 and Squeezenet1_0 is 99.7%, 99.62%, 99.47% and 99.17%. Moreover, the results given by four models with the XAI method SHAP are also compared.

II. Work done on the multi-channel EEG dataset from TUH

1) Signals are converted into spectrogram images, and the visual inspection of the images is improved by applying filters
2) The performance of three DNNs - Densenet121, Resnet18 and VGG16- is compared for epileptic signal identification and explanation using the XAI method SHAP.
3) Patient-specific top contributing EEG channels in the classification decision are identified using the XAI method SHAP, which can help locate the epileptogenic brain regions of epilepsy patients.

1.9 Thesis outline

This section describes the organisation of the thesis contents chapter-wise.

Chapter 1: This chapter gives general information about the epilepsy disorder like, its prevalence, causes, symptoms, etc. It also describes the patterns observed in epileptic and non-epileptic EEG signals and the International 10-20 system used to record EEG. Moreover, it discusses the need for an automated epilepsy detection system and the significance of XAI in these systems. Further, this chapter provides a short overview of the research objective, thesis contribution, and the organisation of the thesis contents chapter-wise.

Chapter 2: This chapter provides a survey of literature on EEG signal pre-processing methods, ML-based approaches, DL-based approaches, and transfer learning approaches used for epilepsy detection. Various XAI techniques used for explaining the detection of epilepsy, and single and multi-channel datasets used in epilepsy detection research are reviewed. Further, it discusses the primary and secondary research questions with research objectives identified during the literature review.

Chapter 3: This chapter elaborates on the datasets used and the five new methods developed. These methods are: (i) implementation of LRP in Bi-LSTM network, (ii) implementation of LRP in VGG16 network, (iii) comparison of the XAI methods SHAP, Guided Backpropagation and LIME, (iv) SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0 and (v) SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channels.

Chapter 4: This chapter provides a detailed analysis of the experimental outcomes generated by the newly developed methods introduced in Chapter 3. It provides a comparative analysis of the XAI methods SHAP, Guided Backpropagation and LIME. Moreover, it describes the performance analysis of four DNNs - Resnet18, Densenet121, VGG16 and Squeezenet1_0. Further, it elaborates on the use of the SHAP method for the identification of the top contributing channels, which in turn identify the epileptogenic brain regions.

Chapter 5: This chapter discusses the findings observed during the results analysis of newly developed methods in Chapter 4. This chapter compares the results achieved by

the new methods with the results of existing epilepsy detection systems. It also discusses how the integration of XAI techniques can support the adoption of new methods in clinics for epilepsy detection. Moreover. it describes the limitations of these methods.

Chapter 6: This chapter provides a short overview of the new methods developed. It lists out the comparison results of various XAI techniques, i.e., LRP, SHAP, Guided backpropagation, and LIME. Moreover, the results of performance comparisons of four DNNs, i.e., Resnet18, Densenet121, VGG16 and Squeezenet1_0, are listed. Furthermore, the future enhancements section discusses the future work that can further improve the developed methods.

At last, the bibliography is provided.

Summary:

Epilepsy is a neurological disorder that is particularly prevalent in low- and middle-income countries. It is commonly diagnosed using EEG, where electrodes are placed on the patient's scalp according to the International 10–20 system to ensure standardised electrode placement and reliable signal acquisition. However, manual interpretation of EEG signals for seizure detection is labour-intensive, time-consuming, and prone to human error, highlighting the need for automated seizure detection systems. DNNs have shown great potential in this domain, but their black-box nature limits their interpretability. To address this limitation, XAI methods are employed to

interpret and justify the decisions of DNNs, thereby enhancing the transparency, trustworthiness, and clinical acceptability of AI-based epilepsy detection systems.

2. Review of Existing Approaches in Epilepsy Detection

This chapter reviews seizure detection techniques and epilepsy monitoring methods. It includes clinical biomarkers, manual assessment, AI-based and XAI-based methods. The methods that utilise EEG data are focused on according to the research scope. Various pre-processing techniques used to process EEG data are explored. The implementation of various non-DL and DL-based systems for epilepsy detection has been reviewed. The use of various XAI techniques to explain the prediction of epileptic signals has been reviewed. Moreover, the research studies that utilise the widely used EEG datasets, the Bonn University and TUH datasets, are reviewed.

2.1 Literature survey

2.1.1 Survey of literature on EEG signal pre-processing methods

Pre-processing of EEG signals comes as the first step when implementing an epilepsy detection system. In the literature, different processing techniques were applied to the EEG. For example, in (San-Segundo et al. 2019), the Fourier, wavelet, and Empirical Mode Decomposition (EMD) transforms were applied to process the EEG signals. In (Hu et al. 2020), the authors processed EEG signals using the local mean decomposition and statistical feature extraction to preserve the non-stationary nature of EEG signals and decrease the computational burden. (Thara D.k. et al. 2019)

normalised the dataset using four distinct feature scaling techniques, each paired with a separate loss function. In (Cao et al. 2020), probability features were learned using stacked Convolutional Neural Networks (CNN). In (Chen et al. 2020), the Epilepsy EEG Graph Convolutional Network was proposed. This network used correlations between channels to mine data information deeply. In (Mathur and Chakka 2020), epileptic seizures were detected using the graph discrete Fourier transform-based features. (Wang et al. 2020) segmented the intracranial EEG signals and used the directed transfer function algorithm to calculate their information flow features. Then, reconstructed these features as channel-frequency maps. (Sood et al. 2020) preprocessed the EEG data by discrete cosine transform and filter conversion, and extracted the features from them. In (de Vargas et al. 2023), EEG signals were classified using multispectral analysis feature selection and a classifier-building approach. This methodology reduced the number of features for building a classifier. In (Prabin Jose et al. 2021), noise was removed from the EEG using a bandpass filter. Then, features were acquired from each channel of the EEG. The probabilistic Principal Component Analysis (PCA) was used to reduce the dimensionality of features. In (Guerrero et al. 2021), features were extracted from the EEG using Fourier analysis through a transformation in the frequency domain. In (Ding and Zhao 2024), features were extracted using the techniques in the frequency domain, time-frequency domain, and nonlinear dynamics. (Ahmad, Liu, et al. 2024) extracted time and frequency domain features from EEG data using a one-dimensional CNN. In the study by (Ahmad, Zhu, et al. 2024), EEG signals were decomposed using the Dual-Tree Complex Wavelet Transform to extract the real and imaginary eigenvalue features. In (Faiq Ahmad Khan et al. 2024), EEG signals were segmented into small pieces of 5 sec. duration and

features were extracted from each channel. In (Bouazizi and Ltifi 2024), the authors identified the relevant features for decision-making from the Power Spectral Density (PSD). In the study by (Ijaz Ahmad et al. 2024), EEG signals were decomposed using the Discrete Wavelet Transform (DWT), followed by the extraction of various eigenvalue features and selection of the most optimal ones. (Khan et al. 2024) segmented EEG signals into epochs of 5 sec. duration and statistical features were extracted from them. In (Gowtham Reddy N et al. 2023), ordered weighted aggregation-based statistical features were used as a feature set to identify the epilepsy disorder. In the study by (Peker et al. 2016), EEG data were processed using the dual-tree complex wavelet transform, and features were extracted using feature vectors. In (Mathur et al. 2021), the single-channel EEG signal was divided into multiple EEG blocks. These blocks were projected onto a particular Ramanujan Periodic Subspace (RPS). Using the energy of projection as a feature, the block was labelled as either epileptic or non-epileptic. In (Sharmila and Geethanjali 2016), statistical features were derived from EEG using DWT. In (Wang et al. 2017), DWT was used to derive the features from EEG. In (Abdelhameed and Bayoumi 2018), the two-dimensional Deep Convolutional Autoencoder (DCAE) was used to identify the spatial features. In (Bekbalanova et al. 2019), two methods, DWT and EMD, were used for frequency space transformation.

In all the methods mentioned above, preprocessing techniques were applied directly to EEG signals. In the literature, some methods have been reported to have converted EEG signals into images to achieve better performance. For example, in (Wang et al. 2019), a multichannel EEG's subband Mean Amplitude Spectrum (MAS) image was used to represent the EEG. In (Hu et al. 2021), the MAS, mean PSD, and

wavelet packet features were derived from EEG signals and fused into an image feature. In (Raghu et al. 2020), the EEG data were converted into spectrogram stack images and classified into non-seizure and seven types of seizure signals. EEG signals were processed by (Gao et al. 2020) to obtain Power Spectrum Density Energy Diagrams (PSDED), which were then used for classification. In the study by (Qu and Yuan 2019), time-frequency images were obtained from EEG signals using a Continuous Wavelet Transform (CWT) to classify them as focal or non-focal. In (Shankar et al. 2022), EEG signals were first decomposed into subcomponents using Hilbert vibration decomposition. The first three subcomponents were used to generate 2D images using CWT, which were then combined into a single image. In (Rashed-Al-Mahfuz et al. 2021), EEG signals were converted into images using CWT and Short-time Fourier Transform (STFT) for the purpose of seizure identification. In (Liang et al. 2020), EEG signals were converted into images for automatic epileptogenic zone recognition and localisation. In (Choi et al. 2019), STFT was used to extract spectral features from EEG signals, which were converted into 2D images for seizure classification. In the literature, before analysing and processing medical images, their quality was improved by applying image enhancement methods (Salem et al. 2019) (Islam and Mondal 2019).

Converting EEG signals into images offers several benefits. CNNs trained on 2D images of EEG, like spectrograms, can capture spatial patterns and frequency correlations very effectively. Spectrograms show how the frequency content changes over time. Moreover, images make it easier to combine multiple EEG channels into a single visual map. Pre-trained CNNs, which are trained on the ImageNet dataset, can be fine-tuned on EEG image datasets, saving time and improving accuracy even with small medical datasets.

2.1.2 Survey of literature on ML-based approaches for epilepsy detection

In the literature, many ML-based models were used for epilepsy detection. For example, (Thara D.k. et al. 2019) employed ML models like Support Vector Machines (SVM) and K-Nearest Neighbours (KNN) for epileptic seizure detection. In (Cao et al. 2020), the kernel Extreme Learning Machine (ELM) algorithm was used for epileptic signal classification. In (Mathur and Chakka 2020), a crisp rule-based classification method was used to detect epileptic seizures. In (Guerrero et al. 2021), the classification was performed using Logistic Regression (LR), Artificial Neural Networks (ANN), SVM, and CNN. In (Qu and Yuan 2019), EEG signals were classified as focal and non-focal using SVM. In (Wong and Kuhlmann 2020), an extremely randomised tree classification-based seizure prediction algorithm was proposed. Designed for resource-constrained environments, the algorithm targets implementation on small-scale implantable or wearable EEG devices. In (Ahmad, Liu, et al. 2024), a multi-view forest classifier was used to classify nonseizure, transition and seizure classes from EEG data. In (Ahmad, Zhu, et al. 2024), Stacking Ensemble Classifiers were used to classify EEG signals. In (Bouazizi and Ltifi 2024), predictions of multiple echo state networks were combined to predict the stroke. In (Abhishek S. et al. 2024), EEG signals from Bern-Barcelona, Khas and Bonn datasets were used and classified as preictal, ictal, interictal, focal and non-focal using different ML algorithms like Random Forest (RF), Ada Boost, Gradient Boosting (GB), Extra Tree, SVM, and Naive Bayes (NB) classifier. In (Xiaofei Ye et al. 2024), the early seizure predictor for children with idiopathic focal epilepsy was developed using different ML algorithms. The ML algorithms used were SVM, Decision Tree (DT) Classifier, RF Classifier, Xgboost Classifier, CATboost Classifier, and ADAboost Classifier. In (Ijaz Ahmad et al. 2024), epileptic seizures were

identified using a Bagged Tree-based classifier. (Khan et al. 2024) utilised SVM, KNN, and RF classification techniques to categorise EEG signals as seizure or non-seizure. In (Chapatwala et al. 2022), epilepsy was diagnosed using different ML models like SVM, RF and XGBoost. In (Gowtham Reddy N et al. 2023), epileptic seizures were detected using different ML models like LR, KNN, SVM, and RF. In (Mridha et al. 2023), classifiers like LR, KNN, SVM, NB Classifier, RF Classifier, and GB were used to classify epileptic seizures. In (Peker et al. 2016), Complex-valued Neural Networks (CVANN) were used for epilepsy detection. In the study by (Mathur et al. 2021), epileptic signals were classified using an SVM classifier. To detect epilepsy, EEG signals were classified using NB and KNN algorithms in the approach proposed in (Sharmila and Geethanjali 2016). (Wang et al. 2017) employed multi-domain feature extraction techniques along with nonlinear analysis to classify EEG signals. In (Bekbalanova et al. 2019), EEG signals were classified using non-DL methods such as SVM, KNN, and a DT classifier. In (Torse et al. 2019), ANN, Probabilistic Neural Network (PNN) and SVM were used for seizure detection.

2.1.3 Survey of literature on DL-based approaches for epilepsy detection

DL-based methods achieved better performance in many areas like healthcare, finance, retail and e-commerce, transportation and autonomous vehicles, the manufacturing industry, agriculture, education, entertainment and media, NLP applications, cybersecurity, environment and sustainability, and scientific research. DL methods were also used for epilepsy detection and reported to achieve promising results. For example, in (San-Segundo et al. 2019), a DNN was used for classifying focal and non-focal EEG signals and achieved a classification accuracy of 98.9%. For

the epileptic seizure recognition dataset, DNN achieved non-seizure vs. seizure classification accuracy of 99.5%, healthy, non-focal and seizure classification accuracy of 96.5%, and seizure, healthy and focal classification accuracy of 95.7%. Seizure detection was performed with the help of a deep Bi-LSTM network using the Children's Hospital Boston - Massachusetts Institute of Technology (CHB-MIT) scalp EEG database in the study by (Hu et al. 2020). In (Thara D.k. et al. 2019), seizure signals were detected using a DNN. In (Sood et al. 2020), EEG signal classification was performed using a Bayesian optimisation-tuned Bi-LSTM model. In (Prabin Jose et al. 2021), epileptic seizure detection was performed using the deep-stacked autoencoder, which was trained using an Adaptive rag-Rider optimisation algorithm. A CNN combined with a Long Short-Term Memory (LSTM) model was employed in (Shankar et al. 2022) to extract features and classify EEG signals into five seizure and non-seizure categories. In (Liu and Woodson 2019), seizure EEG signals were classified using a CNN. In (Shanmugam and Dharmar 2023), the EEG signals were classified as normal, ictal and interictal signals using a one-dimensional CNN-LSTM model. In (Jemal et al. 2022), a DL model was used for epileptic seizure prediction. In (Ding and Zhao 2024), an LSTM was used for seizure detection. In (Faiq Ahmad Khan et al. 2024), epileptic seizure was detected using a fuzzy DL architecture. The classification of EEG signals as seizure and non-seizure was performed using the hybrid model, which was the integration of CNNs and LSTM networks in the study by (Amrani et al. 2024). In (Rashed-Al-Mahfuz et al. 2021), CNN, VGG16 and ResNet50 were used to perform the classification of EEG signals in three different methods. The CNN model was trained from scratch in the first method. In the second method, only the top layers of the models were trained. In the third method, the networks were fine-tuned. In

(Daoud and Bayoumi 2019), four DL models were used to extract features and perform classification. Multi-Layer Perceptron (MLP), Deep Convolutional Neural Network (DCNN) combined with MLP, DCNN combined with Bi-LSTM, and DCAE combined with Bi-LSTM. In (Abdelhameed and Bayoumi 2018), a Bi-LSTM network was used for classification. The CNN model was used for seizure prediction from EEG in (Liang et al. 2016). In (Xu et al. 2021), the CNN model was used for EEG signal classification to identify epileptic seizures. (Liang et al. 2020) applied an 18-layer Long-Term Recurrent Convolutional Network to automatically identify and localise epileptogenic zones. Epileptic EEG signals were classified using a Recurrent Neural Network (RNN) in the study by (Aliyu et al. 2019). A neural memory network enhanced with neural plasticity was utilised by (Ahmedt-Aristizabal et al. 2020), to distinguish between various seizure types. In (Truong and Kavehei 2019), the CNN was used to evaluate the impact of bit reduction in an analogue-to-digital converter used in an EEG recorder for seizure onset detection. In (Wang et al. 2018), the multivariate brain connectivity detection problem was solved using an RNN-based Granger causality estimator. It was useful in diagnosing epileptic foci. In (Antoniades et al. 2017), a CNN was used to detect interictal epileptiform discharge. In (Olokodana et al. 2020), the distributed Kriging-Bootstrapped DNN model was used to detect seizures. (Taqi et al. 2017) applied three DNNs, LeNet, AlexNet, and GoogLeNet, to distinguish between focal and non-focal EEG signal patterns. A 3D-CNN combined with a bidirectional Gated Recurrent Unit was utilised by (Choi et al. 2019) to extract features from EEG-based 2D images for seizure detection.

DL-based models offer significant advantages in automated epilepsy detection by eliminating the need for manual feature engineering, effectively modelling nonlinear

temporal-spatial EEG patterns, and achieving superior classification accuracy. They support end-to-end learning, enable transfer learning for small datasets, and integrate seamlessly with explainable AI frameworks to provide interpretability. In contrast, non-DL models rely on handcrafted features, struggle with complex high-dimensional EEG data, and often show limited scalability and adaptability.

2.1.4 Survey of literature on transfer learning-based approaches for epilepsy detection

Some epilepsy detection systems used a transfer learning approach and reported achieving improved results. In (Wang et al. 2019), classification was performed using three CNNs with deep transfer learning. Experiments were performed on the CHI-MIT database for the classification of seizure, interictal, and three preictal states. In (Hu et al. 2021), features were extracted from a fused image feature using five DNNs with a transfer learning approach. The classification was performed using a 7-layer fully connected Hierarchical Neural Network (HNN). In (Raghu et al. 2020), ten pre-trained networks were used with a transfer learning approach to classify EEG signals into seven seizure types and a non-seizure class. (Gao et al. 2020) adopted a DCNN-based model with transfer learning to process and classify EEG recordings from the CHB-MIT database. A transfer learning-based feature extraction technique using the AlexNet architecture was implemented by (Qu and Yuan 2019) on EEG image data. In (Daoud and Bayoumi 2019), a transfer learning based semi-supervised approach with DCAE was used for patient-specific epileptic seizure prediction. In (Abdelhameed and Bayoumi 2018), epileptic seizures were predicted, in which the weights for the patient-specific networks were initialised using the transfer learning technique. In (Liang et al.

2016), the seizure signals present in the EEG were predicted using the transfer learning approach.

Transfer learning significantly enhances automated epilepsy detection systems by enabling high accuracy with limited EEG data, reducing computational demands, and improving model generalisation. By leveraging pretrained models from large visual datasets, transfer learning captures robust feature hierarchies transferable to EEG-derived images such as spectrograms. It also supports cross-dataset adaptability and integration with XAI frameworks, making it a powerful and practical approach for developing clinically reliable epilepsy detection systems.

2.1.5 Survey of literature on XAI techniques for explaining epilepsy prediction

With the improvement in the architecture of deep networks, they became deeper and more complicated and turned into black boxes. To understand the predictions of these black boxes, developers used various XAI methods in different domains, as per the situations and requirements (Yang 2022) (Linardatos et al. 2020) (Das and Rad 2020) (Pocevičiūtė et al. 2020) (Samek et al. 2021). XAI methods make a substantial contribution while processing medical images using DL-based methods (van der Velden et al. 2022) (Singh et al. 2020) (Prajakta Rathod and Shefali Naik 2022). Some of the commonly used XAI methods are LRP (Montavon et al. 2019), Deep Learning Important Features (DeepLIFT) (Shrikumar et al. 2019), LIME (Ribeiro et al. 2016), Guided backpropagation (Springenberg et al. 2015) and SHAP (Lundberg and Lee 2017).

Some epilepsy detection systems have become transparent with the use of different XAI methods. In the study conducted by (Jemal et al. 2022), LRP was used to

explain the prediction results of a DNN applied to epileptic seizure detection on the CHB-MIT dataset. Moreover, XAI methods were also used to provide explanations for the EEG signal classification results while diagnosing other disorders. For example, in (Banluesombatkul et al. 2021), EEG-based sleep stage classification results were explained using the LRP method. In (Sidulova et al. 2021), the identification of Alzheimer's disease and cognitive impairment was explained with the help of the XAI method LIME, in which EEG signals were classified using ANN and CNN. In (Li and Adeel Khan 2019), one task was presented in front of subjects on two different devices: a head-mounted display and a personal computer. The authors monitored and differentiated the attention level on the two devices. The feature extraction and classification were performed by the compact CNN EEGNet. The extracted features were visualised by the XAI method DeepLIFT. Thus, the application of XAI techniques - LIME, LRP, and DeepLIFT - enabled interpretability of EEG classification outputs and delivered promising interpretive results.

The unification of LIME, LRP and DeepLIFT is an XAI method, SHAP (Lundberg and Lee 2017) (Holzinger et al. 2022). Therefore, the SHAP method is used in the proposed study to provide an explanation for the EEG signal classification results in diagnosing epilepsy. Moreover, the SHAP method was widely used in the literature and provided outstanding results in explaining epileptic signal identification. For example, in (Ding and Zhao 2024), the contribution of every channel in seizure detection was determined with the help of the Shapley value obtained using the SHAP method. This facilitated the selection of relevant channels and decreased the total number required for effective seizure detection. In (Ahmad, Liu, et al. 2024), epileptic seizures were detected using an advanced multi-view DL framework, while SHAP

provided insights into the model's decision-making process. (Ahmad, Zhu, et al. 2024) applied the SHAP method to provide interpretability for the predictions of Stacking Ensemble Classifiers. In (Faiq Ahmad Khan et al. 2024), SHAP and LIME methods were used to provide interpretability for the predictions of the fuzzy DL model. (Bouazizi and Ltifi 2024) understand the influential features in the model prediction with the help of the SHAP and LIME methods. (Amrani et al. 2024) applied the SHAP method to provide insights into the decision-making of the hybrid model. (Abhishek S. et al. 2024) explained the predictions of the model with the help of the SHAP and LIME methods. In (Xiaofei Ye et al. 2024), important features in seizure detection were analysed using the SHAP method. In the study by (Ijaz Ahmad et al. 2024), the SHAP method was employed to interpret how the Bagged Tree-based classifier made its decisions. In (Khan et al. 2024), the transparency of the system was enhanced with the help of the SHAP and LIME methods. In (Chapatwala et al. 2022), the classification results of RF, SVM and XGBoost classifiers were explained with the help of the SHAP method while diagnosing epilepsy. In (Gowtham Reddy N et al. 2023), the SHAP method was used to show the impact of features in epileptic seizure identification using an RF classifier. In (Mridha et al. 2023), the decision-making process of the non-DL models for epilepsy detection was explained with the help of the SHAP and LIME methods. In the study by (Rashed-Al-Mahfuz et al. 2021), the SHAP was used to detect frequency features unique to seizure-related EEG signals compared to normal EEG.

The integration of XAI in automated epilepsy detection systems transforms DL models from black boxes into clinically interpretable tools. XAI enhances trust, transparency, and reliability by identifying key EEG features that drive model predictions and validating biological relevance. It also facilitates regulatory compliance

and clinical decision-making, ensuring that AI-driven epilepsy detection is both accurate and clinically acceptable.

2.1.6 Survey of literature on single and multi-channel datasets used for epilepsy detection studies

The publicly available dataset from Bonn University (Andrzejak et al. 2001) was used by many researchers for EEG signal classification and seizure identification. They applied different preprocessing and classification techniques to EEG data. For example, in (Thara D.k. et al. 2019), epilepsy detection was performed on the Bonn University dataset using ML models and DNN. SVM classified the EEG signals with 94% accuracy, while KNN achieved an accuracy of 79.7%. The DNN achieved a classification accuracy of 97.21%. Using the Bonn University EEG dataset, (Qu and Yuan 2019) achieved 95.00% accuracy in classifying signals as either focal or non-focal. In (Liu and Woodson 2019), EEG signals were obtained from the Bonn University dataset and classified with more than 90% accuracy on average. (Ahmad, Liu, et al. 2024) perform experiments on the Bonn University dataset and classify EEG signals into non-seizure, transition and seizure classes with an accuracy of 99.40%. In (Ahmad, Zhu, et al. 2024), the Bonn University dataset was used for epileptic signal identification and achieved an accuracy of 99.20%. In (Peker et al. 2016), the Bonn dataset was used for epilepsy detection and achieved a classification accuracy of 99.30%. In (Mathur et al. 2021), an accuracy of 98.6% was achieved in detecting epilepsy from EEG signals in the Bonn dataset. In (Sharmila and Geethanjali 2016), EEG signals from the Bonn University dataset were used for epilepsy detection by non-DL models and achieved an accuracy of 96.5%. (Wang et al. 2017) performed

experiments on the Bonn dataset, classifying EEG signals as seizure and non-seizure with an accuracy of 99.25%. Epilepsy detection based on the Bonn University dataset was conducted by (Xu et al. 2021), with a reported classification accuracy of 95.8%. Epileptic and non-epileptic EEG signals were distinguished with 94% accuracy in the work of (Sudhamayee et al. 2023), using the Bonn University dataset. They used a simplicial approach with an SVM classifier. (Gowtham Reddy N. et al. 2024) used the Bonn University dataset and classified EEG signals as healthy, seizure-free, and seizure using the fuzzy pattern tree classifier. They achieved an accuracy of 97.8%. Epilepsy detection using EEG signals from the Bonn University dataset was carried out by (Abedin et al. 2019), achieving 97.33% accuracy with an ANN model. Using the Bonn University EEG dataset (Bekbalanova et al. 2019) applied SVM, KNN, and DT classifiers to distinguish between normal, pre-seizure, and seizure conditions. In (Jose et al. 2020), EEG signals from the Bonn University dataset were classified as normal and seizure signals using the ELM classifier with an accuracy of 98.5%. (Aliyu et al. 2019) used the Bonn University EEG dataset to recognise epileptic patterns. In the study by (Xie et al. 2019), epileptic EEG recognition was performed using the Bonn University dataset. In (Torse et al. 2019), the Bonn University dataset was used for seizure detection employing ANN, PNN, and SVM techniques. (Olokodana et al. 2020) performed the seizure detection using the Bonn University dataset with the help of a distributed Kriging-Bootstrapped DNN and achieved an accuracy of 98.53%.

Many researchers have used the multichannel EEG dataset from the TUH (Veloso et al. 2017). This dataset has many subsets, which were used to solve different research problems. For example, in (Prabin Jose et al. 2021), epileptic seizure detection was performed on the TUEP EEG dataset and achieved a classification accuracy of

91.5%. In (Guerrero et al. 2021), the TUEP EEG dataset was used to carry out experiments and achieved an epileptic vs non-epileptic classification accuracy of 86.00%. (Raghu et al. 2020) performed experiments on the EEG signals from the TUH dataset and classified non-seizure and seven types of seizure signals, and achieved an accuracy of 77.66%. In (Sudhamayee et al. 2023), epileptic and non-epileptic EEG signals from the TUEP dataset were classified with a classification accuracy of 92.86%. They used a simplicial approach with an SVM classifier. In (Huang and Duan 2023), the TUEP dataset was used to detect epileptic discharges. The features from the EEG data were extracted using the graph CNNs. The system could achieve an accuracy of 92.20%. Using the TUEP dataset, (Gowtham Reddy N. et al. 2024) achieved 86% accuracy in classifying EEG signals into epileptic and non-epileptic classes. In (Jaffino et al. 2022), EEG signals from the TUEP dataset were used for epilepsy detection and achieved a classification accuracy of 93.6%. In the study by (Tawhid et al. 2022), EEG signals from the TUEP dataset were segmented into time frames of 3-sec. duration and classified using a convolutional LSTM network, reaching an accuracy of 92.17%. EEG signals from the TUEP dataset were classified as epileptic and healthy with an accuracy of 97% in the work done by (Omari et al. 2023). LConvNet combined CNNs for extracting spatial features and an LSTM for temporal dependencies identification. (Tang et al. 2024) performed epilepsy detection using the TUEP dataset, where features were extracted using the path signature algorithm and classified using a Bi-LSTM network with an attention mechanism, achieving an accuracy of 95.60%. Using the TUH EEG seizure corpus (Ahmedt-Aristizabal et al. 2020) performed classification of multiple seizure types.

The Bonn EEG dataset provides clean, noise-free, and well-structured data ideal for algorithm development, benchmarking, and proof-of-concept studies. In contrast, the TUH EEG corpus offers large-scale, multi-channel, clinically annotated EEG recordings reflecting real-world variability, making it highly suitable for developing and validating robust DL and XAI models. Together, these datasets offer a balanced foundation - Bonn for model training and methodological exploration, and TUH for testing scalability, generalisation, and clinical applicability in automated epilepsy detection systems.

This literature review provided insight into the current progress in automated epilepsy detection. At the same time, it helped to identify the primary and secondary research problems, which are addressed through specific research objectives. These research problems and corresponding research objectives are elaborated in Section 2.2.

2.2 Research problem

The literature review highlighted a broad spectrum of methodologies applied in the detection of epileptic seizures from EEG signals, ranging from traditional pre-processing techniques and non-DL approaches to the more advanced DL and transfer learning frameworks. Pre-processing methods, including filtering, artefact removal, and feature extraction, have been crucial for improving signal quality and reducing noise, thereby enabling more reliable downstream analysis. Classical ML methods like SVM, KNN, RF and DT demonstrated moderate classification success but were often limited in handling the non-linear and high-dimensional nature of EEG signals. These models often struggled to capture the temporal and spatial complexities in EEG data. By contrast, DL approaches, especially CNNs and RNNs, have proven highly effective

at automatically extracting hierarchical and discriminative features, achieving improved seizure detection accuracy.

Additionally, transfer learning approaches have proven effective in addressing the challenges posed by limited annotated EEG datasets, enabling models trained on larger datasets to be adapted for more specific seizure detection tasks. However, despite the improved performance of DL models, a major limitation identified in the literature is the lack of interpretability, which hampers their clinical adoption and trustworthiness. To address this, XAI techniques have been explored, offering a way to uncover the decision-making processes of complex models and providing clinicians with meaningful insights into seizure-related EEG patterns. Moreover, the Bonn and TUH datasets, widely used benchmarks in epilepsy research, have played a pivotal role in evaluating and comparing different methodologies, yet gaps remain in developing approaches that are interpretable and hence clinically acceptable.

Despite significant progress in automated EEG-based seizure detection, a critical challenge remains in identifying which EEG channels are most significant for distinguishing epileptic from non-epileptic patients and in linking these channels to the underlying epileptogenic brain regions. Existing DL models have achieved high accuracy in seizure classification, but they typically treat all channels equally or focus on global feature extraction without explicitly highlighting channel-level contributions. Consequently, there is a pressing need for research that integrates robust DL models with XAI methods to determine significant EEG channels, improve classification performance, and provide clinically interpretable insights into epileptogenic regions. Addressing this gap will directly contribute to both computational neuroscience and

clinical decision support in epilepsy management. Building upon these insights, the primary and secondary research questions and research objectives are as follows:

2.2.1 Primary research question

Research Question: How can a clinically acceptable automated system be developed to distinguish between epileptic and non-epileptic patients while also identifying the most influential channels contributing to the detection of epilepsy?

Research Objective 1: To develop a DL-based, interpretable, automated epileptic seizure detection system.

Research Objective 2: To evaluate the system performance across various datasets.

Research Objective 3: To identify the top contributing EEG channels in determining epileptic and non-epileptic patients.

2.2.2 Secondary research question

1. Research Question 1: How to develop an automated epilepsy detection system with the help of a DNN-based EEG signal classifier, which is reliable and acceptable in the medical domain?

   Research Objective 1: To select a DNN suitable for EEG data to classify epileptic and non-epileptic EEG signals.

   Research Objective 2: To make the developed system interpretable with the help of the XAI methods, which will turn it into a more reliable and acceptable system in the medical domain.

2. Research Question 2: How to enhance the performance of the epileptic signal classifier with the help of technical improvements in the preprocessing method?

   Research Objective 1: To convert EEG signals into images to achieve higher classification accuracy of epileptic and non-epileptic EEG signals.

   Research Objective 2: To improve the visual quality of EEG signal images by applying filters.

3. Research Question 3: How to identify the most suitable among the widely used XAI methods and DNNs to select the most appropriate XAI and DNN combination for optimal performance with various datasets? How to make it more useful for neurologists in identifying the epileptogenic brain regions of epilepsy patients?

   Research Objective 1: To evaluate the performance of widely used XAI methods for explaining the classification of EEG signal images and select the best-performing XAI method.

   Research Objective 2: To evaluate the performance of widely used DNNs in classifying EEG signal images and producing explanations with the best-performing XAI method, and select the best-performing DNN.

   Research Objective 3: Select the best-performing DNN and XAI method to identify the top contributing channels in classifying epileptic and non-epileptic patients, which can be used to locate epileptogenic brain regions.

Summary:

This chapter presents a comprehensive review of the existing research on epileptic seizure detection. It begins by exploring various preprocessing techniques that have been applied to EEG data to enhance the quality and interpretability of signals. Additionally, studies that convert EEG signals into image representations during preprocessing are examined to understand their impact on classification performance. Both non-DL and DL-based EEG signal classification methods from the literature are reviewed to highlight the evolution of techniques used in this domain. Furthermore, different approaches employing transfer learning for seizure detection are discussed, emphasising their potential to perform well with a small dataset and achieve a good model accuracy. The chapter also investigates the application of various XAI methods - such as Integrated Gradients, LRP, DeepLIFT, LIME, Guided Backpropagation, and SHAP - in the context of epilepsy detection. Among these, SHAP has been found to be widely used and particularly effective in providing meaningful explanations for epileptic seizure identification. Lastly, research studies utilising the Bonn University and TUH EEG datasets are reviewed, as these datasets are among the most frequently employed benchmarks in seizure detection research.

3. Methodological Framework for Epilepsy Detection and Explainability Strategy

This chapter describes the dataset used to conduct the research. Moreover, this chapter presents the new methods developed to meet multiple research objectives outlined in Section 2.2. These methods are: (i) implementation of LRP in Bi-LSTM network, (ii) implementation of LRP in VGG16 network, (iii) comparison of the XAI methods SHAP, Guided Backpropagation and LIME, (iv) SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0 and (v) SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channels. The first method works directly with EEG signals. The other four methods convert EEG signals to spectrogram images before using them for model training. The first four methods work with a single-channel EEG dataset, and the fifth method works with a multi-channel EEG dataset.

3.1 Dataset

For this study, we made use of two datasets: (1) the Bonn University dataset, and (2) the Temple University Hospital (TUH) dataset. Further information regarding these datasets can be found in Sections 3.1.1 and 3.1.2.

3.1.1 Bonn University dataset

This study utilises the open-source Bonn University Dataset (Andrzejak et al. 2001), which consists of single-channel EEG recordings, as outlined in Table 3.1. The dataset is divided into five subsets: A, B, C, D, and E, also referred to as Z, O, N, F, and S, respectively. Sets A and B contain EEG data recorded from healthy individuals with eyes open and eyes closed, respectively. Sets C, D, and E include data from individuals diagnosed with epilepsy. Among these, Sets C and D represent interictal EEG activity, while Set E comprises seizure data. The recordings in Set C were obtained from the hemisphere opposite to the epileptogenic focus, whereas those in Set D were recorded from within the epileptogenic zone. Each subset contains 100 trial segments, with each segment comprising 4097 samples and lasting 23.6 seconds. The EEG signals were recorded using a single channel at a sampling frequency of 173.6 Hz and subsequently filtered using a bandpass filter with a frequency range of 0.53–40 Hz.

Table 3.1 Bonn University Dataset Description

Set	File Name	Patient	Setup	Data Points	Time Duration
A	Z	Healthy	Surface EEG	4097	39.33 min.
B	O	Healthy	Surface EEG	4097	39.33 min.
C	N	Epilepsy	Intracranial EEG	4097	39.33 min.
D	F	Epilepsy	Intracranial EEG	4097	39.33 min.
E	S	Epilepsy	Intracranial EEG	4097	39.33 min.

3.1.2 Temple University Hospital (TUH) dataset

In this research, the TUH (Veloso et al. 2017) dataset, which is an open-access multi-channel EEG dataset, is used. It consists of various subsets that can be used to solve various research problems related to epileptic seizures. Our research objective is based on epileptic and non-epileptic EEG signal classification and seizure

identification. Subset TUEP from the TUH dataset was suitable for our research objective. Therefore, the TUEP subset is selected to perform the research study. This subset has 200 subjects' data files in European Data Format (EDF) format. Among these 200 subjects, 100 have epilepsy, and 100 do not have epilepsy. Various factors are considered to identify epileptic and non-epileptic patients. These factors are patients' clinical history, multiple features that can be used to identify epileptic signals, and medicines given to the patients at the time of recording the signals. To maintain uniformity, EEG signal recording from the common 21 channels, which are present in all EDF files, is considered for this research. Accordingly, data from 100 patients are selected, 50 epileptic and 50 non-epileptic, having common EEG channels. Table 3.2 describes the data. It consists of 47 male and 53 female subjects, and their age range is 43 to 79. The International 10-20 system is used to record the signals. Two different montages are used to record the data: average reference and linked ear reference. Recording from the average reference montage is used to maintain uniformity. The signals are sampled at the rate of 250 Hz. The Biomedical Research Ethics Committee at Ahmedabad University granted approval to conduct this research using publicly available and anonymized datasets for experimentation.

Table 3.2 Data from the TUH Dataset Used in the Research Study (Prajakta Rathod et al. 2025)

	Epileptic	**Non-epileptic**
No. of Patients	50	50
No. of Files	50	50
Male/Female Ratio	21:29	26:24
Age Group	49-77	43-79
EEG Channels	21	21
Sampling Rate	250 Hz	250 Hz
Montage	Average Reference	Average Reference

3.2 Methods

3.2.1 Implementation of LRP in Bi-LSTM network

3.2.1.1 Data preprocessing

The primary objective of this research is to identify epileptic and non-epileptic patients. Therefore, data from sets B and E are used in this research study. The duration of recording in each set is 39.33 min. Table 3.3 presents comprehensive information about the dataset employed in this research study.

Table 3.3 Dataset Description (Prajakta Rathod et al. 2022)

Signals	**Set**	**Data Points**	**Time Duration**
Normal EEG signals	B	4097	39.33 min.
Epileptic EEG signal	E	4097	39.33 min.

3.2.1.2 Proposed method

The method has three steps: (i) Preprocessing the EEG data, (ii) Training and testing the Bi-LSTM network on EEG data, and (iii) implementing the LRP method in the Bi-LSTM network to explain the classification results. An overview of the system architecture is presented in Fig. 3.1. EEG data is pre-processed and normalised before being fed to the network for classification. EEG signal data is in the form of numerical values. EEG signals are classified as epileptic or non-epileptic using a Bi-LSTM network consisting of one Bi-LSTM layer and one dense layer. The dense layer comprises two neurons, each corresponding to the epileptic and non-epileptic classes, respectively. Training is conducted over 100 epochs using the Adaptive Moment Estimation (Adam) optimiser, with categorical cross-entropy as the loss function and sigmoid activation function. The random_state value is set to 4, batch size is 32 and

learning rate is 0.01. The dataset is randomly divided into 80% training and 20% testing sets. After classification of EEG signals into epileptic and non-epileptic classes, the XAI method LRP is used to explain the network's decisions.

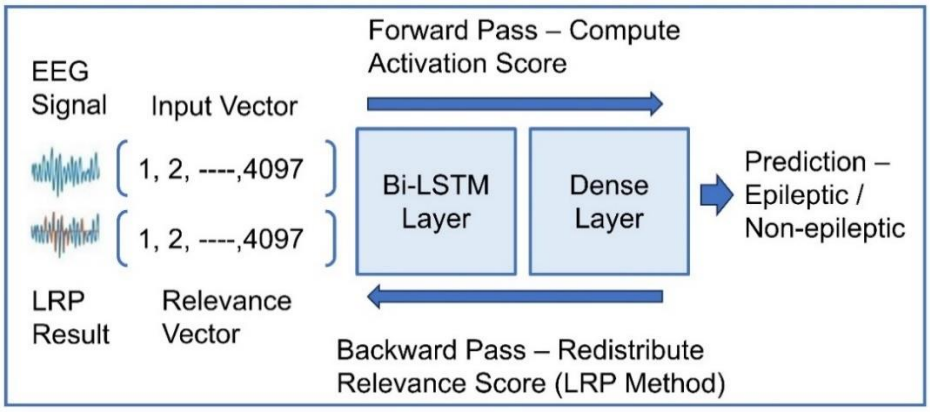


Fig. 3.1 System architecture (Prajakta Rathod et al. 2022)

3.2.1.2.1 Implementation of LRP method in Bi-LSTM network

When EEG signal data is input to the Bi-LSTM network for classification, it produces an activation score at the output layer. This activation score gives a prediction about the output class. The output layer has two neurons for epileptic and non-epileptic output classes. Relevance values are set for these two neurons. The target class neurons' activation score obtained at the output layer is set as its relevance value. The neuron of another class in the output layer is assigned a relevance value of zero. These relevance scores are propagated in the reverse direction, layer by layer. Finally, these relevance scores reach the lower-layer neuron (Wu et al. 2022) (Montavon et al. 2019). In a Bi-LSTM network, connections are formed in two ways: many-to-one weighted linear connections and two-to-one multiplicative interactions. These two connections have different formulas to compute relevance values propagated to the lower layers. When

the relevance values are propagated from the output layer in the reverse direction, their values are computed using formulas which is specific to the connection involved. Relevance propagation through both connections is described below.

1) Many-to-one weighted linear connections: Let $z_i$be lower-layer neurons and z_j be upper-layer neurons. w_{ij} are the connection weights, and b_j are the biases. The relevance score of neurons at lower layer i is R_i and at the upper layer j is R_j. The upper-layer neuron z_j is connected to the 'N' number of lower-layer neurons. The relevance score is redistributed to lower-layer neurons in two steps. Relevance messages $R_{i\leftarrow j}$ from upper layer neurons to lower layer neurons, i.e., from z_j to z_i are computed in the first step. Then, for each lower-layer neuron z_i, all incoming messages $R_{i\leftarrow j}$ are summed up. As mentioned in (Arras et al. 2017), $R_{i\leftarrow j}$ is computed by using equation 3.1

$$R_{i\leftarrow j} = \frac{z_i \cdot w_{ij} + \frac{\epsilon \cdot sign\left(z_j\right) + \delta \cdot b_j}{N}}{z_j + \epsilon . sign(z_j)} \cdot R_j \tag{3.1}$$

In this equation, ϵ serves as a stabilising term with a small positive value of 0.001. δ is a multiplicative factor. Its value can be either 1.0 or 0.0. When the value of δ is set to 0.0, the biases absorb a fraction of the relevance, leading to an approximately conservative propagation rule. When the value of δ is set to 1.0, then for all the neurons that belong to the same layer, the total relevance is conserved. The default value of δ is 0.0 (Arras et al. 2017).

2) Two-to-one multiplicative interactions: Two-to-one multiplicative interactions represent another form of connection, takes place between neurons in the lower layers. Consider the upper-layer neuron to be z_j. The lower-layer neurons are z_g and z_s, i.e., gate and source neurons. In the forward pass, the value of z_j is computed by using the formula $z_j = z_g \cdot z_s$. The output value is generated by the sigmoid activation function. This value lies in the range of 0 to 1, and it is assigned to the gate neuron. Suppose the relevance of the neuron z_j in the upper layer is R_j, then it is redistributed to lower-layer neurons z_g and z_s as $R_g = 0$ and $R_s = R_j$ (Arras et al. 2017).

3.2.1.2.2 Relevance vector generation

To test the working of the LRP method, a test input vector is given to the model as shown in Fig. 3.1. The network classifies the test input correctly in the forward pass. The LRP method works in the backward pass to explain the classification results generated by the network. LRP method backpropagates the relevance score layer by layer in the reverse direction. Finally, this relevance score reaches the input layer. Data in the test input vector is in numerical form. For each value in the test input vector, a relevance value is generated. A relevance vector is obtained for the test input vector from these relevance values. Normalisation is applied to these relevance values, constraining them to the range of 0.0 to 1.0. A relevance value of 0.0 indicates that its corresponding value in the test input vector has the least contribution to the output prediction of the network. On the other hand, relevance value 1.0 indicates that its corresponding value in the test input vector has a higher contribution to the output prediction of the network. From the test input vector, values for which the

corresponding relevance values are greater than 0.5 are selected. These values are marked in red colour by the LRP method as shown in Fig. 4.2 and Fig. 4.4.

3.2.2 Implementation of LRP in VGG16 network

3.2.2.1 Data preprocessing

The objective of this study is to distinguish between seizure and non-seizure EEG signals and to elucidate the reasoning behind its decisions. Therefore, data from two sets, set B, which contains healthy people's EEG data, and set E, which contains epilepsy patients' EEG data during a seizure attack, are considered. One trial file is of 23.6 sec. duration long, and each set has 100 trial files. Therefore, sets B and E have 39.33 min. duration recording, which is used in this research study. Table 3.4 describes the dataset.

Table 3.4 Data Used in Research Study (Prajakta Rathod et al. 2024)

Set	Setup	Signal Type	Duration
B	Surface EEG	Non-seizure (Normal)	39.33 min.
E	Intracranial EEG	Seizure	39.33 min.

EEG signals are divided into small segments of 2 sec. duration. There is an overlap of 1 sec. duration, which helps to preserve information. It also helps to increase the training samples for the network. One trial file has 23.6 sec. duration signal. Therefore, 22 data segments of 2 sec. duration are obtained from one trial file. Two sets, B and E, have 100 trial files each. So, sets B and E gave 2200 signal segments each. A total of 4400 data segments are used to carry out this research study. These small segments are converted into spectrogram images by applying STFT. EEG signal variations in amplitude and frequency over time are illustrated using a spectrogram

image. In this research study, 4400 spectrogram images are used for binary classification.

Fig. 3.2a shows a sample signal segment of 2 sec. duration from set E, i.e., epileptic EEG signal, and Fig. 3.2b shows its spectrogram. Similarly, Fig. 3.3a shows a sample signal segment of 2 sec. duration from set B, i.e., non-seizure signal, and Fig. 3.3b shows its spectrogram. In a spectrogram, the horizontal axis corresponds to time, while the vertical axis corresponds to frequency. The bottom region of the image illustrates the low-frequency components of EEG signals, whereas the top region shows

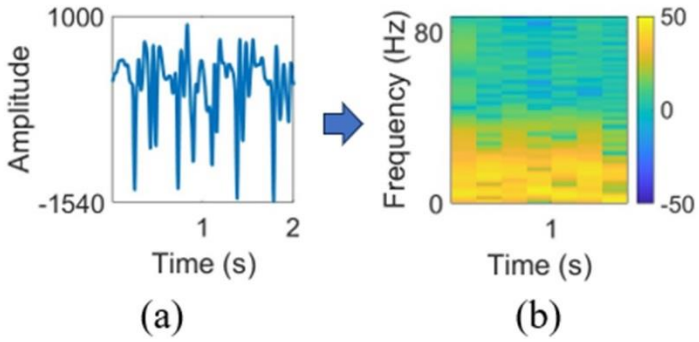


Fig. 3.2 Seizure signal and its spectrogram. Fig. a displays the seizure signal, and Fig. b displays its spectrogram (Prajakta Rathod et al. 2024)

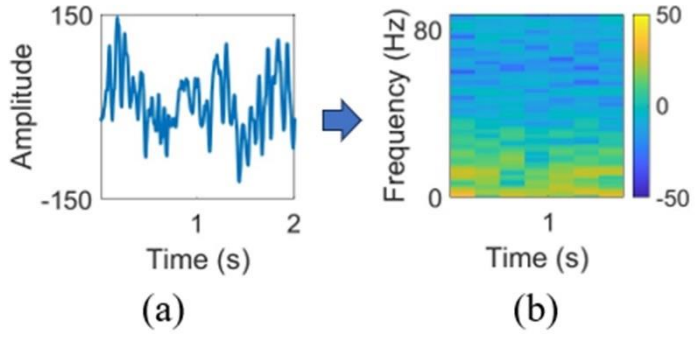


Fig. 3.3 Non-seizure signal and its spectrogram. Fig. a shows a non-seizure signal, and Fig. b shows its spectrogram (Prajakta Rathod et al. 2024)

the high-frequency components. Colour in the spectrogram image represents the amplitude of the EEG signal. The blue colour shows low amplitude, and the yellow colour shows high amplitude. The change in amplitude from low to high changes the colour from blue to yellow in the spectrogram image. The colorbar range for amplitude is -50 to 50. As shown in Fig. 3.2a and Fig. 3.3a, the amplitude is higher in the case of an epileptic seizure signal than a non-seizure signal. Hence, the spectrogram for the epileptic seizure signal in Fig. 3.2b shows more yellow colour in the lower frequency region than the non-seizure signal spectrogram in Fig. 3.3b. The classifier learns this difference in amplitude and performs classification of the EEG as an epileptic seizure signal and a non-seizure signal.

3.2.2.2 Proposed method

The system architecture is illustrated in Fig. 3.4. EEG signals are preprocessed and converted into an image form using STFT. To distinguish between epileptic seizure and non-seizure EEG images, the VGG16 architecture is employed for feature extraction and classification, leveraging a fine-tuned transfer learning approach. The size of spectrogram images is changed to 224 × 224 pixels to give input to the VGG16 network, as this is the required input size for the VGG16 network.

The original VGG16 network is trained on the ImageNet dataset. This dataset has images from 1000 different classes. The VGG16 model is robust in identifying curves, edges and colours in the images. The CNN model is initialised with weights from the VGG16 model. Then, the CNN model is trained on the EEG image dataset from scratch. Previous learning of the VGG16 model is adapted to a new EEG dataset.

The CNN model's output layer is modified to identify images from the two classes: seizure and non-seizure.

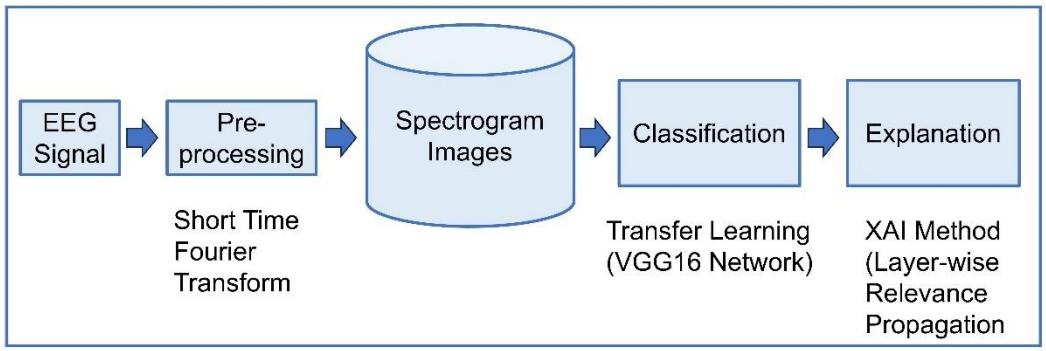


Fig. 3.4 Architecture of the proposed system (Prajakta Rathod et al. 2024)

The dataset is divided into an 80% training set and a 20% testing set using a random selection of samples. During training, the Adam optimiser solver is used. The learning rate is 0.00001, the batch size is 100, and 15 epochs are used for training. The loss function is Cross-Entropy. Since no random seed was specified during model training, the system's entropy source - an operating system component that collects unpredictable environmental data - was used by default to generate random numbers. This network is trained to identify images from seizure and non-seizure classes.

After the network has classified the image into seizure or non-seizure class, the XAI method LRP is used to explain the classification results. The LRP method is implemented in the network, which backpropagates the relevance value and marks the positively contributing pixels in the identification of the image as a seizure or non-seizure class image. Section 3.2.2.3 describes the LRP method in detail.

3.2.2.3 Layer-wise Relevance Propagation (LRP)

The VGG16 network's classification outcomes are explained by implementing the LRP method in the network. Fig. 3.5 shows the working of the LRP method. It has two passes, a forward pass, and a backward pass. The activation score is computed during the forward pass. The output layer's activation score is used to identify the class of the input, and this score is then propagated backward through the network in a layer-wise manner. Pixels that have a positive contribution to image class prediction are identified at the input layer. Each pixel's contribution to the final prediction is represented by its relevance value, which is calculated at every input layer and sorted. These relevance values explain the classification results.

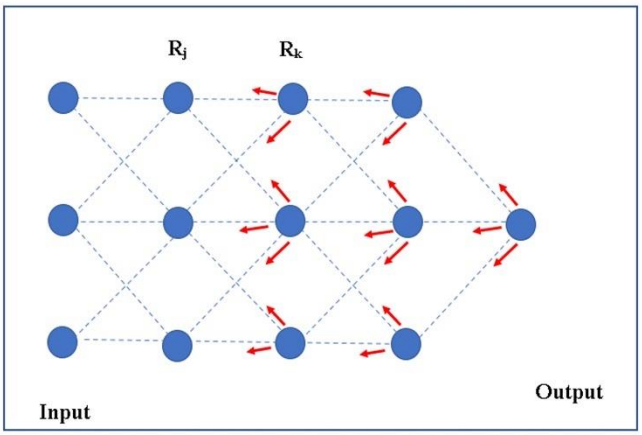


Fig. 3.5 Working of the LRP method implemented in the VGG16 network (Prajakta Rathod et al. 2024)

Let neuron activation be given by a_j and neuron bias by w_{ij} and w_{jk}. Let neuron k be at a higher layer and neuron j be at a lower layer. The relevance score R_j propagated from neuron k to neuron j is given by equation 3.2

$$R_j = \sum_k \frac{z_{jk}}{\sum_j z_{jk}} R_k \tag{3.2}$$

Where,

R_j – Relevance propagated from neuron k to j

k and j – Neurons in the higher and lower layers

Z_{jk} – Contribution of neuron j in making neuron k relevant

R_k – Relevance score at higher layer, i.e., layer of neuron k

Three different rules are used to propagate relevance in the reverse direction during the implementation of this method. These three rules are LRP-0, LRP-ε and LRP-γ which are a special case of the generic rule given by equation 3.3

$$R_j = \sum_k \frac{a_j . \rho(w_{jk})}{\epsilon + \sum_{0,j} a_j . \rho(w_{jk})} R_k \tag{3.3}$$

Where,

ρ – Function used to transform the weights

ε - Small positive value for increment

1) **Basic Rule (LRP-0):** The LRP-0 rule propagates relevance backward through the network. Each input contributes to activating neurons. The relevance score redistributed by the LRP-0 method is proportional to that contribution. LRP-0 rule is given by equation 3.4

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{o,j} a_j w_{jk}} R_k \tag{3.4}$$

When the LRP-0 rule is uniformly used throughout the network, the resulting explanation corresponds to that produced by the Gradient × Input

approach. The gradient of DNN is noisy. Therefore, we need a more robust rule to propagate relevance.

2) **Epsilon Rule (LRP-ε):** The LRP-ε rule modifies the LRP-0 rule by adding a small positive term, ε, to the denominator. The term ε absorbs the relevance when the contribution is negative. The factors that have major contributions to output prediction can only survive when ε becomes larger. It gives sparser and less noisy explanations. The Epsilon rule is given by equation 3.5

$$R_j = \sum_k \frac{a_j w_{jk}}{\epsilon + \sum_{0,j} a_j w_{jk}} R_k \tag{3.5}$$

3) **Gamma Rule (LRP– γ):** In the gamma rule, parameter γ is added. This parameter favors the effect of positive contribution. Also, it helps to remove negative contributions. This rule is given by equation 3.6

$$R_j = \sum_k \frac{a_j \cdot \left(w_{jk} + \gamma w_{jk}^{+}\right)}{\sum_{0,j} a_j \cdot \left(w_{jk} + \gamma w_{jk}^{+}\right)} R_k \tag{3.6}$$

4) **Z^B Rule:** At the input layer, pixel values need to be handled. The Z^B rule properly handles the pixel values. Therefore, the Z^B rule is applied at this layer, which is given by equation 3.7

$$R_i = \sum_j \frac{x_i w_{ij} - l_i w_{ij}^{+} - h_i w_{ij}^{-}}{\sum_i x_i w_{ij} - l_i w_{ij}^{+} - h_i w_{ij}^{-}} R_j \tag{3.7}$$

Where,

R_i – Input layer relevance score

x_i – Neurons at the input layer

l_i - Lower bounds of pixel values, i.e., -1

h_i - Upper bounds of pixel values, i.e., +1

The uniform LRP-0 rule produces explanations that are complex and not faithful, or easy to understand. This rule picks up some of the local artifacts. Uniform LRP-ε rule produces explanations that are noise-free. It considers limited matching features and produces faithful explanations. But these explanations are not easy to understand. The uniform LRP- γ rule produces easy-to-understand explanations, but at the same time, some unrelated concepts are picked by this rule. Because of all these disadvantages associated with the usage of a single rule, the composite LRP rule is used in this research study. In the upper layers, the LRP-0 rule is used. At these layers, many concepts from different classes are entangled. LRP-0 rule is insensitive to these entanglements. Therefore, the LRP-0 rule is implemented at these layers. At the middle layers, the LRP-ε rule is used. These layers may have spurious variations that are filtered by the LRP-ε rule. Therefore, the LRP-ε rule is implemented at these layers. At lower layers, the LRP-γ rule is implemented. This rule uniformly spreads relevance to the whole feature. It produces explanations that are easy to understand. The input layer receives input in a different form than the intermediate layers. For intermediate layers, Rectified Linear Unit (ReLU) activations are input. Whereas for input layers, the input is pixel or real values. Hence, these layers require a special rule. In this research study, the Z^B rule is implemented at these layers. Thus, the LRP composite rule, i.e., the combination of LRP-0, LRP-ε and LRP-γ rule, is implemented in the VGG16 network to explain the epileptic seizure signal identification by the network.

3.2.3 Comparison of the XAI methods SHAP, Guided Backpropagation and LIME

3.2.3.1 Data preprocessing

This research study classifies EEG signals as seizure and non-seizure and explains the classification results. Therefore, data from sets B and E are used to conduct the research. Table 3.5 describes the data used in the research.

Table 3.5 Data Used in Research Study (Rathod et al. 2025b)

Signal Type	Set	Setup	Time Duration
Non-seizure	B	Surface EEG	39.33 min.
Seizure	E	Intracranial EEG	39.33 min.

As described in Section 3.2.2.1, 4400 signal segments are obtained from these two sets. By applying STFT to them, these signal segments are converted into spectrogram images. Fig. 3.6a shows a sample seizure signal of 23.6 sec. duration, and Fig. 3.6b shows its spectrogram after applying STFT to it. Similarly, Fig. 3.7a shows a sample non-seizure signal of 23.6 sec. duration and Fig. 3.7b shows its spectrogram.

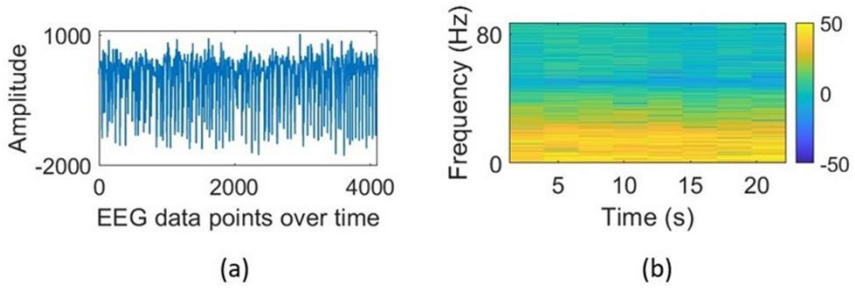


Fig. 3.6 Seizure signal and its spectrogram. Fig. a shows a seizure signal of 23.6 sec. duration. It has 4097 data points over time. Fig. b shows the spectrogram of Fig. a (Rathod et al. 2025b)

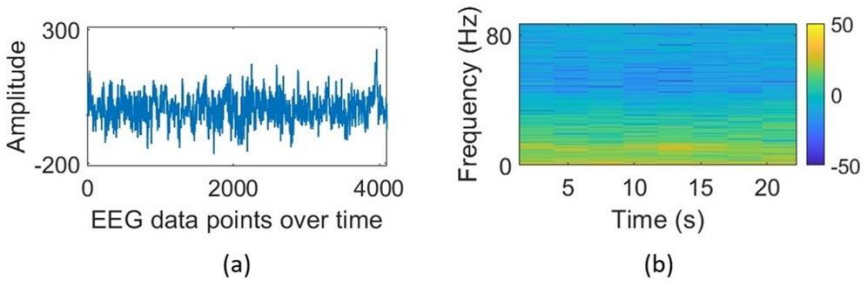


Fig. 3.7 Non-seizure signal and its spectrogram. Fig. a shows a non-seizure signal of 23.6 sec. duration. It has 4097 data points over time. Fig. b shows the spectrogram of Fig. a (Rathod et al. 2025b)

As outlined in Section 3.2.2.1, the classifier discriminates between seizure and non-seizure images based on colour variations observed in the lower section of the spectrogram. Moreover, the XAI methods use this colour difference to explain the classification results.

3.2.3.2 Proposed method

The architecture of the system is shown in Fig. 3.8. EEG signals are segmented into small pieces. Each piece is 2 sec. duration long. STFT is applied to these segments to obtain spectrogram images. The dataset of 4400 seizure and non-seizure spectrogram images is created. The VGG16 network is used to classify these images. Images are resized to 224 × 224 pixels before feeding them to the VGG16 network. A fine-tuning approach to transfer learning is applied using the VGG16 network for classification.

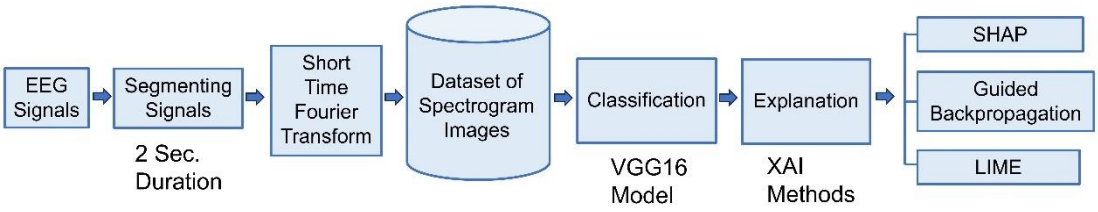


Fig. 3.8 Block diagram of the system. EEG signals are segmented into 2 sec. duration pieces and converted into images using STFT. Classification is performed using the VGG16 model. The performance of three XAI methods - SHAP, Guided Backpropagation, and LIME – is compared to explain the classification results (Prajakta Rathod et al. 2025b)

As described in Section 3.2.2.2, the VGG16 model is used to initialise the weights of the CNN model. Then, the EEG image dataset is used to train the CNN model from scratch. The CNN model's output layer is modified to identify images from the two classes: seizure and non-seizure.

The dataset is separated into training and testing sets, with 80% used for training and 20% for testing. Single-stage cluster sampling is used to select samples. The entire dataset is segregated into five clusters. The proportion of images from each class is maintained across all five clusters, consistent with the original dataset. One cluster is selected at random for testing the model. The model is trained using the remaining four clusters. During training, all hyperparameters are kept the same as in Section 3.2.2.2. The network is trained to classify images into seizure and non-seizure classes. After classification is performed, XAI methods are used to explain the classification results of the network. These methods marked the pixels or regions in the image that have a positive contribution to image classification in a particular class. Three different XAI methods used to explain the classification results are SHAP, guided propagation and LIME.

3.2.3.3 Working of XAI methods

This section describes the working of the XAI methods. XAI method SHAP is described in Section 3.2.5.2.2.1. The XAI methods Guided backpropagation and LIME are described below.

3.2.3.3.1 Guided Backpropagation

Guided Backpropagation is an XAI technique used to visualise and interpret the inner workings of DNNs. It helps understand which parts of an input image contribute most to a model's decision. The method is based on the process of backpropagation, which is normally used during training to compute gradients of the loss function with respect to the input. In guided backpropagation, the gradient flow is modified in a special way during the backward pass. Normally, during standard backpropagation, both positive and negative gradients are propagated through the ReLU activation function. However, guided backpropagation changes this rule by allowing only positive gradients to pass through neurons that had positive activations in the forward pass. This dual gating mechanism - considering only positive activations and positive gradients - ensures that only the features that positively influence the output are highlighted.

As a result, Guided Backpropagation suppresses the negative or irrelevant gradients, focusing on the most important input features responsible for the final classification. It combines the advantages of standard backpropagation and the deconvolutional network approach to produce sharper, more meaningful saliency maps. These maps visually depict which parts of the input image most strongly influenced the network's decision, providing a clearer understanding of the decision-making process. In the context of EEG-based epilepsy detection, Guided Backpropagation can reveal which frequency components the model relies upon to distinguish between epileptic and non-epileptic signals. The method enhances model interpretability without altering its architecture or accuracy.

3.2.3.3.2 LIME

The LIME method is one of the most widely used XAI techniques for understanding the decision-making process of complex black-box models. LIME is designed to generate human-understandable explanations for individual predictions made by any ML or DL model, regardless of its internal architecture.

The central idea behind LIME is that while a complex model may be globally non-linear and difficult to interpret, it can be approximated locally around a specific prediction using a simpler, interpretable model. This local approximation allows users to understand why a model predicted a certain class for a particular instance. LIME works by slightly perturbing the input data around the instance of interest and observing how the model's output changes in response to these perturbations. By creating multiple perturbed samples, LIME builds a dataset that represents the local neighbourhood of the original instance. A simple, interpretable model, such as a linear regression or DT, is then trained on this locally sampled data to approximate the behaviour of the complex model near that specific point.

Each feature of the input data is assigned a weight or importance score based on how much it contributes to the prediction in this local region. For image classification tasks, LIME divides an image into several segments known as superpixels and perturbs them by turning certain regions on or off to assess their influence on the model's output. The regions that most strongly affect the prediction are highlighted visually, helping users understand which parts of the image were most influential in the model's decision. Overall, LIME bridges the gap between complex DL architectures and human interpretability by providing transparent, locally faithful explanations for individual predictions, making it a powerful tool in the field of XAI.

3.2.4 SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0

3.2.4.1 Data preprocessing

An open-access Bonn University dataset is used in this research study. Previous methods identified epileptic seizure and non-seizure signals. The identification of interictal signals also plays a major role in epilepsy treatment plans. Therefore, this method distinguishes between epileptic seizure, interictal, and non-seizure EEG patterns, and utilises SHAP, the XAI method, to attribute and visualise feature contributions to the classification decisions. Hence, three sets, B, D, and E, are used in this research study. Sets B, D and E contain non-seizure, interictal, and epileptic seizure signals, respectively. Table 3.6 describes the sets used in this research study.

Table 3.6 Data Used in Research Study (Prajakta Rathod et al. 2025a)

Set	Signal Type	Duration
B	Non-Seizure	39.33 min.
D	Interictal	39.33 min.
E	Seizure	39.33 min.

As described in Section 3.2.2.1, small pieces of 2 sec. duration are obtained from the EEG signals. These segments are converted into spectrogram images using STFT. Sets B, D and E gave 2200 signal segments each. Therefore, a total of 6600 images are used for the classification of three classes. In this method, a comparative analysis of four CNN models is performed to identify the best-performing model for the identification of interictal, non-seizure and epileptic seizure class signals. Moreover, the best-performing model with the XAI method SHAP is identified. These four CNN models are Resnet18, Densenet121, VGG16 and Squeezenet1_0.

Sample EEG signals of 5-second duration from the interictal, non-seizure, and epileptic seizure classes are shown in Fig. 3.9 a, b, and c, respectively. Fig. 3.9 d, e, and f show their corresponding spectrograms. The interictal and non-seizure signal has a lower amplitude than the epileptic seizure signal, as shown in Fig. 3.9 a, b and c. The EEG signals' amplitude increases during a seizure attack. Moreover, when interictal and non-seizure signals are compared for amplitude, it is observed that the non-seizure signal has a higher amplitude in a higher frequency region than interictal signals.

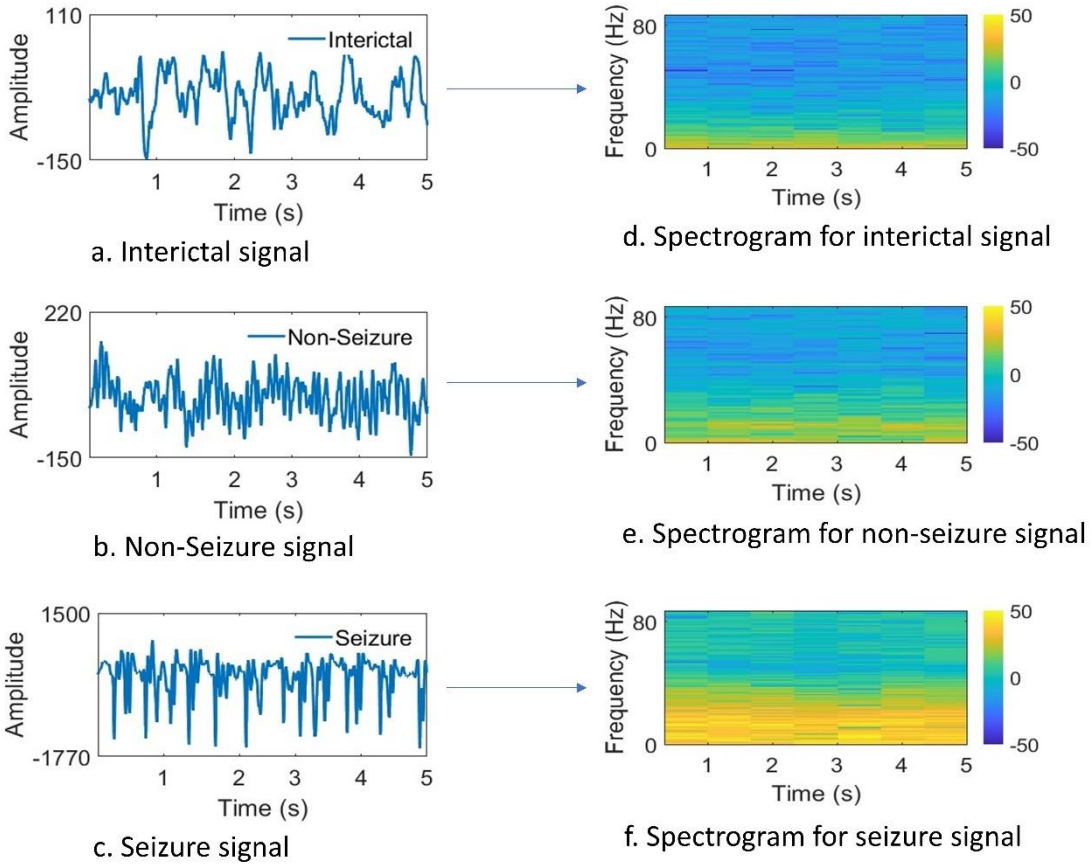


Fig. 3.9 a, b and c represent interictal, non-seizure and seizure signals of 5 sec. duration and d, e and f represent their spectrograms, respectively. Colours in the spectrogram image represent the amplitude of a signal. The yellow colour in the image shows a higher amplitude, and the blue colour shows a lower amplitude (Prajakta Rathod et al. 2025a)

As described in Section 3.2.2.1, the lower portion, i.e., the lower frequency region of spectrogram images, shows the main difference among interictal, non-seizure and seizure signals. The yellow colour in spectrogram images, which indicates amplitude, is more in seizure signals than in non-seizure and interictal signals. This difference in amplitude is learned by the classifier and used to identify signals from three classes. All spectrogram images are normalised before they are used to train the networks.

Fig. 3.10 a, b and c show 2 sec. duration spectrogram images representing interictal, non-seizure, and epileptic seizure classes, respectively. The corresponding normalised versions of these images are presented in Fig. 3.10 d, e and f.

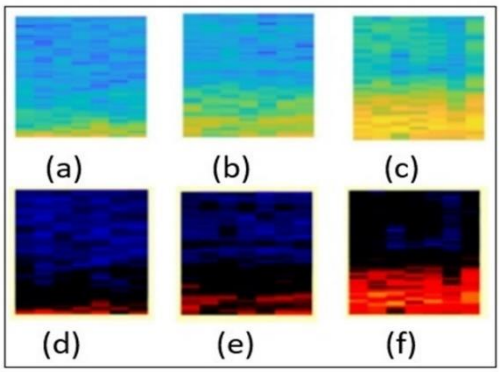


Fig. 3.10 a, b and c represent spectrogram images for interictal, non-seizure and epileptic seizure signals of 2 sec. duration, respectively, before normalisation, and d, e and f represent the corresponding normalised versions of these images (Prajakta Rathod et al. 2025a)

3.2.4.2 Proposed method

The workflow of the proposed system is shown in Fig. 3.11. EEG signals are converted into small segments of 2 sec. duration, which is converted into spectrogram images using STFT. This dataset of images is used to train four CNN models, Resnet18, Densenet121, VGG16 and Squeezenet1_0, to identify interictal, non-seizure and

epileptic seizure signals. Classification results are explained using the XAI method SHAP. The performance of all four networks for EEG signals classification, and the SHAP method results are analysed and compared to identify the best-performing model.

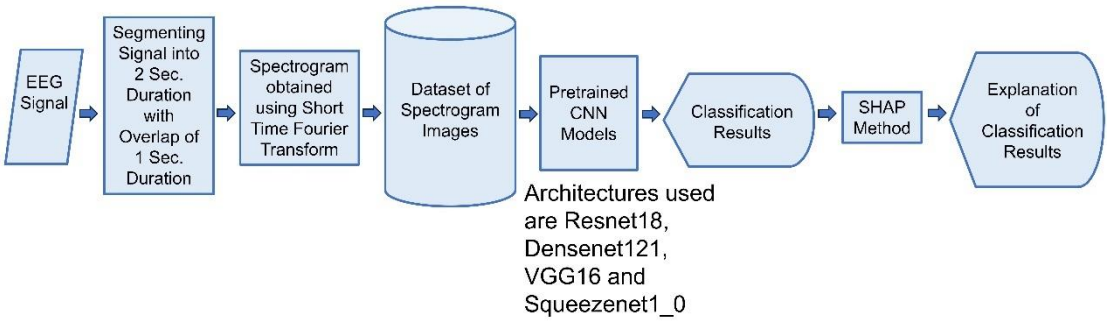


Fig. 3.11 Flow diagram of the system. EEG signals are preprocessed using STFT to obtain spectrogram images. Classification is performed using four CNN models, and the explainable AI method SHAP is used for explaining the classification results (Rathod et al. 2025a)

The EEG image dataset is used to train all four CNN models from scratch using a transfer learning approach. During training, the fully connected layer is modified and set to identify images from three classes: interictal, non-seizure and epileptic seizure class.

As described in Section 3.2.3.2, 80% of the data is allocated for training and the remaining 20% for testing. The training and testing sets are obtained using the single-stage cluster sampling. The dataset is partitioned into five clusters. One cluster is selected at random for testing, and four clusters are selected for training. During training, all hyperparameters are kept the same as in Section 3.2.2.2.

3.2.4.3 Validation methods

5-fold cross-validation experiments are performed on a training set to evaluate network performance. A total of 20 experimental runs are carried out, and the average result is used for evaluation. Among all four networks, Resnet18 is found to be achieving the highest accuracy. Therefore, the training set undergoes 10-fold cross-validation to evaluate Resnet18 once more. The experiments are conducted with 20 repetitions, and the average of all 20 repetitions is taken. To evaluate all four networks' performance on unseen data, a test set is used. For that, an entire training set is used to train all four networks again, and a test set is used to test them.

3.2.5 SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channel

3.2.5.1 Data pre-processing

In this research study, the TUH dataset is used, which is described in Table 3.2. EEG signals are pre-processed in four steps: (i) Artefact removal using a band-pass filter, (ii) Application of STFT to obtain images from signals, (iii) Vertically concatenating all 21 channels, and (iv) Applying a filter to images to improve the visual quality of images. A comprehensive overview of these steps is presented in the subsequent sub-sections.

3.2.5.1.1 Band-pass filtering

EEG signals obtained from the dataset may contain some artefacts like eye movement artefacts, eye blink artefacts, muscle artefacts, and 50 Hz Line noise. These artefacts should be removed before further processing of signals. Therefore, a band-

pass filter with a frequency range of 0.1 to 45 Hz is applied to eliminate these artefacts. The reason behind selecting this frequency range is that the frequencies below 0.1 Hz typically contain slow drifts and baseline fluctuations caused by electrode movement, perspiration, and sensor instability, which can obscure meaningful neural information. Similarly, frequencies above 45 Hz are often dominated by muscle artefacts and external electrical interference, such as 50 Hz power line noise. The selected range effectively preserves the clinically relevant EEG frequency bands, including delta (0.5 - 4 Hz), theta (4 - 8 Hz), alpha (8 - 12 Hz), beta (12 - 30 Hz), and low gamma (30 - 45 Hz), which are crucial for identifying epileptic patterns. This filtering ensures that transient epileptic discharges and rhythmic abnormalities are clearly represented without distortion. Furthermore, the 0.1 - 45 Hz range enhances the signal-to-noise ratio, improving both visual interpretability and the reliability of subsequent feature extraction. The chosen limits are consistent with standard EEG preprocessing practices recommended in clinical and research studies. Overall, this frequency range provides an optimal balance between artefact removal and retention of physiologically significant brain activity for accurate epileptic signal classification. To minimise the processing overload, the data originally sampled at 250 Hz is down-sampled to 100 Hz.

3.2.5.1.2 Signal to image conversion using STFT

In the original dataset, during the data collection process, different subjects' EEG data are recorded with different numbers of channels, and channel names are also different. Therefore, we consider recordings from the common 21 channels, which are present in all EDF files. Names of these common 21 channels are, FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, A1, A2, FZ, CZ, PZ.

EEG signals are divided into small pieces of 4 sec. duration without overlap. In the previous methods, the EEG signals are divided into segments of 2 sec. duration. The Bonn dataset used in the previous methods has a comparatively smaller amount of data. In order to obtain a larger number of data samples for the model training, signals are divided into 2 sec. duration segments with an overlap of 1 sec. The TUH dataset is comparatively larger than the Bonn dataset. Therefore, EEG signals are divided into 4 sec. duration segments without overlap, as enough data samples could be obtained for model training with this size as well.

From each patient's EDF file, 20 minutes of recording is processed to carry out experiments. Hence, 300 small pieces of signals, i.e., 300 samples, are obtained from 20 minutes of recording from one patient. The duration of each sample is 4 sec. From 100 patients, 30,000 samples are obtained, among which 15,000 are epileptic and 15,000 are non-epileptic. STFT is applied to these small segments to convert them into images. The parameters used to compute STFT are a Hamming window of length 256, 50% overlap, and a fast Fourier transform length of 256.

3.2.5.1.3 Channels concatenation

EEG data consists of 21 channels. Initially, a spectrogram for an individual channel is obtained. Then, spectrograms of all 21 channels are concatenated vertically. Fig. 3.12(a) and Fig. 3.12(b) show a sample epileptic signal and its spectrogram. Similarly, Fig. 3.13(a) and Fig. 3.13(b) show a sample non-epileptic signal and its spectrogram. The sample epileptic and non-epileptic signals have a duration of 4 sec. In the spectrogram of the EEG signal, frequency is present on the X-axis, and vertically concatenated 21 channels are present on the Y-axis. The power distribution of the signal

over different frequencies and channels is depicted using colours. As the power of the signal changes from low to high, the colour in the image changes from blue to red. The colorbar range is -230 to -10. The spectrogram is created using the 'Jet' colormap.

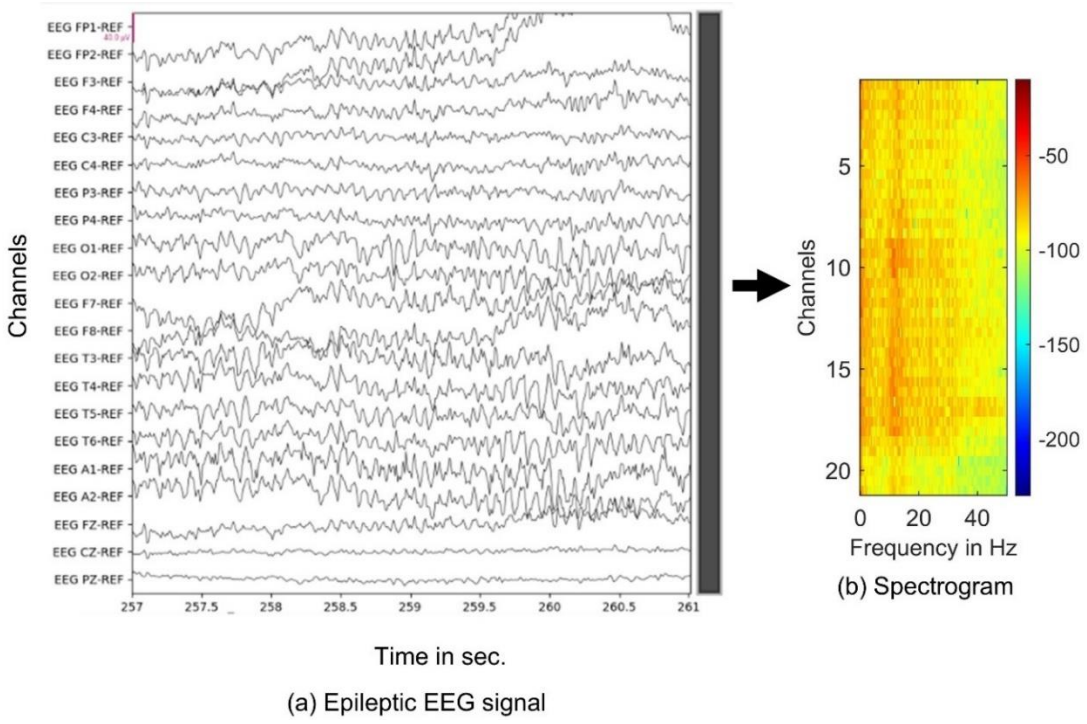


Fig. 3.12 (a) represents an epileptic signal of 4 sec. duration with 21 channels used in the study. (b) represents its spectrogram. The channels in the spectrogram are in the same sequence as that of Fig. 3.12 (a). Colours in the spectrogram image represent the power of a signal at different frequencies in different channels. The red colour in the spectrogram image shows a higher power, and the blue colour shows a lower power (Prajakta Rathod et al. 2025)

3.2.5.1.4 Image filtration

Medical images can be enhanced before analysis to aid in disease diagnosis. Enhancing medical images is a vital process that improves their visual quality, thereby enhancing the human evaluator's ability to interpret them accurately. According to the existing body of research, a wide range of image enhancement techniques have been employed to refine the visual quality of medical images for better interpretation and

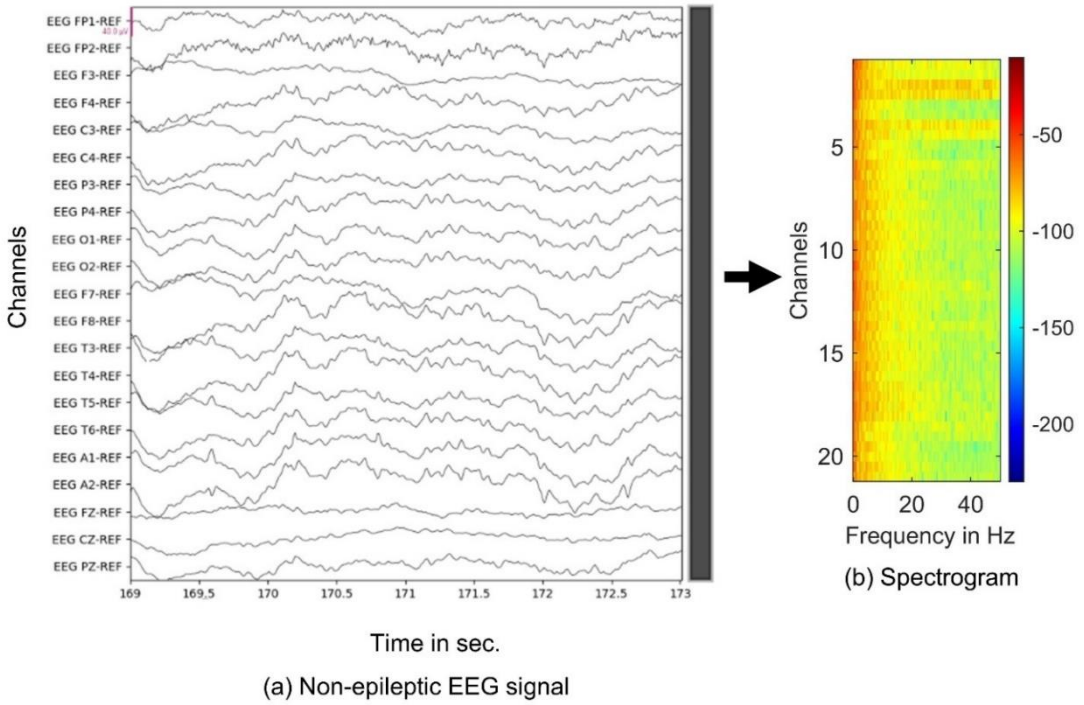


Fig. 3.13 (a) represents a non-epileptic signal of 4 sec. duration with 21 channels used in the study. (b) represents its spectrogram. The channels in the spectrogram are in the same sequence as that of Fig. 3.13 (a). Colours in the spectrogram image represent the power of a signal at different frequencies in different channels. The red colour in the spectrogram image shows a higher power, and the blue colour shows a lower power (Prajakta Rathod et al. 2025)

analysis (Salem et al. 2019) (Islam and Mondal 2019). This study applies sharpening and contrast enhancement techniques to spectrogram images in order to improve the visual distinction between epileptic and non-epileptic signals. These methods are described in Subsections 3.2.5.1.4.1 and 3.2.5.1.4.2.

3.2.5.1.4.1 Sharpening

The Matlab function B=imsharpen(A) is used to sharpen the spectrogram images. In this function, the input argument A is the spectrogram image. It takes an input image in RGB format and sharpens it using the unsharp masking method ("Imsharpen," n.d.). This function takes three arguments: radius, amount and threshold. The value of the radius argument is one. The amount argument represents the

sharpening effects' strength. Its value is 0.8. The minimum contrast value should be specified for the threshold argument if the pixel needs to be considered as an edge pixel. This minimum value is zero. A sharpened image is returned in the form of a numeric array by this method. The unsharp masking method subtracts the blurred version of an image from itself.

Three colour channels, R, G and B, are present in the RGB image. Before applying sharpening, the RGB image is transformed to the L*a*b colour space. In L*a*b colour space, channel L is dedicated to luminosity, and two channels, a and b, are dedicated to colour information. Because of this, luminosity and colours can be processed separately without affecting each other. During sharpening, only the luminosity layer is used. After sharpening, the image is converted back to the RGB colour space.

3.2.5.1.4.2 Contrast enhancement

After sharpening, the spectrogram image is further enhanced by applying contrast enhancement using histogram equalisation ("Contrast Enhancement Techniques," n.d.). The histogram represents the intensity distribution of an image graphically. Each intensity value has some pixels. In the histogram, the distribution of pixel counts across different intensity levels is illustrated. In the histogram equalisation, the intensity range of an image is stretched out. Some intensity values occur frequently in an image. In histogram equalisation, these intensity values get stretched.

Before contrast enhancement is performed, the spectrogram image is converted to the L*a*b colour space using the MATLAB function lab = rgb2lab(RGB). The input argument to this function is a spectrogram image in sRGB format, which is converted

to L*a*b* values (“Rgb2lab,” n.d.). Contrast adjustment is performed on the luminosity layer. The manipulation performed on the luminosity layer only affects the intensity of the pixels and helps to preserve the original colours in the image. The range of luminosity values is 0 to 100. These values are scaled to the range [0 1]. The Matlab function J = histeq(L) is used to perform histogram equalisation (“Adjust Image Contrast Using Histogram Equalization,” n.d.). The input argument to this function is the luminosity channel L. After histogram equalisation, luminosity values are again rescaled to the range 0 to 100. Then, the Matlab function rgb=lab2rgb(lab) is applied to an image to convert it to the RGB colour space. This function converts L*a*b values to sRGB values (“Lab2rgb,” n.d.).

Sample images of epileptic and non-epileptic signals are presented in Fig. 3.14(a) and Fig. 3.14(d), respectively. After applying the sharpening technique, the enhanced versions of these images are shown in Fig. 3.14(b) and Fig. 3.14(e). Subsequently, Fig. 3.14(c) and Fig. 3.14(f) display the results after contrast enhancement is applied to the sharpened images in Fig. 3.14(b) and Fig. 3.14(e), respectively.

During spectrogram image creation, X-axis and Y-axis labels, i.e., frequency range and channel numbers, are made invisible. This helps to provide a proper visual appearance of images after applying filters to them.

3.2.5.2 Proposed method

The proposed method performs classification of epileptic and non-epileptic signals and utilises the SHAP method of XAI for explanation of the results. This proposed approach is structured into four key steps: (i) EEG data pre-processing, (ii)

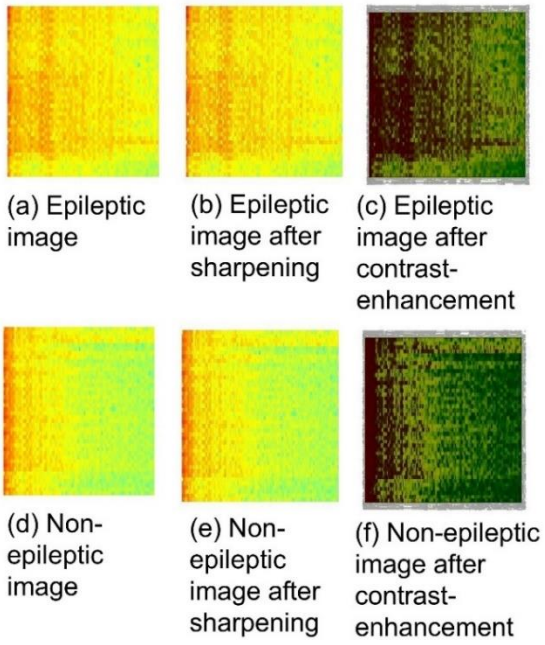


Fig. 3.14 (a) and (d) show epileptic and non-epileptic spectrogram images, respectively. Fig. (b) and (e) show epileptic and non-epileptic spectrogram images after sharpening. Fig. (c) and (f) show epileptic and non-epileptic spectrogram images after applying contrast enhancement (Prajakta Rathod et al. 2025)

classification using DL models, (iii) explanation of classification results using the XAI method SHAP, and (iv) clinical relevance of technical findings. The following sub-sections provide a detailed explanation of all these steps. The system's flow diagram is presented in Fig. 3.15.

3.2.5.2.1 Classification

The proposed method uses three CNN models, Densenet121, Resnet18 and VGG16, to classify EEG signals into epileptic and non-epileptic categories. Transfer learning with fine-tuning is applied to each of the three CNN models. EEG signals are preprocessed to remove artefacts. Then, STFT is applied to these signals to convert them into spectrogram images. These images are resized to 224 × 224 pixels and input to three CNN models. The visual quality of images is improved by applying filters to

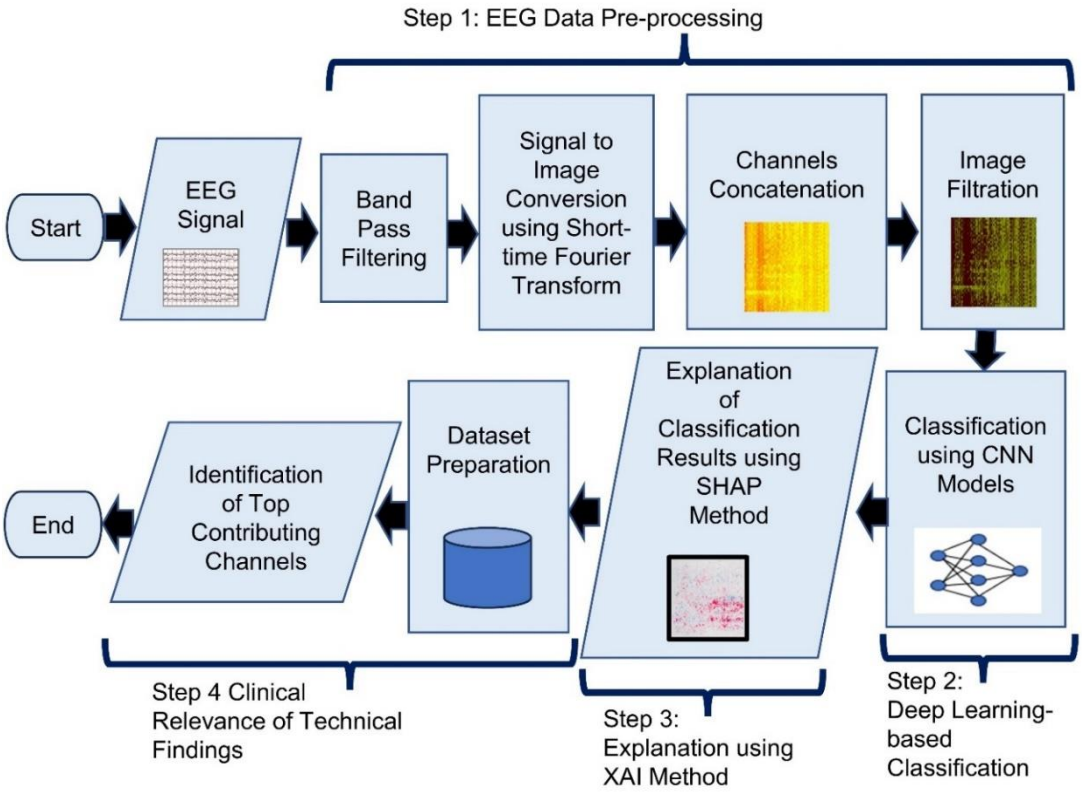


Fig. 3.15 Flow diagram of the proposed system. EEG signals are pre-processed using STFT to obtain spectrogram images, which are then filtered to enhance their visual quality. Three CNN models - Densenet121, Resnet18 and VGG16- are used for classification, and the explainable AI method SHAP is used for explaining the classification results. The heatmap generated using the SHAP method is evaluated using the AOPC value. Top contributing channels in the classification of epileptic and non-epileptic patients are identified using the relevance matrix generated by the SHAP method (Prajakta Rathod et al. 2025)

these images. The architecture of the VGG16 network is shown in Fig. 3.16. The three CNN models are trained from scratch on an EEG spectrogram image dataset having images from the two classes, epileptic and non-epileptic, as described in Section 3.2.2.2. During training on EEG data, a fully connected layer is updated to identify images from epileptic and non-epileptic classes. Pre-trained weights of the networks are fine-tuned using backpropagation. During training, all hyperparameters are kept the same as in Section 3.2.2.2. The data is split into 80% for training and validation, and 20% as an untouched test set. The performance of these three CNN models is analysed and compared for the classification of EEG signal images.

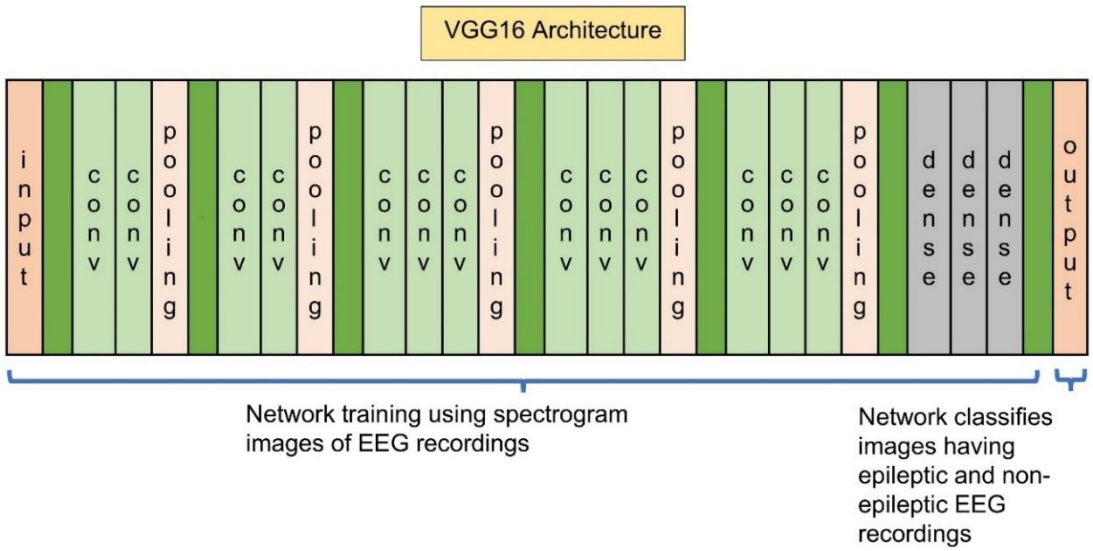


Fig. 3.16 Architecture of the VGG16 network. Network layers are trained on image datasets of epileptic and non-epileptic EEG signal images. The output layer is changed to identify two classes (Prajakta Rathod et al. 2025)

3.2.5.2.2 Explanation

The predictions generated by CNN models are explained by the XAI method SHAP. Some features in the spectrogram image have a positive contribution, and some features have a negative contribution in classifying images in a particular class. Shapley value is calculated by the SHAP method for these features according to their contribution. Different colours are used to represent the positive and negative contributions of features in image classification in a particular class. The SHAP method is described in detail in the following subsection.

3.2.5.2.2.1 Working of the SHAP method

In this research study, the SHAP method (Lundberg and Lee 2017), from the field of XAI, is used to provide an explanation for the identification of a particular image as epileptic or non-epileptic by the CNN model. The concept of Shapley value, which is calculated by the SHAP method, comes from game theory. Suppose n players

are playing a game. In the game, a coalition of players participates together to obtain certain rewards. The contribution of each player in the coalition is different. Some players have more contributions than others. The Shapley value find out the amount of contribution from each player. According to the contribution, the obtained reward is distributed fairly among the players. This concept of Shapley value is used in the CNN model. Predictions generated by the network bear a resemblance to the game. One prediction is similar to one game. Different features of the data bear a resemblance to the players in the game. Each feature's contribution to making an output prediction is determined by the SHAP value. Thus, the SHAP method is used to determine each feature's contribution, positive or negative, in identifying the class of a particular image.

The contribution of each feature is determined by all possible coalitions of features. This is known as a power set of features, comprising 2^n possible coalitions when there are a total of 'n' features. For each coalition of features, a separate model is required to be trained. Hence, the number of models required is 2^n.

Suppose all features of the data are represented by the set F. When Shapley regression values are used by the method, then it requires the model to be trained on all possible subsets S of features, where $S \subseteq F$. It assesses the extent to which each feature impacts the prediction made by the model. Two models are trained to compute this contribution. Suppose i is the feature for which the contribution is to be computed; then two models are trained, one with the feature set $f_{s \cup \{i\}}$ and another with f_s excluding feature i. Suppose the values of the input features present in set S are represented by X_s. For a current feature under consideration, two models' prediction is compared, with and without the feature. The difference value is given by $f_{s \cup \{i\}}(X_{s \cup \{i\}}) - f_s(X_s)$.

This difference value is computed for every subset of features $S \subseteq F \setminus \{i\}$. When the model's prediction is obtained without a feature, it is affected by the other features present in the model. Equation 3.8 gives the weighted average for all possible differences.

$$\varphi_i = \sum_{S \subseteq F \setminus \{i\}}^{\infty} \left\{ \left\{ \frac{|S|!\,(|F|-|S|-1)!}{|F|!} \right\} \times \left\{ f_{s \cup \{i\}}\left(X_{s \cup \{i\}}\right) - f_s\left(X_s\right) \right\} \right\} \quad (3.8)$$

Sampling approximation is applied to equation 3.8. The feature under consideration is removed from the model, and its effect is approximated by integrating over the samples from the training dataset. In this case, less than $2^{|F|}$ differences are required to be computed. Also, retraining of the model is not required. Any model can be explained using these Shapley sampling values. The heatmap generated by the SHAP method has different values for different features of the dataset. This heatmap is evaluated qualitatively and quantitatively. The deepExplainer method present in the SHAP Python package is used to implement the SHAP method.

3.2.5.2.3 Clinical relevance of technical findings

We have identified potential EEG channels that have a higher contribution in determining whether a signal is epileptic or non-epileptic. For this purpose, the results from the XAI method SHAP are used for analysis. The process is divided into two steps: (i) dataset preparation and (ii) identification of top-contributing channels. The following sub-sections provide a detailed description of these steps.

3.2.5.2.3.1 Dataset preparation

The test dataset is used to identify the top contributing channels. We ensure that images from all 100 subjects (50 epileptic and 50 non-epileptic) are included in the dataset to conduct experiments. Therefore, images from one epileptic and one non-epileptic patient are put in one test set. 50 such test sets are created. Images in all 50 test sets are classified using the CNN model, and classification results are explained using the SHAP method.

3.2.5.2.3.2 Identification of top contributing channels

The steps in identifying the top contributing channels are shown in Fig. 3.17. Each filtered spectrogram image has 21 channels. Pixels corresponding to each of these 21 channels are identified. X and Y coordinates for these pixels are obtained. Spectrogram images are resized to 224 × 224 during model training. When the SHAP method is used for explaining the results, it generates a relevance matrix of size 224 × 224. This relevance matrix contains relevance values for all pixels in the filtered spectrogram image. One relevance value for one pixel in the spectrogram image. If a pixel has a positive contribution in classifying the image in a particular class, then its relevance value is positive. If a pixel has a negative contribution in classifying the image in a particular class, then its relevance value is negative. All pixels having positive relevance values are considered in the top contributing channel identification. For one image, all pixels in one channel are considered, and only positive relevance values of pixels are added together. This is done for all 21 channels of one image. Thus, 21 values for 21 channels are obtained for one image after summation. This process is

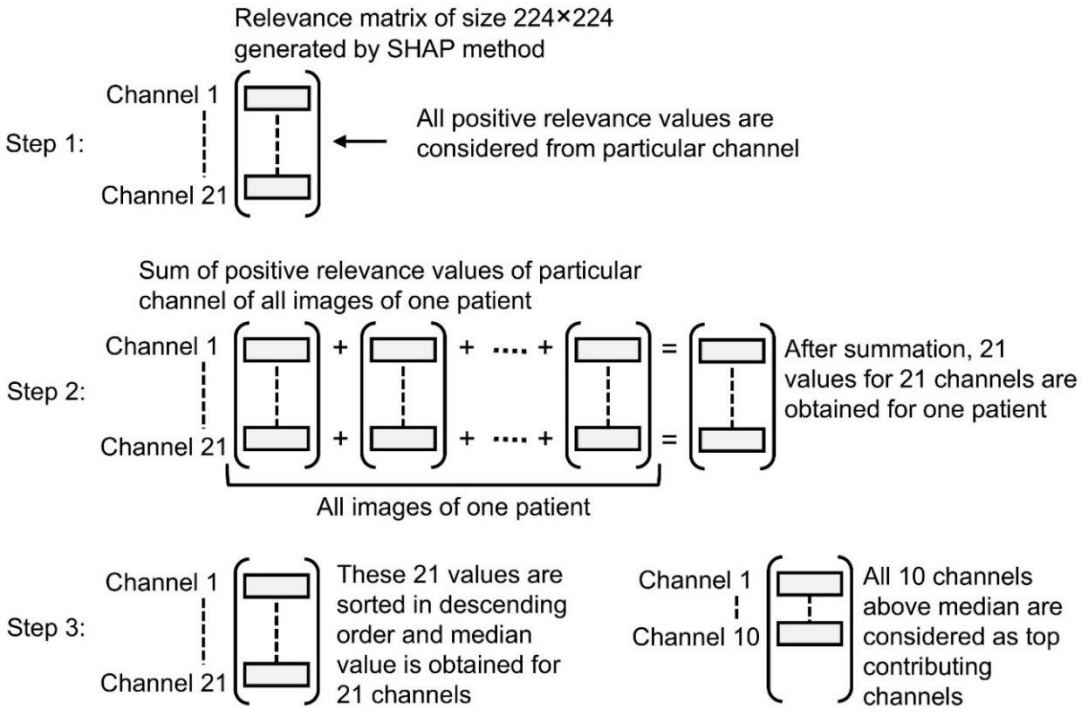


Fig. 3.17 Identification of the top contributing channels using the relevance matrix generated by the SHAP method. In step 1, all positive relevance values from a particular channel are considered. In step 2, all positive relevance values of a particular channel of all images of one patient are added. In step 3, these values are sorted in descending order, and the median value is obtained. Channels with relevance value above the median are considered as the top contributing channels (Prajakta Rathod et al. 2025)

repeated for all images of one patient. Then, values for a specific channel in all images are added together. This gives one value for one channel for all images of one patient. The process is applied to all channels in the same manner. Hence, 21 values are obtained for 21 channels for all images of one patient. The positive contribution of each of the 21 channels in patient identification as epileptic or non-epileptic is represented by these 21 values. The values are first ordered from highest to lowest, and then their median is determined. The channels whose relevance sum is above the median value are identified. These channels are selected as the top contributing channels. This process is repeated for all 50 test sets, each set having one epileptic and one non-epileptic patient. Thus, the top contributing channels are identified for all 100 subjects.

3.2.5.3 Validation methods

The classification performance of the CNN models is measured using the K-fold cross-validation method. In addition, the SHAP method's performance is evaluated through an Area Over Perturbation Curve (AOPC) analysis.

3.2.5.3.1 K-fold cross-validation

Data is divided into an 80% training set and a 20% testing set. Samples are selected at random. The 5-fold cross-validation technique is used to evaluate the performance on the training set. The dataset used for training was segmented into five folds. CNN models are trained using four folds. One fold is designated as the validation set in each iteration to evaluate the CNN models, and the process is repeated across all five folds. In five repetitions, five folds become a validation set once. Accuracy is noted on all five folds, and its average value is taken. This entire process is repeated 10 times, and the average of all 10 repetitions is found. To perform K-fold cross-validation, the KFold function present in the model_selection module of the Scikit-learn (Sklearn) Python library is used.

Among the three CNN models, VGG16 achieved the highest accuracy using 5-fold cross-validation. Afterwards, the VGG16 model is tested again using 10-fold cross-validation by dividing the training dataset into ten equal folds. The model is trained using nine folds of the data, and its performance is evaluated on the tenth fold, which acts as the validation set. The accuracies obtained from each of the 10 folds are averaged to obtain the final result. This entire process is repeated 10 times, and the average of all 10 repetitions is found.

After ten repetitions, the CNN model's performance is evaluated on unseen data. The evaluation involves training the CNN model on 80% of the dataset and assessing its performance on the remaining 20% test set. This process is repeated for all three CNN models. The results obtained using the three CNN models are described in Chapter 4.

3.2.5.3.2 Heatmap evaluation using AOPC values

The XAI method SHAP is used to explain classification results generated by three CNN models. The SHAP method generates the heatmap, which shows significant pixels in network prediction. In (Samek et al. 2017), a method to evaluate the heatmap generated by the XAI method is proposed. This method is used to evaluate the heatmap generated by three CNN models in the proposed research method. This heatmap has positive and negative relevance values for each pixel in the filtered spectrogram image. If the pixel is more significant in the classification decision, then its relevance value is higher, and for a less significant pixel, the relevance value is lower. These relevance values are sorted in descending order. Therefore, higher relevance values appear at the beginning of the sorted vector, and lower relevance values appear at the end of the sorted vector.

When these significant pixels in the image are perturbed, then it is expected that classification accuracy will also be reduced in proportion to the perturbation amount. Experiments are performed to perturb the significant pixels in iterative steps and measure the decrease in accuracy. During perturbation, a randomly generated value from the uniform distribution is used to replace the pixel value. 10 repetitive steps are used to carry out the perturbation. The first 1000 pixels using a sorted array of relevance

values are perturbed in the first iteration. The first 2000 pixels are perturbed in the second iteration, and so on. The first 10000 pixels are perturbed in the tenth iteration. The size of the image is set to 224*224 pixels during training. Therefore, an image consists of a total of 224*224=50176 pixels. 10000 pixels is 19.92% of the original image. In (Samek et al. 2017), 15.7% of pixels of the original image are perturbed to carry out similar experiments. Pixels perturbed in the proposed research study, i.e., 19.92%, are close to 15.7% used in (Samek et al. 2017).

Evaluation of the heatmap produced by the SHAP method is performed on the 20% test set. From this test set, 200 images -100 epileptic and 100 non-epileptic- are selected at random. The first classification accuracy is found on this 200-image set without perturbing any pixel. Then, the iterative steps of the perturbation process are carried out. In 10 iterative steps, 10000 pixels are perturbed. Pixel perturbation increases incrementally by 1000 in each step, starting with 1000 in the first step and reaching 10,000 by the tenth step. At last, 10 different perturbed image sets were obtained for 10 iterative steps. Classification accuracy is found for all 10 perturbed image sets. A reduction in classification accuracy with respect to the original image set without perturbation is obtained.

After repeating the process ten times, the average of the ten outcomes is taken. This helps to reduce the effect of randomness. 10 subsets are created to carry out 10 repetitions. Each subset contains 200 images, 100 epileptic and 100 non-epileptic. These images are selected at random from the test dataset. 10 repetitive steps are performed to perturb 10000 pixels of all images in these ten subsets.

A perturbation curve is plotted with perturbation steps on the X-axis and the average reduction in accuracy on the Y-axis. The XAI method SHAP generates a

heatmap when used to explain the network's classification decision. The quality of this heatmap is given by the AOPC value. If the AOPC value is bigger, then the quality of the heatmap is better. In the sorted array of relevance values, the most significant pixel values appear first, and the least significant pixel values appear last. When the most significant pixels are perturbed, the classification accuracy decreases quickly. This process of perturbing pixels and finding a reduction in classification accuracy is repeated for all CNN models. The AOPC curve is plotted for all CNN models. The results are described in Chapter 4.

In all the above methods, MATLAB is used to convert EEG signals into images using STFT. PyTorch is used to implement the proposed methods. Experiments are performed in Kaggle and Google Colab on a single GPU. The LRP method results for both epileptic and non-epileptic signals, as discussed in Section 3.2.1, are displayed using Matlab. In Section 3.2.5, images are enhanced by applying filters to them in Matlab.

The MNE Python library is used for analysing, processing, and visualising EEG data. For the implementation of the LRP method in the Bi-LSTM network TensorFlow library is used. For this purpose keras.models module in the Keras API is used. The os Python module is used to interact with the computer's operating system. The csv module is used to read from and write data to Comma-Separated Values (CSV) files. The Pandas library in Python is used for data analysis and data manipulation. The sklearn.model_selection module in the scikit-learn (or sklearn) Python library is used to split datasets for training, testing, validation, and perform cross-validation. The sklearn.preprocessing module in scikit-learn is used to transform the data.

Various packages from the torch and torchvision libraries from the PyTorch framework are used to perform different tasks. The torch. cuda module is used to enable GPU acceleration. The torchvision.models module is used to import pretrained models. The torch.nn PyTorch module is used to build and train DL models. The sklearn.metrics module in the scikit-learn library is used to evaluate the performance of ML models. The Python library numpy is used for mathematical operations. Python's built-in statistics module is used to perform basic statistical calculations. Python Imaging Library (PIL) is used for loading, editing, and saving images. The torchvision.transforms module is used to preprocess images before feeding them into a model. The shap library from the PyTorch framework is used to work with the SHAP method. The shap.DeepExplainer module is used to compute feature contributions (Shapley values) for each prediction. The shap.image_plot visualisation function is used to see which parts of the image contributed most to the model's prediction. The Matplotlib Python library is used to create a heatmap. The matplotlib.pyplot module is used to plot figures. The seaborn Python library is used to plot model performance metrics, e.g., a confusion matrix heatmap. Python's time module is used to measure durations and work with the current time. The torchmetrics, a PyTorch-friendly library, is used to compute model evaluation metrics directly on tensors, and to support GPU computation and batch updates.

A LIME, an XAI library, is used to implement the XAI method LIME for interpreting the predictions made by the DL model. The lime_image module of the LIME library is used to visualise which parts of an image contributed most to the model's decision for a particular class.

Summary:

This chapter presents the various methods implemented and experiments conducted to address multiple research objectives outlined in Section 2.2. The first method utilises a Bi-LSTM network to differentiate between epileptic and non-epileptic EEG signals. To enhance interpretability, the XAI method LRP is applied to the network's predictions. In the second method, EEG signals are transformed into spectrogram images to improve classification accuracy. A VGG16 network with transfer learning with fine-tuning is used to classify these images, and the LRP method is again used to interpret the network's decisions. The third method focuses on comparing three different XAI techniques - SHAP, Guided Backpropagation, and LIME - when applied to the VGG16 network. This comparison aims to identify the most effective method for explaining classifications of EEG patterns, with SHAP emerging as the most effective explainability method. In the fourth method, the performance of four CNN architectures - Resnet18, Densenet121, VGG16, and Squeezenet1_0 - is compared for classifying EEG signals into three categories: epileptic seizure, interictal, and non-seizure signals. This is accomplished through transfer learning with fine-tuning, and the classification results of all models are explained using the SHAP method.

All experiments in the first four methods are performed using the single-channel EEG dataset from Bonn University. The fifth method uses a multichannel EEG dataset from the TUH EEG corpus. Here, multichannel EEG signals are converted into spectrogram images using STFT. Spectrograms of all 21 channels are concatenated vertically. Images are enhanced by applying filters. A comparative analysis is then

conducted using three CNN models - Densenet121, Resnet18, and VGG16 - evaluating both classification performance and model explainability with SHAP. VGG16 demonstrates the highest classification accuracy, while Densenet121 offers the best explainability using SHAP. Therefore, SHAP with Densenet121 is used to identify the top contributing EEG channels in the classification of epileptic versus non-epileptic patients.

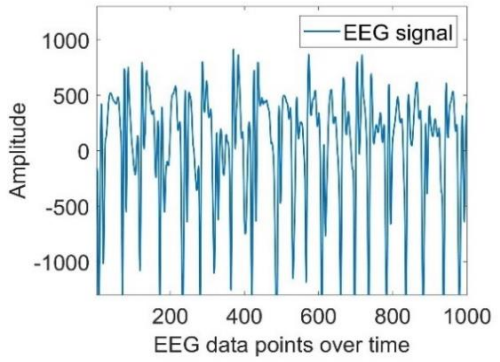
4. Epilepsy Detection Outcomes and Explainability Insights

4.1 LRP in Bi-LSTM network

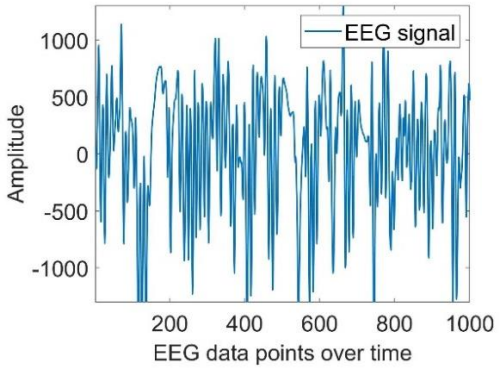
In this method, the Bi-LSTM network is trained using the EEG signal dataset, as described in Chapter 3, Section 3.2.1. The model correctly identified test input data and achieved an accuracy of 87.25%. Network prediction is explained by implementing the XAI method LRP in the Bi-LSTM network.

Fig. 4.1a, 4.1b, and 4.1c display sample epileptic signals, while the LRP method results for these images are provided in Fig. 4.2a, 4.2b, and 4.2c, respectively. Fig. 4.3a, 4.3b, and 4.3c present sample non-epileptic signals, while their respective LRP method results are displayed in Fig. 4.4a, 4.4b, and 4.4c. The X-axis shows the signal's data points over time, and the signal's amplitude is shown on the Y-axis. The duration of the signal is 5.76 sec. The data was collected with a sampling rate of 173.6 Hz. Hence, $5.76 \times 173.6 \approx 1000$ data points are shown on the X-axis.

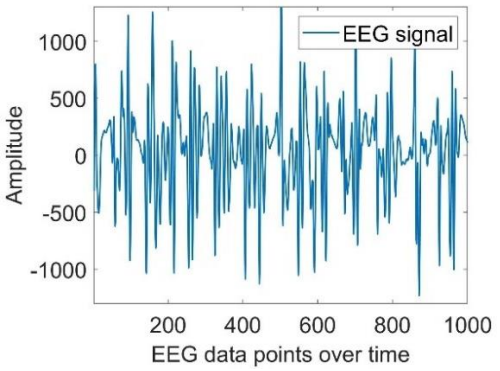
The relevance vector generated by the LRP method has relevance values, which are signed values. Normalisation is applied to these relevance values, mapping them to the range of 0.0 to 1.0. These values have positive and negative relevance. The values that lie in the range of 0.0 to 0.5 have negative relevance, and the values that lie in the range of 0.5 to 1.0 have positive relevance. A positive relevance value means the data point has a positive contribution in classifying the signal in a particular class. Similarly,



(a) Sample epileptic signal

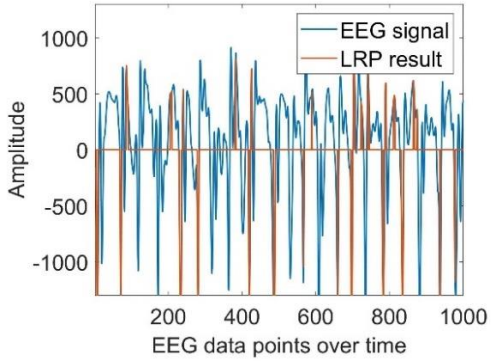


(b) Sample epileptic signal

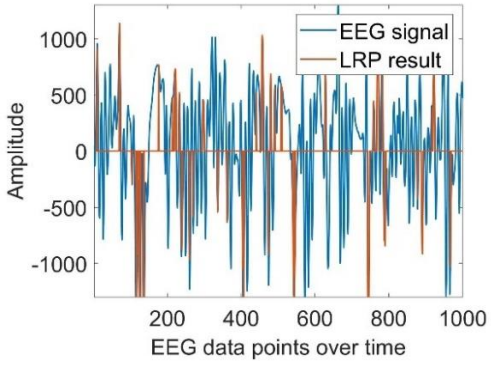


(c) Sample epileptic signal

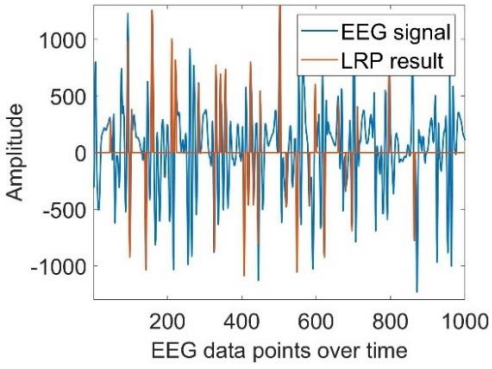
Fig. 4.1 Sample epileptic signals from the dataset. Signal data over time is plotted along the X-axis, and signal amplitude is shown on the Y-axis. The duration of the signal is 5.76 sec. i.e., 1000 data points (Prajakta Rathod et al. 2022)



(a) LRP result for sample epileptic signal 4.1a

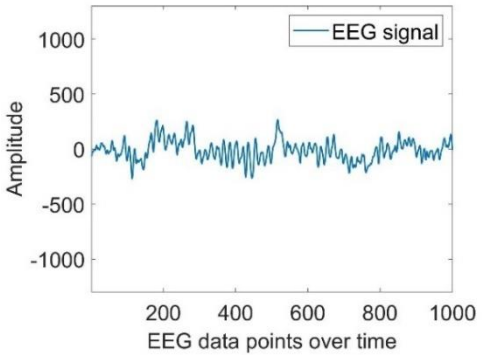


(b) LRP result for sample epileptic signal 4.1b

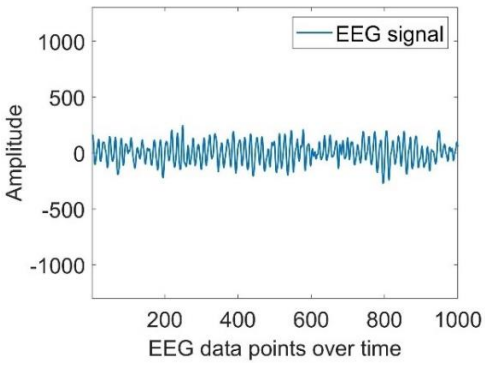


(c) LRP result for sample epileptic signal 4.1c

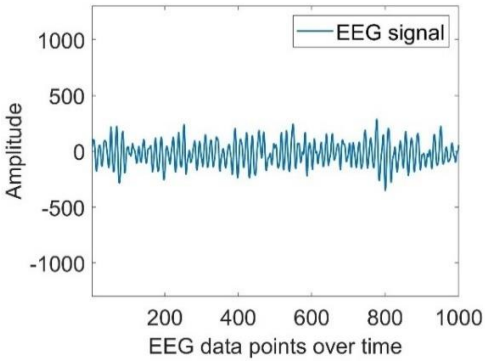
Fig. 4.2 LRP results for sample epileptic signals in Fig. 4.1. Signals are marked in blue colour, and LRP results are marked in red colour, indicating part of the signal classifying it in the epileptic class (Prajakta Rathod et al. 2022)



(a) Sample non-epileptic signal

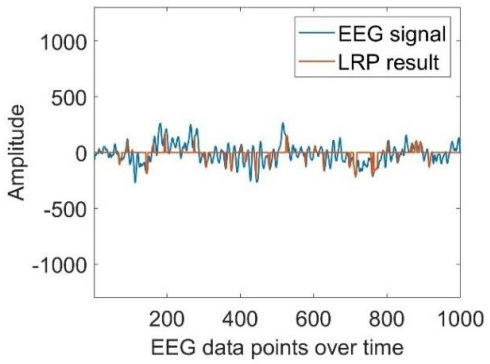


(b) Sample non-epileptic signal

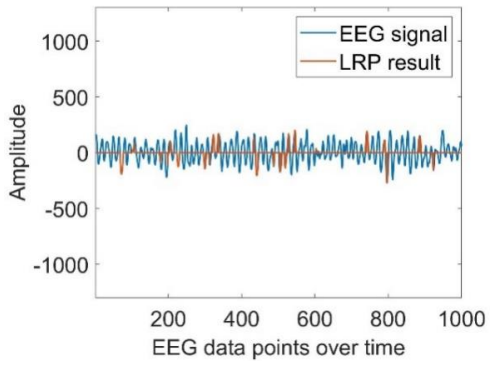


(c) Sample non-epileptic signal

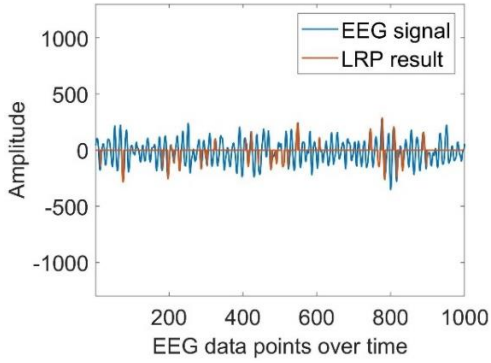
Fig. 4.3 Sample non-epileptic signals from the dataset. Signal data over time is plotted along the X-axis, and signal amplitude is shown on the Y-axis. The duration of the signal is 5.76 sec. i.e., 1000 data points (Prajakta Rathod et al. 2022)



(a) LRP result for sample non-epileptic signal 4.3a



(b) LRP result for sample non-epileptic signal 4.3b



(c) LRP result for sample non-epileptic signal 4.3c

Fig. 4.4 LRP results for sample non-epileptic signals in Fig. 4.3. Signals are marked in blue colour, and LRP results are marked in red colour, indicating part of the signal classifying it in the non-epileptic class (Prajakta Rathod et al. 2022)

negative relevance shows a negative contribution of data points in classifying the signal in a particular class. Data points that have positive relevance are marked in red by the LRP method. Fig. 4.2 and Fig. 4.4 display the results of the LRP method.

Fig. 4.1a, 4.1b and 4.1c show epileptic signals, and Fig. 4.2a, 4.2b and 4.2c show their LRP results, respectively. The epileptic seizure signal amplitude is higher than the non-epileptic signal. Amplitude has a significant role in epileptic and non-epileptic signal identification. The data points that correspond to higher amplitude in upward and downward directions are marked in red by the LRP method. In Fig. 4.2a, 4.2b and 4.2c, original signals are shown in blue, and the data points that correspond to higher amplitudes are marked in red. These data points are responsible for classifying the signals in the epileptic class. The relevance score generated by the LRP method is more than 0.5 for these data points.

Fig. 4.3a, 4.3b, and 4.3c depict non-epileptic signals, and their corresponding LRP outputs are shown in Fig. 4.4a, 4.4b, and 4.4c, respectively. The amplitude of the non-epileptic EEG signal is comparatively lower than the epileptic signal. The data points that correspond to lower amplitude in upward and downward directions are marked in red by the LRP method. In Fig. 4.4a, 4.4b and 4.4c, original non-epileptic signals are displayed in blue, and the data points corresponding to lower amplitude are marked in red. These data points are responsible for classifying the signals in the non-epileptic class. The relevance score generated by the LRP method is more than 0.5 for these data points.

4.2 LRP in VGG16 network

This study uses the VGG16 network to classify epileptic seizure and non-seizure EEG signals. Also, the XAI method LRP is implemented in the network to explain classification results. The VGG16 network gave 99.89% classification accuracy on the test dataset. Table 4.1 shows sensitivity, specificity, precision, recall and F1-score achieved by the VGG16 network on the test dataset.

Table 4.1 Classification Results for VGG16 Network (Prajakta Rathod et al. 2024)

	Non-seizure	**Seizure**
Sensitivity	100.00%	99.77%
Specificity	99.77%	100.00%
Precision	100.00%	100.00%
Recall	100.00%	100.00%
F1-Score	100.00%	100.00%

Fig. 4.5 shows the confusion matrix for the VGG16 network. It shows that the VGG16 network correctly classified all seizure class samples. The network misclassified a few samples from the non-seizure class into the seizure class.

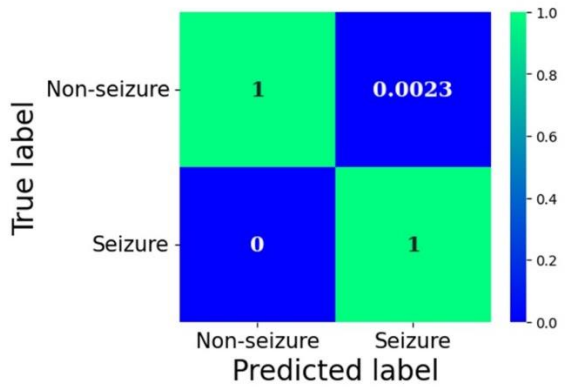


Fig. 4.5 Confusion matrix generated by the VGG16 network for the classification of epileptic seizure and non-seizure class images (Prajakta Rathod et al. 2024)

Fig. 4.6 and Fig. 4.7 show results generated by the LRP method for epileptic seizure and non-seizure signal images, respectively. Fig. 3.2b and Fig. 3.3b show that the lower portion of the spectrogram image shows a major difference between a seizure and a non-seizure class image. The upper portion of the spectrogram image for both classes is similar. Therefore, pink marks generated by the LRP method mostly lie in the lower portion, which is the lower frequency region of the spectrogram image.

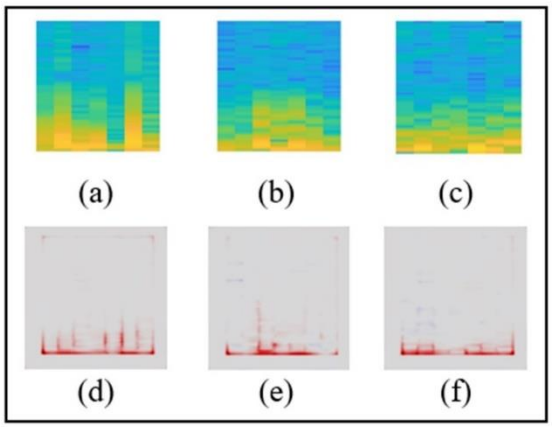


Fig. 4.6 LRP results for epileptic seizure class images. Fig. a, b and c show seizure signal images, and Fig. d, e and f show LRP results for images a, b and c, respectively (Prajakta Rathod et al. 2024)

Fig. 4.6a, 4.6b and 4.6c present three sample images from the seizure class, while Fig. 4.6d, 4.6e, and 4.6f display their corresponding LRP results. Fig. 3.2 and Fig. 3.3 show that the amplitude of the epileptic seizure signal is greater than the non-epileptic seizure signals. Hence, the lower frequency region of the spectrogram image for the seizure class signal has a more yellow colour. This yellow colour, in large amounts, contributes positively to classifying the image in the seizure class. Hence, it is marked in pink by the LRP method.

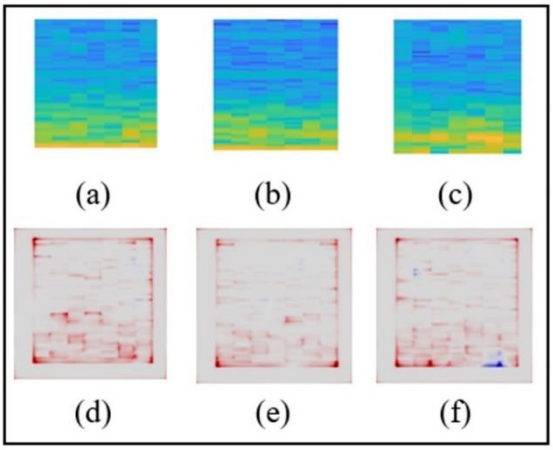


Fig. 4.7 LRP results for non-seizure class images. Fig. a, b and c show non-seizure signal images, and Fig. d, e and f show LRP results for images a, b and c, respectively (Prajakta Rathod et al. 2024)

Fig. 4.7a, 4.7b and 4.7c present three sample images from the non-seizure class, while Fig. 4.7d, 4.7e, and 4.7f display their corresponding LRP results. The non-seizure signal spectrogram image has less yellow colour in the lower frequency region, as the non-seizure signal has a lower amplitude than the seizure signal. This yellow colour in a smaller amount contributes positively to classifying the image in the non-seizure class. Hence, it is marked in pink by the LRP method.

In spectrogram images of seizure and non-seizure classes, axis labels and colorbar are made invisible for the proper visual appearance of images. Therefore, these spectrogram images appear smaller in size than the LRP result images.

4.3 Comparison of the XAI methods SHAP, Guided Backpropagation and LIME

This research study classifies EEG signal images as seizure and non-seizure using the VGG16 network. The Bonn University dataset was used to perform experiments and achieved an accuracy of 98.98%. Table 4.2 summarises the

classification performance of the VGG16 network. Moreover, three XAI methods - SHAP, guided backpropagation and LIME- are used to explain the classification results of the network, and their explanation results are also compared.

Table 4.2 Classification Results for VGG16 Network (Prajakta Rathod et al. 2025b)

	Non-seizure	**Seizure**
Sensitivity	97.9545%	100.00%
Specificity	100.00%	97.9545%
Precision	98.00%	100.00%
Recall	100.00%	98.00%
F1-Score	99.00%	99.00%

Fig. 4.8 and Fig. 4.9 display the results produced by three XAI methods - SHAP, Guided Backpropagation, and LIME - to explain the classification outcomes of the VGG16 model. Fig. 4.8 presents the results for the seizure class, while Fig. 4.9 shows the results for the non-seizure class. Each figure includes five sample images from the respective class. The seizure signal has a higher amplitude than non-seizure signals, as shown in Fig. 3.6 and Fig. 3.7. Therefore, the spectrogram of the seizure signal has a more yellow colour than the spectrogram of the non-seizure signal. The VGG16 model used this difference in yellow colour in the spectrogram image to classify seizure and non-seizure signals. Moreover, XAI methods produced explanations for the classification results by marking this yellow colour region.

The lower yellow colour regions in the spectrogram images are more significant for the classification decision in the case of the seizure class images. Therefore, the SHAP and guided backpropagation methods marked this region. However, the SHAP method produced results with a better visual appearance than the guided backpropagation method. The LIME method marked more significant regions in the

lower portion of the image, but it marked less significant regions as well in the upper portion of the image.

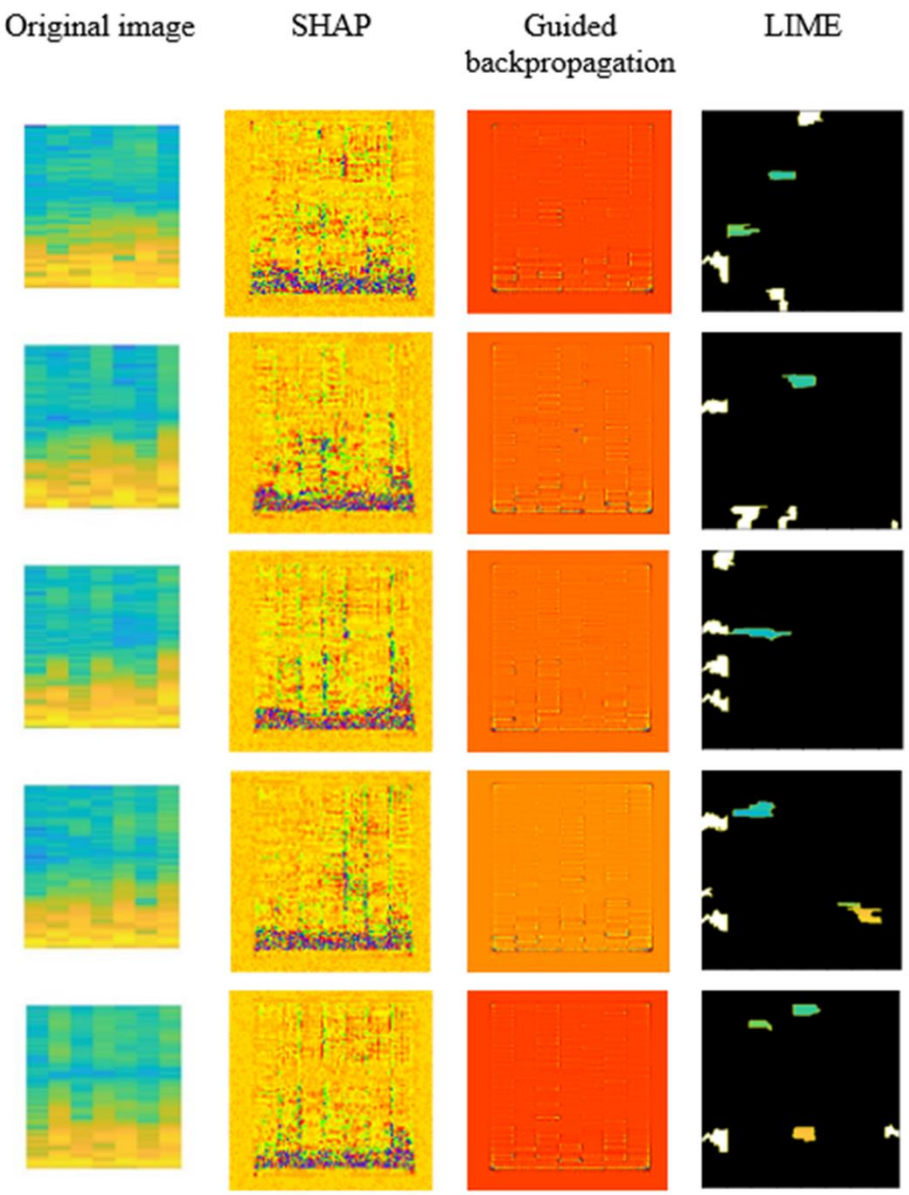


Fig. 4.8 Explanation of classification results produced by three XAI methods for seizure class images. Five sample images from the seizure class are shown. XAI methods used are SHAP, Guided Backpropagation and LIME (Prajakta Rathod et al. 2025b)

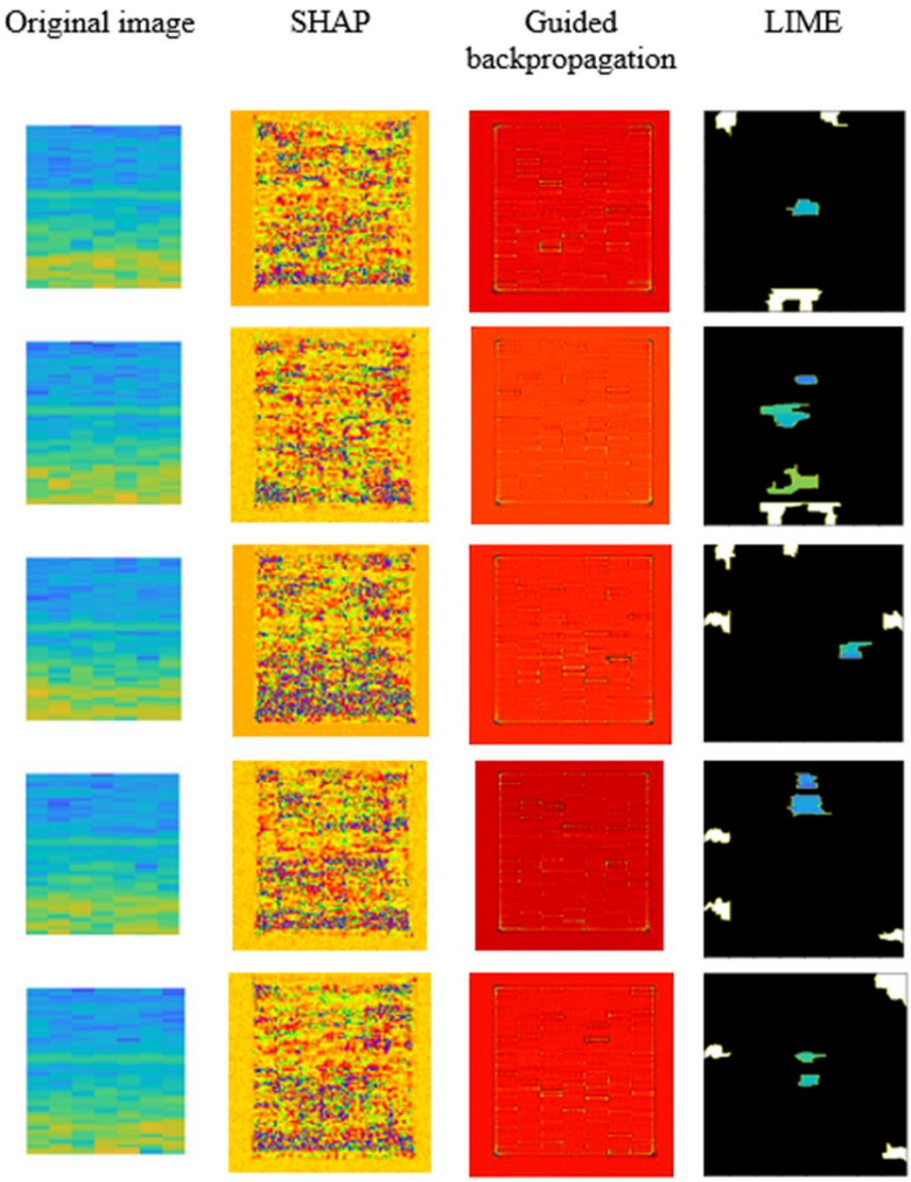


Fig. 4.9 Explanation of classification results produced by three XAI methods for non-seizure class images. Five sample images from the non-seizure class are shown. XAI methods used are SHAP, Guided Backpropagation and LIME (Prajakta Rathod et al. 2025b)

For the non-seizure class images, the SHAP and guided backpropagation methods marked the significant features mainly in the lower portion of the spectrogram images. These methods have also marked some features in the upper portion of the images. Moreover, the SHAP method results have a better visual appearance than the

guided backpropagation method in the case of non-seizure class images as well. The LIME method marked significant regions in the lower portion of the spectrogram image, but it shows some less significant regions in the upper portion of the spectrogram image as well.

Thus, it shows that the XAI method SHAP produced better results than the guided backpropagation and LIME methods for both seizure and non-seizure classes.

4.4 SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0

This method compares the performance of four CNN models for EEG signal classification and explanation. Classification results of four CNN models - Resnet18, Densenet121, VGG16 and Squeezenet1_0 – in distinguishing interictal, non-seizure, or epileptic seizure EEG signals are discussed in this section. Moreover, this section describes the results generated by the SHAP method when used with these four networks.

Table 4.3 shows the 5-fold cross-validation results achieved by Resnet18, Densenet121, VGG16, and Squeezenet1_0 on the validation dataset. The accuracy achieved by these four CNN models on the test dataset is shown in Table 4.4. It also shows the computation time required on the test dataset for these four CNN models.

Table 4.3 5-Fold Cross-Validation Results of Four Pre-Trained CNN Models Using Validation Dataset (Prajakta Rathod et al. 2025a)

Network	**Average Accuracy**
Resnet18	99.3405 ± 0.1764
Densenet121	99.3125 ± 0.3519
VGG16	99.2115 ± 0.1724
Squeezenet1_0	95.852 ± 0.8781

According to Table 4.3 and Table 4.4, Resnet18 outperforms the other models by achieving the highest accuracy of 99.7% on the test dataset. Also, the computation time taken by Resnet18 is the least, i.e., 8 sec. on the test dataset. The accuracy achieved by Densenet121, VGG16 and Squeezenet1_0 on the test dataset is 99.62%, 99.47% and 99.17%, respectively. Computation time taken by Densenet121, VGG16 and Squeezenet1_0 on the test dataset is 11 sec., 12 sec. and 8 sec., respectively.

Table 4.4 Results of Four CNN Models on Test Dataset (Prajakta Rathod et al. 2025a)

Network	**Accuracy on Test Dataset**	**Time Required for Test Dataset**
Resnet18	99.7%	8 sec
Densenet121	99.62%	11 sec.
VGG16	99.47%	12 sec
Squeezenet1_0	99.17%	8 sec.

Table 4.5 shows the sensitivity, specificity, precision, recall and F1-score achieved by four CNN models during the classification of three classes. From Table 4.5, it can be seen that, for the interictal class, Resnet18, Densenet121 and VGG16 achieved 100% precision, recall and F1-score. A 100% precision value is achieved by

Table 4.5 Classification Results of Four Pre-Trained CNN Models (Prajakta Rathod et al. 2025a)

Model	**Class**	**Sensitivity**	**Specificity**	**Precision**	**Recall**	**F1-Score**
Resnet18	Interictal	100.00%	100.00%	100.00%	100.00%	100.00%
	Non-Seizure	99.54%	100.00%	99.00%	100.00%	100.00%
	Seizure	100.00%	99.09%	100.00%	99.00%	100.00%
Densenet121	Interictal	100.00%	100.00%	100.00%	100.00%	100.00%
	Non-Seizure	99.43%	100.00%	99.00%	100.00%	99.00%
	Seizure	100.00%	98.86%	100.00%	99.00%	99.00%
VGG16	Interictal	100.00%	100.00%	100.00%	100.00%	100.00%
	Non-Seizure	99.20%	100.00%	98.00%	100.00%	99.00%
	Seizure	100.00%	98.40%	100.00%	98.00%	99.00%
Squeezenet1_0	Interictal	99.31%	100.00%	99.00%	100.00%	99.00%
	Non-Seizure	99.43%	100.00%	99.00%	100.00%	99.00%
	Seizure	100.00%	97.50%	100.00%	97.00%	99.00%

all four networks for the seizure class. For interictal and non-seizure classes, 100% recall value is achieved by all four networks. For all three classes, Squeezenet1_0 achieved a 99% F1-score. For the interictal and seizure class, the best sensitivity is achieved by Resnet18, Densenet121 and VGG16. For interictal and non-seizure classes, the best specificity is achieved by Resnet18, Densenet121 and VGG16. For the seizure class, Squeezenet1_0 achieved the best sensitivity, and for the interictal and non-seizure classes, it achieved the best specificity.

Fig. 4.10 and Fig. 4.11 show the sensitivity and specificity, respectively, achieved for three classes by all four networks. Fig. 4.12 shows the accuracy achieved on the test dataset by all four networks. Fig. 4.13 shows the confusion matrix generated by all four networks for the classification of three classes. The confusion matrix shows that all samples from interictal and non-seizure classes are correctly classified by all four networks. All four networks misclassified a few samples from the seizure class into the interictal and non-seizure classes.

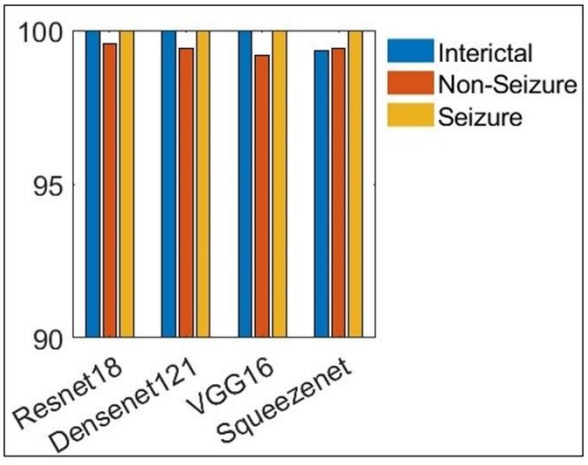


Fig. 4.10 Sensitivity graph for four CNN models. The X-axis represents four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0 and three classes: interictal, non-seizure and seizure. The Y-axis represents the sensitivity value (Prajakta Rathod et al. 2025a)

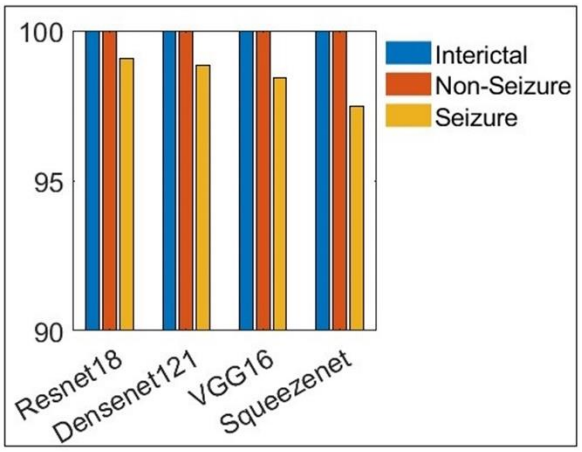


Fig. 4.11 Specificity graph for four CNN models. The X-axis represents four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0 and three classes: interictal, non-seizure and seizure. The Y-axis represents the specificity value (Prajakta Rathod et al. 2025a)

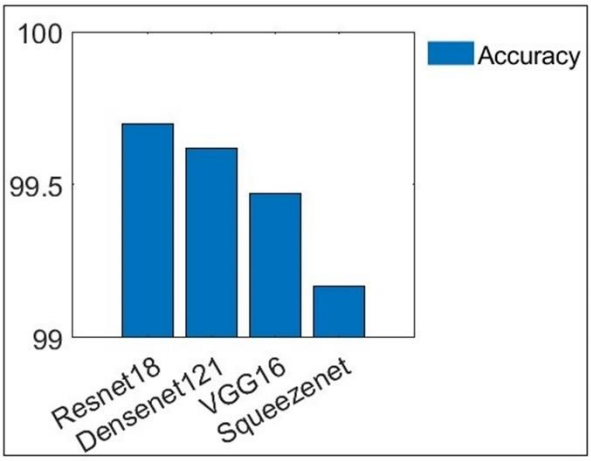


Fig. 4.12 Accuracy graph for four CNN models. The X-axis represents four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0. The Y-axis represents the accuracy value (Prajakta Rathod et al. 2025a)

Results generated by the explainable AI method SHAP for all four networks are shown in Fig. 4.14, in which two sample images from each class are shown. When a classifier classifies a spectrogram image in a particular class, some features in the image contribute positively, and some features in the image contribute negatively. The SHAP method marks positively contributing features in pink colour and negatively contributing features in blue colour. Fig. 4.14 demonstrates that when the SHAP method is used with VGG16 and Squeezenet1_0 networks, it produces good results. When the

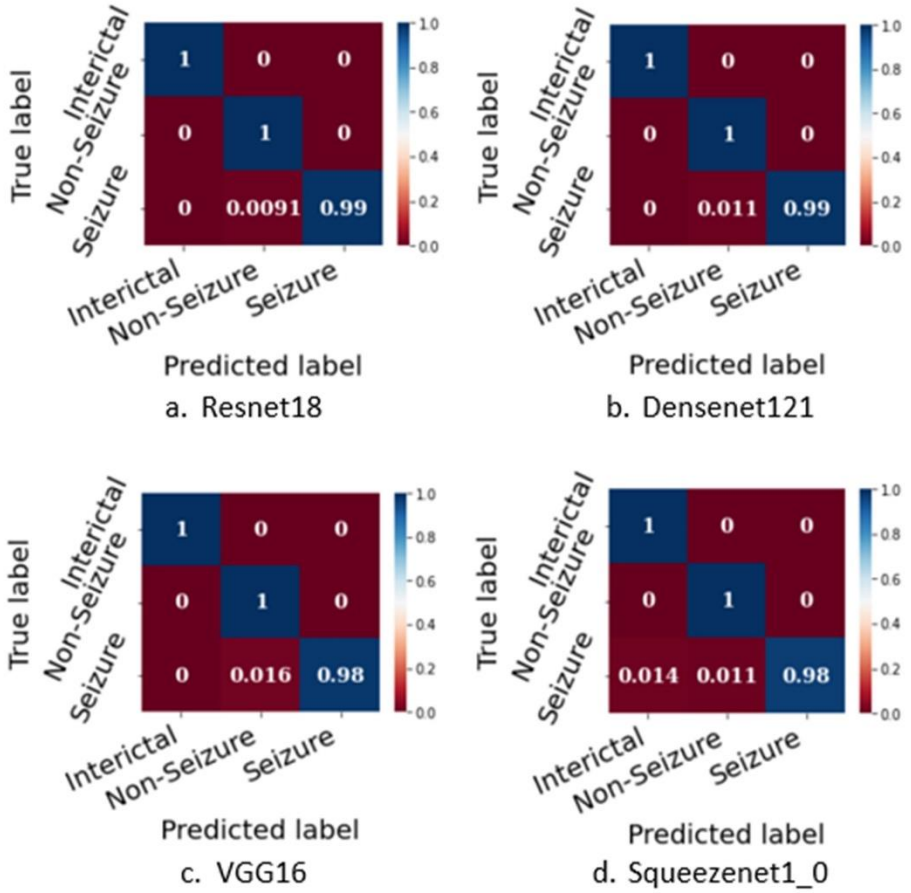


Fig. 4.13 Confusion matrix for four CNN models, Resnet18, Densnet121, VGG16 and Squeezenet1_0 obtained for three classes: interictal, non-seizure and seizure (Prajakta Rathod et al. 2025a)

SHAP method is used with Densenet121 and Resnet18 networks, the results produced are visually poor. Moreover, Fig. 4.14 shows that the SHAP method generated better results for the epileptic seizure class than the non-seizure and interictal classes. The SHAP method produced pink marks for the features in the image, due to which the image is classified in a particular class.

In this research, Resnet18, Densenet121 and VGG16 achieved an accuracy of 99.7%, 99.62% and 99.47%, respectively. In this method, the accuracy achieved by Resnet18, Densenet121 and VGG16 on the Bonn University dataset is higher than the accuracy achieved by other systems on the same dataset.

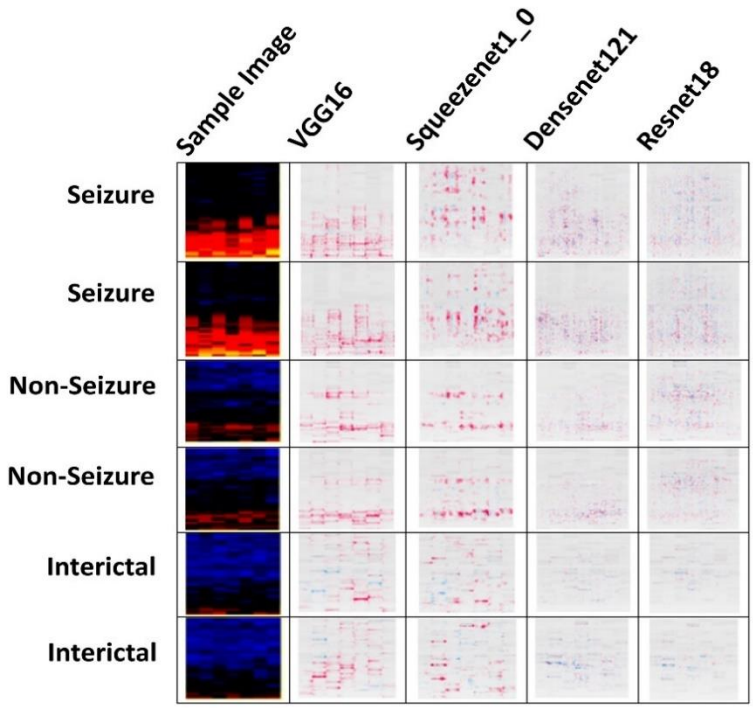


Fig. 4.14 SHAP method results for four CNN models: Resnet18, Densnet121, VGG16 and Squeezenet1_0. Two sample images from each of the three classes, interictal, non-seizure and seizure, are shown. The SHAP method marks positively contributing features in pink shade and negatively contributing features in blue shade for classification in a particular class (Prajakta Rathod et al. 2025a)

4.5 SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channels

Three CNN models are used to classify EEG signals as epileptic and non-epileptic. The weights of the three CNN models are updated with the EEG signal image dataset, and their classification outcomes are interpreted using the XAI method SHAP. This section gives the CNN models' comparative results.

Table 4.6 summarises the CNN models' performance based on 5-fold cross-validation using the validation dataset. It also shows the accuracy obtained on the test dataset by the same three networks. Table 4.6 shows that VGG16 attains the highest accuracy of 97.55% on the test dataset. Moreover, the VGG16 model achieved the highest accuracy of 97.1 ± 0.2522 with k=5, on the validation dataset. Therefore,

VGG16 is again evaluated on the validation dataset with k=10 and obtained an accuracy of 97.57 ± 0.1580. VGG16 model's k-fold cross-validation results on the validation dataset with k=5 and k=10 are shown in Table 4.7.

Table 4.6 Performance of Three CNN Models on the Validation Dataset Using 5-Fold Cross-Validation and on the Test Dataset (Prajakta Rathod et al. 2025)

Network	**Average Accuracy on Validation Dataset**	**Accuracy on Test Dataset**
Densenet121	89.441 ± 0.3162	91.29%
Resnet18	94.005 ± 0.2151	95.0%
VGG16	97.1 ± 0.2522	97.55%

Table 4.7 K-Fold Cross-Validation Results of VGG16 Network on the Validation Dataset for k=5 and k=10 (Prajakta Rathod et al. 2025)

Network	**k value**	**Average Accuracy on Validation Dataset**
VGG16	k=5	97.1 ± 0.2522
VGG16	k=10	97.57 ± 0.1580

Table 4.8 summarises the classification performance of the same three networks for identifying epileptic and non-epileptic signal images. VGG16 yields the highest precision, recall, and F1-score for both classes, whereas Densenet121 shows the lowest performance on all three metrics.

Table 4.8 Classification Results of Three CNN Models, Densenet121, Resnet18 and VGG16, for Epileptic and Non-epileptic Signal Images (Prajakta Rathod et al. 2025)

Network	**Class**	**Sensitivity**	**Specificity**	**Precision**	**Recall**	**F1-Score**
Densenet121	Epileptic	95.6%	86.9%	95.0%	87.0%	91.0%
	Non-epileptic	86.9%	95.6%	88.0%	96.0%	92.0%
Resnet18	Epileptic	94.4%	95.6%	94.0%	96.0%	95.0%
	Non-epileptic	95.6%	94.4%	96.0%	94.0%	95.0%
VGG16	Epileptic	97.6%	97.5%	98.0%	97.0%	98.0%
	Non-epileptic	97.5%	97.6%	98.0%	98.0%	98.0%

Fig. 4.15a, 4.15b and 4.15c show a graphical representation of accuracy, sensitivity and specificity, respectively, obtained by the same three networks on the test dataset for classifying epileptic and non-epileptic signal images.

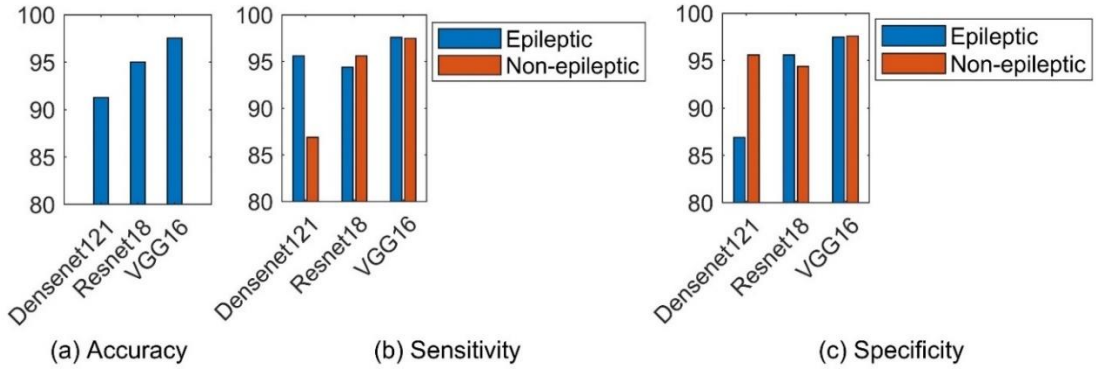


Fig. 4.15 a, b and c show the accuracy, sensitivity and specificity graphs, respectively, for Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025)

Fig. 4.16 shows the confusion matrix for all three networks in classifying epileptic and non-epileptic signal images. VGG16 has correctly classified more samples, whereas Densenet121 has misclassified more samples.

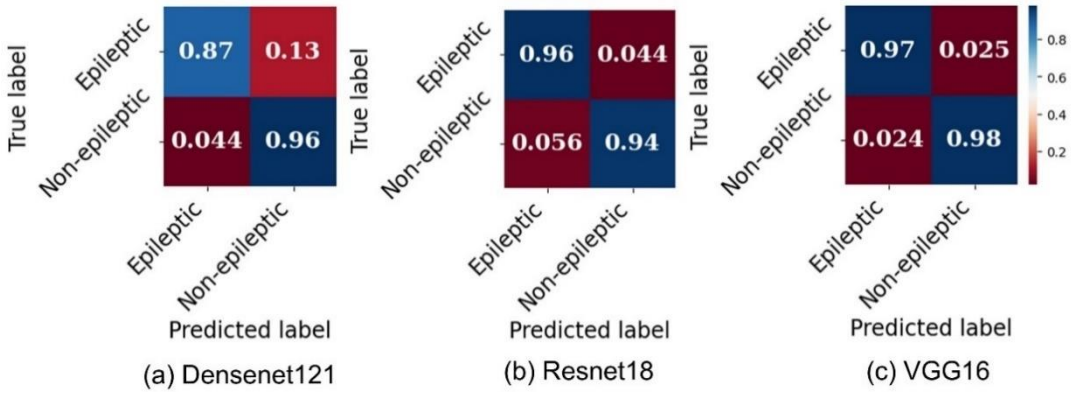


Fig. 4.16 Confusion matrix for three CNN models, Densenet121, Resnet18 and VGG16, obtained for two classes, epileptic and non-epileptic. (Prajakta Rathod et al. 2025)

The results of the XAI method SHAP for epileptic and non-epileptic classes are shown in Fig. 4.17 and Fig. 4.18, respectively, for all three networks. Some features in the image contribute positively, and some contribute negatively to the classification. Positively contributing features are marked in pink colour, and negatively contributing features are marked in blue colour by the SHAP method. Fig. 4.17 and Fig. 4.18 show five sample images from each class.

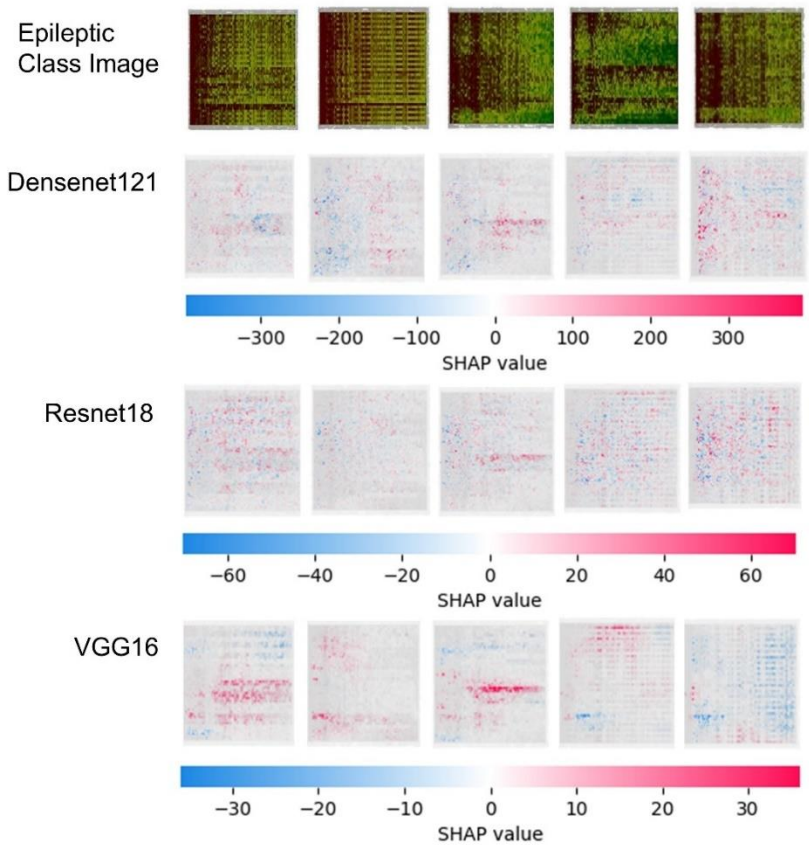


Fig. 4.17 SHAP method results for epileptic signal images using three CNN models: Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025)

SHAP method results are quantified using AOPC values. Section 3.2.5.3.2 describes heatmap evaluation using AOPC values. AOPC curves obtained for all three networks are shown in Fig. 4.19. AOPC values obtained for all three networks are

shown in Table 4.9. Densenet121 achieved the largest AOPC value, which means that Densenet21 has produced the best explainability for the XAI method SHAP. Resnet18 has a larger AOPC value than VGG16, which indicates that Resnet18 produced better explainability than VGG16 for epileptic and non-epileptic image classification using the XAI method SHAP.

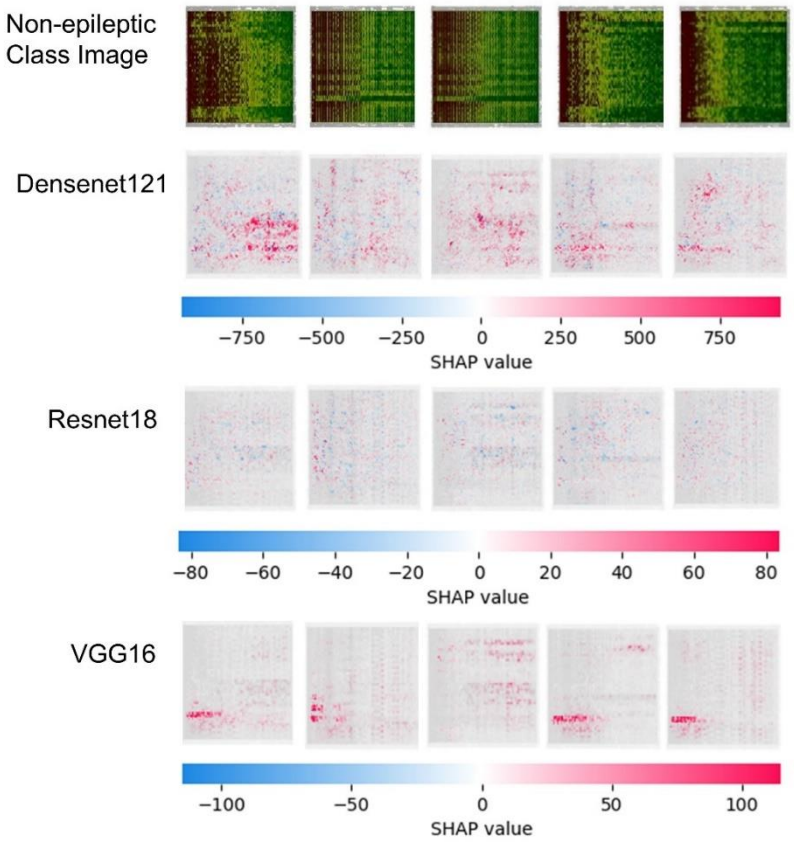


Fig. 4.18 SHAP method results for non-epileptic signal images using three CNN models: Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025)

Densenet121 has produced the best explainability, as shown in Fig. 4.19 and Table 4.9. Therefore, it is used to identify the top contributing EEG channels in the classification. The key EEG channels that contributed most to identifying all epileptic and non-epileptic patients in the test set are shown in Table 4.10.

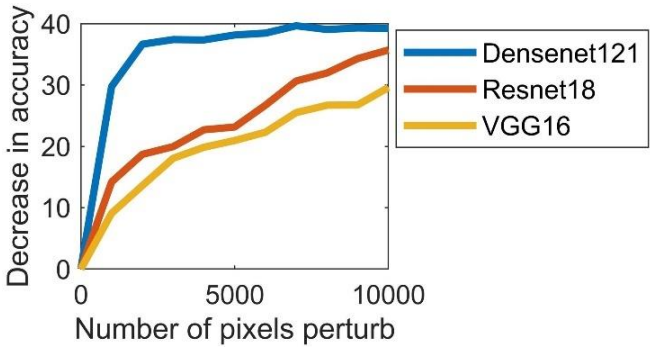


Fig. 4.19 AOPC curves obtained for all three networks: Densenet121, Resnet18 and VGG16 (Prajakta Rathod et al. 2025)

Table 4.9 Evaluation of the Heatmap of Three CNN Models Using AOPC Value (Prajakta Rathod et al. 2025)

Network	**AOPC value**
Densenet121	355350
Resnet18	240150
VGG16	197600

Table 4.10 Top Contributing EEG Channels in Epileptic and Non-epileptic Patients' Classification (Prajakta Rathod et al. 2025)

Epileptic		**Non-epileptic**	
Patient1	F8, T5, T4, F7, O2, T3, P4, P3, A1, T6	Patient1	C4, C3, P3, O2, T4, F7, P4, O1, F8, T3
Patient2	O2, P4, F7, P3, F8, T4, T5, O1, C4, T3	Patient2	O2, F7, F4, P3, T4, F8, C4, T3, T5, P4
Patient3	T4, F8, T3, T5, O2, P4, F7, A1, A2, T6	Patient3	C3, F7, F8, P3, P4, F4, O2, T4, T3, O1
Patient4	F8, T3, F7, O2, T4, P3, P4, O1, C3, C4	Patient4	F8, T3, F7, T4, C4, C3, O2, P3, T5, P4
Patient5	T4, T5, F8, A1, F7, P3, C4, P4, T3, A2	Patient5	T4, F4 F8, C3, F7, P4, C4, T5, T3, P3
Patient6	T4, O2, T3, T5, F7, F8, P4, P3, T6, O1	Patient6	O2, F7, C3, C4, T4, T3, P4, P3, F8, T5
Patient7	T4, T5, F8, A1, F7, A2, FZ, T6, T3, O2	Patient7	F4, C4, P3, C3, F3, F7, O2, F8, A1, P4
Patient8	O2, F7, P3, F8, O1, P4, T5, T4, T3, C4	Patient8	T5, P4, P3, T4, O2, A1, T6, F7, C4, F8
Patient9	P4, F7, O2, P3, O1, C3, C4, F8, T4, F4	Patient9	P4, T5, T4, F7, P3, O1, O2, F8, A1, C4
Patient10	O2, F7, P4, T4, F8, O1, T5, P3, T3, A1	Patient10	O2, F8, F7, T3, T4, T5, A1, A2, T6, O1
Patient11	T4, T5, F7, P4, F8, A1, C4, O2, T3, T6	Patient11	C3, C4, P4, P3, O1, O2, F7, F4, F8, T4
Patient12	O2, F8, F7, P4, O1, T4, T5, P3, T3, A1	Patient12	T5, T4, O2, F8, A1, F7, T6, T3, O1, A2
Patient13	C3, F4, P3, P4, C4, F7, F8, O2, T4, T5	Patient13	T4, T5, P3, P4, C4, F8, O2, O1, T6, T3
Patient14	C3, F4, C4, F3, P3, P4, F7, F8, T4, O2	Patient14	C3, F4 C4, P3, P4, F8, O2, F7, O1, T4
Patient15	F8, T4, F7, O2, T5, P4, A1, T3, T6, O1	Patient15	C3, C4, F4, F3, P4, P3, O1, T5, T4, F8
Patient16	P3, P4, O2, T4, F8, F7, O1, T3, T5, C4	Patient16	T4, F8, T3, T5, F7, A1, T6, A2, O2, FZ
Patient17	T5, A1, T4, A2, T6, FZ, F8, T3, F7, CZ	Patient17	T4, O2, F4, F7, F8, O1, T5, P4, P3, T3
Patient18	F7, T4, F8, T5, P3, P4, F4, O2, T3, C4	Patient18	C4, P3, F4, C3, F3, P4, T4, F7, F8, O2
Patient19	P3, P4, O2, O1, F7, C4, F8, T4, F4, T3	Patient19	C3, F4, C4, P3, P4, O2, O1, F7, F8, T5
Patient20	T4, O2, F7, T5, F8, O1, A1, T6, P4, T3	Patient20	F4, F3, C3, C4, P3, O2, O1, P4, F7, F8
Patient21	F7, T4, F8, O2, T3, P4, P3, T5, O1, A1	Patient21	F8, F7, T4, O2, T3, T5, FZ, A1, A2, T6
Patient22	T5, T4, F8, A1, T6, F7, T3, A2, O2, FZ	Patient22	F4, F3, C3, P3, T4, P4, F8, C4, A1, T5
Patient23	C3, P3, F4, C4, O2, P4, F7, O1, F8, F3	Patient23	C3, C4, O2, P3, F7, P4, F8, O1, T4, T3
Patient24	T4, F7, F8, T5, O2, T3, A1, T6, P4, A2	Patient24	P4, F7, O2, P3, F8, T4, O1, T5, T3, C4
Patient25	F8, T4, F7, O2, T3, P4, O1, T5, P3, A1	Patient25	P3, P4, O1, O2, A1, T5, T6, T4, F8, C3

Patient26	F8, F7, F4, O2, P3, T4, C3, P4, T5, T3	Patient26	O2, P4, T4, F8, F7, O1, P3, T5, C3, C4
Patient27	C4, P3, F4, C3, F3, P4, F7, O2, F8, O1	Patient27	F8, F7, P3, O2, T3, T4, C3, P4, T5, O1
Patient28	C4, A1, O2, P3, P4, F8, F7, C3, T4, A2	Patient28	P3, C3, C4, F4, P4, O2, O1, F7, F3, F8
Patient29	F8, F7, O2, P3, C4, T4, P4, T3, T5, O1	Patient29	P3, F8, C3, C4, T4, F7, T3, T5, P4, O2
Patient30	F7, F8, O2, P4, T4, T5, O1, P3, T3, C3	Patient30	C4, P3, O2, F7, P4, O1, C3, F8, F4, T4
Patient31	T4, O2, T5, F7, F8, A1, A2, T6, T3, P4	Patient31	C3, P4, O2, O1, F7, C4, P3, F8, T3, T4
Patient32	P3, P4, T4, T5, C4, T3, C3, A1, F4, F8	Patient32	F4, C3, P3, F3, P4, C4, O1, T4, O2, F7
Patient33	T4, O1, P4, T5, P3, F8, F7, O2, A2, A1	Patient33	C3, P3, C4, P4, T4, T5, F4 O1, F3, O2
Patient34	P3, T5, T4, A1, F7, P4, F8, C4, A2, FZ	Patient34	F4, C3, F3, C4, F7, P3, O2, F8, P4, A2
Patient35	T4, F8, T3, T5, A1, F7, A2, O2, T6, P3	Patient35	P3, F4, T5, C3, P4, C4, A1, T4, F8, T6
Patient36	T4, F8, T5, F7, A1, T6, O2, T3, P3, P4	Patient36	T4, F8, O2, A1, T5, F7, P4, P3, T3, A2
Patient37	O2, F7, F8, P4, T4, F4, O1, T5, T3, P3	Patient37	C3, C4, F4, F8, P3, F3, F7, T4, T3, P4
Patient38	P4, O2, O1, T4, P3, T5, F8, F7, C4, A1	Patient38	T4, T5, P3, F7, C3, F8, P4, T3, F4 O2
Patient39	T5, A1, T4, F8, A2, F7, T6, O2, T3, P4	Patient39	F4, C3, F3, C4, P3, P4, F8, F7, O2, T4
Patient40	A1, T4, T5, A2, T6, F7, F8, FZ, O2, T3	Patient40	F4, F3, A2, FZ, PZ, A1, T4, CZ, T5, T6
Patient41	T4, F8, T5, F7, A1, T3, T6, A2, O2, FZ	Patient41	T4, F8, F7, O2, T5, T3, A1, T6, P4, A2
Patient42	T4, T5, A1, P4, O2, F8, F7, A2, C3, P3	Patient42	C3, T4, O2, P4, P3, F7, F8, T5, C4, T3
Patient43	F7, C3, O2, F8, P4, T4, T5, F4, P3, O1	Patient43	T4, F8, T5, T3, C3, F7, O2, T6, A1, P4
Patient44	T4, T5, F8, O2, F7, T3, O1, T6, P4, C4	Patient44	T5, T6, A1, T4, A2, T3, C3, F8, FZ, C4
Patient45	P4, P3, F8, T4, T5, F7, O2, C4, O1, T6	Patient45	P3, C3, C4, T4, T5, F4, P4, F8, T3, O2
Patient46	F7, T4, T5, O2, P4, A1, O1, F8, T6, P3	Patient46	P4, P3, T5, C3, T4, F7, A1, C4, T6, O2
Patient47	P4, F8, P3, O1, O2, T4, F7, C3, T3, C4	Patient47	P3, C3, C4, P4, F4 O1, O2, F7, T4, F8
Patient48	F8, F4, T4, T5, C3, F7, T3, O2, A1, T6	Patient48	C4, C3, P3, P4, O2, F7, F8, F4, F3, T4
Patient49	O2, O1, F7, F8, T4, P3, P4, T5, T3, A1	Patient49	P3, P4, C4, O2, O1, F7, T4, F8, T5, A1
Patient50	P4, F3, F4, P3, F7, C3, O1, F8, O2, T4	Patient50	P3, C4, C3, O1, T5, P4, A1, O2, T4, T6

Summary:

This chapter describes the results generated by the five methods discussed in Chapter 3. In the first method, the LRP approach is applied to the Bi-LSTM network, where the method highlights significant parts of the signal that positively contribute to classification in red. The Bi-LSTM network achieved 87.25% classification accuracy on the Bonn University dataset. In the second method, the LRP method is implemented in the VGG16 network, which achieved a remarkably high classification accuracy of 99.89% on the same dataset. The third method focuses on comparing three XAI methods - SHAP, Guided Backpropagation, and LIME - within the VGG16 network, with SHAP demonstrating the best explainability among them. The fourth method involves a comparison of four CNN models - Resnet18, Densenet121, VGG16 and

Squeezenet1_0, where Resnet18 outperformed the others by achieving 99.7% accuracy on the test dataset and 99.34% accuracy on the validation dataset. However, SHAP provided superior explainability with VGG16 and Squeezenet1_0 compared to Densenet121 and Resnet18, particularly for the seizure class over the non-seizure and interictal classes. While the experiments corresponding to the first four methods were conducted on the single-channel Bonn University dataset, the fifth method is applied to the multichannel TUH EEG dataset. In this method, a comparative evaluation of three CNN models - VGG16, Resnet18, and Densenet121 - reveals that VGG16 achieved the best performance with 97.55% accuracy on the test dataset and 97.1% ± 0.2522 on the validation dataset. Furthermore, Densenet121 achieved better results with the XAI method SHAP than Resnet18 and VGG16, and was therefore used with SHAP to identify the top contributing EEG channels that distinguish epileptic and non-epileptic patients.

5. Evaluation of Explainable Epilepsy Detection and Clinical Implications

5.1 Justification for the datasets

5.1.1 Bonn University dataset

This research study uses a single-channel Bonn University EEG dataset illustrated in Chapter 3, Section 3.1.1. This dataset has a limited number of EEG data samples. Nevertheless, numerous methods in the literature have utilised this dataset for EEG signal classification and seizure identification. DNNs have been utilised by some of these methods for detecting seizures. The models in this research are trained using a transfer learning approach with fine-tuning, which allows effective learning from fewer training samples. This study uses the VGG16, Resnet18, Densenet121 and Squeezenet1_0 models, which are already trained on a large dataset and learned to identify curves, edges and colours in the images. Hence, the Bonn University dataset is used to conduct experiments in this research.

5.1.2 TUH dataset

This research uses the TUEP subset from the TUH dataset, which is described in Chapter 3, Section 3.1.2. The TUH dataset has been used by many methods in the literature. This dataset has many subsets that can be used to solve different epilepsy

research problems. Each of these subsets has a large number of EEG data files. Hence, we used the TUH dataset to achieve better accuracy and enable the comparison of our results with literature methods.

5.2 EEG patterns during epileptic episodes

Brain activity is commonly described in terms of distinct frequency bands; each associated with specific states and functions. Delta waves (0.5 - 4 Hz) dominate during deep non-Rapid Eye Movement (REM) sleep and anaesthesia. Theta waves (4 - 8 Hz) are prominent in the hippocampus during movement and REM sleep, as well as in the frontal midline during attention and working-memory tasks. Alpha rhythms (8 - 12 Hz), especially visible over posterior regions when the eyes are closed, are linked to relaxed wakefulness, visual idling, and inhibitory control of attention, while the sensorimotor mu rhythm in the same frequency range supports motor-related processing. Beta oscillations (12 - 30 Hz) are expressed in the sensorimotor cortex during wakefulness and motor planning. Gamma activity (30 - 90 Hz), emerges during sensory processing, attention, and working-memory operations, supporting communication across brain regions. In addition to these canonical bands, sleep spindles (around 12 - 15 Hz) appear during stage 2 non-REM sleep and ripples (80 - 250 Hz) observed in the hippocampus during quiet wakefulness and non-REM sleep, represent brief bursts of high-frequency activity that contribute to memory replay and consolidation. Together, these rhythms highlight how different oscillatory frequencies serve complementary roles: slower bands coordinate large-scale state transitions, whereas faster oscillations enable rapid local computations and effective communication between neuronal populations (Roohi-Azizi et al. 2017).

During an epileptic seizure, EEG signals undergo marked deviations from normal background activity. The most prominent change is the emergence of paroxysmal discharges, which appear as sudden bursts of high-amplitude spikes, sharp waves, or spike-and-wave complexes that are not present in a healthy EEG. In the time domain, the waves become more synchronous, often displaying rhythmic, repetitive patterns that evolve dynamically throughout the seizure. In the frequency domain, seizures are usually characterised by a shift toward lower-frequency oscillations such as theta and delta during generalised seizures, whereas focal seizures may exhibit bursts of higher-frequency activity, including beta and gamma oscillations. As the seizure progresses, the EEG typically shows an increase in amplitude and rhythmicity, followed by a gradual slowing down and flattening (post-ictal suppression) once the seizure terminates (Li et al. 2020). These signal characteristics form the basis for modern machine-learning and deep-learning approaches to automated seizure detection. These dynamic changes in amplitude, frequency, and synchrony in an EEG can be used for seizure detection and localisation of the epileptogenic region (Shoeb and Guttag 2010).

5.3 Justification for selecting EEG signal segments of 2 sec. and 4 sec. duration

In this study, EEG signals from the Bonn dataset were divided into 2 sec. segments with an overlap of 1 sec. The 2 sec. window was chosen because it provides a fine temporal resolution, enabling the detection of short-duration epileptic transients and abrupt changes in brain wave patterns. The overlap ensures continuity between segments and prevents loss of critical information occurring at the window boundaries. Since the Bonn dataset consists of single-channel, noise-free recordings with relatively

short signal durations, a shorter window length is sufficient to preserve meaningful temporal and frequency information without redundancy. In contrast, EEG recordings from the TUH dataset were divided into 4 sec. segments. The longer window size was selected to account for the greater variability, complexity, and multi-channel nature of clinical EEG recordings. A 4 sec. duration allows for capturing the full evolution of epileptic discharges, which often span several seconds in real-world data, while maintaining an optimal balance between temporal resolution and spectral accuracy. This segment length also ensures that lower frequency components, such as delta and theta rhythms, are adequately represented in the spectrograms. By converting these 2 sec. and 4 sec. EEG segments into spectrogram images, both temporal and spectral characteristics are effectively preserved, facilitating robust feature learning by DNNs. Hence, the selected window lengths are empirically and theoretically justified to enhance seizure detection performance across both EEG datasets.

5.4 Discussion of the results

5.4.1 LRP in the Bi-LSTM network

The proposed method uses a Bi-LSTM network to identify epileptic seizure signals and the XAI method LRP to explain network prediction, as shown in Chapter 3, Fig. 3.1. To explain the classification, the LRP method highlights EEG signal patterns characteristic of seizure and non-seizure events. Fig. 4.1 and Fig. 4.3 display sample epileptic and non-epileptic signals, respectively. LRP method results for sample epileptic signals in Fig. 4.1 are shown in Fig. 4.2. Also, LRP method results for sample non-epileptic signals in Fig. 4.3 are shown in Fig. 4.4.

This research focuses on distinguishing between epileptic and non-epileptic EEG signals. Accordingly, data from set B and set E, as detailed in Table 3.3, are utilised. The proposed system is compared with existing approaches in Table 5.1. In (Raghu et al. 2020), pre-trained networks and SVM were used for epileptic signal classification. In (Daoud and Bayoumi 2019) and (Gao et al. 2020), a DCNN-based approach was used. In (Abdelhameed and Bayoumi 2018), epileptic signals were classified using DCAE and Bi-LSTM. In (Wang et al. 2019) (Hu et al. 2021) (Qu and Yuan 2019), epileptic signals were classified using pre-trained networks. These systems identified the epileptic signals correctly, but they lack an explanation of network prediction. This limitation of existing systems is addressed by the proposed study, in which the XAI method LRP is implemented in a Bi-LSTM network to explain the network predictions. The proposed approach provides neurologists with clear insights into why the network has determined whether a patient is epileptic or non-epileptic using EEG signals. The correctness of the technological tool used in clinical practice can be verified by using the LRP method. The ability to verify network prediction helps medical practitioners build up trust in using DL-based epileptic seizure detection systems in daily practice.

Table 5.1 Comparison with Existing Systems Using DL Methods (Prajakta Rathod et al. 2022)

Author	**Method**
(Raghu et al. 2020)	Pretrained Networks + SVM
(Daoud and Bayoumi 2019)	DCNN
(Gao et al. 2020)	DCNN
(Abdelhameed and Bayoumi 2018)	DCAE and Bi-LSTM
(Hu et al. 2021)	Pretrained Networks
(Qu and Yuan 2019)	AlexNet
(Wang et al. 2019)	Pretrained Networks
Proposed Method	Bi-LSTM+XAI method, LRP for explaining results

Despite the correct explanations produced by the system for network prediction, the system has some room for further improvement. Epileptic signals in Fig. 4.1a, 4.1b and 4.1c and their LRP results in Fig. 4.2a, 4.2b and 4.2c, respectively, show that some higher amplitude data points in the upward and downward direction are not captured by the LRP method. These data points that correspond to higher amplitude, which have a positive contribution in classifying signals in the epileptic class, are not marked by the LRP method. Again, some lower amplitude data points are marked in red by the LRP method. The data points that correspond to lower amplitude do not have a positive contribution in classifying the signal in the epileptic class and should not be marked in red colour. Similarly, non-epileptic signals in Fig. 4.3a, 4.3b and 4.3c and their LRP results in Fig. 4.4a, 4.4b and 4.4c, respectively, show that some lower amplitude data points in upward and downward directions are not captured by the LRP method. These data points that correspond to lower amplitude, which have a positive contribution in classifying signals in the non-epileptic class, are not marked by the LRP method. Again, some higher amplitude data points that do not have a positive contribution in classifying the signals in the non-epileptic class are marked in red colour by the LRP method. Therefore, the proposed system needs further improvement to generate more accurate results.

5.4.2 LRP in the VGG16 network

This research employs transfer learning with fine-tuning to train the VGG16 network on an EEG image dataset. The approach effectively decreased the training time and the amount of data needed for model training. Using this approach, the VGG16

network could achieve 99.89% accuracy on the image dataset with 4400 images belonging to epileptic seizure and non-seizure classes.

In this method, EEG signals are divided into small segments of 2 sec. duration. There is an overlap of 1 sec. duration. The dataset is divided into an 80% training set and a 20% testing set using a random selection of samples. However, there is a limitation in this approach that some samples may have overlapping patterns due to the random selection of samples, causing a higher accuracy. This limitation is resolved in subsequent methods by adopting a cluster-based approach for training and testing sample selection.

The XAI method LRP is also implemented in the VGG16 network to explain the classification results of the network. Such explanations enable neurologists to comprehend the rationale behind the network's classification of an EEG signal as an epileptic seizure or non-seizure. Thus, the neurologist can verify the correctness of the decision of the model. Thus, this method integrates the advantages of transfer learning with those of the XAI method.

The proposed method is evaluated against existing approaches using the same Bonn University dataset through a comparative analysis. This comparative analysis is shown in Table 5.2, which shows that the proposed method achieved higher classification accuracy than other existing methods on the same dataset.

The proposed approach has several limitations. A few spectrogram images did not produce proper results with the LRP method. Therefore, the implementation of the LRP method can be improved. This method is tested with a limited dataset of Bonn University, which may limit the generalizability of the method. A large dataset is essential to test the proposed approach.

Table 5.2 Comparison with Existing Systems Using the Same Dataset (Prajakta Rathod et al. 2024)

Author	**Classification Method**	**Accuracy**
(Xu et al. 2021)	CNN	95.8%
(Sharmila and Geethanjali 2016)	NB and KNN classifier	99.25%, 98.25%
(Mathur et al. 2021)	RPS	98.6%,
Proposed Method	VGG16 + XAI method LRP for explaining results	99.89%

5.4.3 Comparison of the XAI methods SHAP, Guided Backpropagation and LIME

This research uses the VGG16 network to distinguish between seizure and non-seizure EEG signals. The experiments are conducted on the Bonn University dataset. Moreover, this study compares the explanation results produced by the three XAI methods: SHAP, guided backpropagation and LIME. EEG signals are segmented into small pieces and converted into spectrogram images by applying STFT to them. The VGG16 model is trained on these images using transfer learning with a fine-tuning approach. As mentioned in Section 5.4.2, by using this approach, both training time and the necessity for a large training dataset were significantly reduced. A classification accuracy of 98.98% was achieved by the VGG16 model with 4400 training images on the Bonn University dataset with the help of the proposed preprocessing technique and a transfer learning approach.

Explanation of the classification results produced by the three XAI methods is compared, and it is found that explanations produced by the SHAP method are better than the guided backpropagation and LIME methods for both seizure and non-seizure classes. Such explanations can help neurologists and EEG technicians understand the decisions of the model. They can verify and trust the model's decisions, which helps to increase its usability.

A comparison of the proposed method and existing approaches using the same Bonn University dataset is presented in Table 5.3. Results indicate that the proposed method achieved a higher accuracy compared to existing approaches on the dataset.

Table 5.3 Comparison with Existing Systems Using the Same Dataset (Prajakta Rathod et al. 2025b)

Author	**Classification Method**	**Accuracy**
(Xu et al. 2021)	CNN	95.8%
(Abedin et al. 2019)	ANN	97.33%
(Bekbalanova et al. 2019)	DT	62.5%
(Jose et al. 2020)	ELM	98.5%
(Mathur et al. 2021)	SVM	98.6%
Proposed Method	VGG16 + Comparative analysis of XAI methods for explaining network prediction	98.98%

The results produced by the guided backpropagation and LIME methods can be further improved by experimenting with different pre-processing techniques.

5.4.4 SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0

In the proposed method, EEG signals are converted into spectrogram images using STFT. These images are input to four CNN models: Resnet18, Densenet121, VGG16, and Squeezenet1_0 for classification. The classification results of four CNN models are explained using the XAI method SHAP. Such explanations are very helpful for neurologists and EEG technologists as they can understand why the signal is classified as an interictal, non-seizure and epileptic seizure.

A major challenge with EEG data is the distribution mismatch between the source (training set) and target (testing set) domains. A transfer learning mechanism can handle such differences. A fine-tuned transfer learning approach is applied to train all four CNN models on the image dataset. As mentioned in Section 5.4.2, by using this

approach, both training time and the necessity for a large training dataset were significantly reduced. All four networks are trained from scratch on an EEG image dataset to identify three classes: interictal, non-seizure and epileptic seizure. Previous learning of all four networks is adapted to new EEG data.

Spectrogram images are normalised before using them to train the network. In Fig. 3.10, original spectrogram images and images after normalisation are shown. The original spectrogram image has three colours, yellow, green, and light blue, which are converted to red, black, and dark blue, respectively, during normalisation. It can be seen from Fig. 3.9 that the amplitude of the epileptic seizure signal is higher than the interictal and non-seizure signal. Higher amplitude is represented by a red colour in the seizure class image. This red colour has a positive contribution to classifying the image in the seizure class. The SHAP method marks positively contributing features in pink shades. Therefore, the red colour in the seizure class image is marked in pink shades by the SHAP method. In the case of a non-seizure class image, black and red colours contribute more to identifying the image as a non-seizure class image. Therefore, the SHAP method marked black and red colours in pink shades. In the case of interictal class image, the blue colour represents low amplitude and contributes positively to the classification of the image in this class. Therefore, for interictal class images, the SHAP method marked blue colour with pink shades.

EEG technologists can interpret explanations of the classification results produced on spectrogram images of EEG signals with little training and practice. EEG technologists can learn to identify higher and lower amplitudes in spectrogram images with training and can understand that the classification of a signal as epileptic or non-epileptic is influenced by its amplitude, with higher amplitudes typically indicating

epileptic activity. EEG signals are divided into 2 sec. duration segments during preprocessing. These signals can be divided into segments of 1 min. or 2 min. duration so that it can identify a small period of seizure attack.

The proposed system has been evaluated against existing systems using the same dataset. This comparison is shown in Table 5.4. As mentioned in Table 4.4, using the proposed pre-processing techniques and transfer learning approach, Resnet18 achieved 99.7% accuracy on the Bonn Dataset.

Table 5.4 Comparison with Existing Systems Using the Same Dataset in Terms of Classification Accuracy (Prajakta Rathod et al. 2025a)

Author	**Set Combination**	**Output Classes**	**Pre-Processing Technique**	**Classification Method**	**Accuracy**
(Peker et al. 2016)	AB-CD-E	Non-seizure, Interictal, Seizure	Dual-tree complex wavelet transformation	CVANN	97.79%
(Mathur et al. 2021)	B-E, D-E	Non-seizure, Interictal, Seizure	Projection of multiple nonoverlapping EEG blocks obtained from a single-channel EEG on a particular RPS	SVM	98.6%, 97.5%
(Sharmila and Geethanjali 2016)	BD-E	Non-seizure, Seizure	DWT	NB and KNN	96.5%, 96.33%
(Wang et al. 2017)	ABCD-E	Non-seizure, Seizure	DWT	Feature extraction across multiple domains and nonlinear analysis	99.25%
Proposed Method	B-D-E	Non-seizure, Interictal, Seizure	Spectrogram image obtained using STFT	Resnet18, Densenet121, VGG16 and Squeezenet1_0 + XAI method SHAP for explaining the classification results	99.7%, 99.62%, 99.47% and 99.17%

In existing literature, some epileptic seizure detection systems have employed transfer learning techniques for EEG signal classification. A comparison with such systems is shown in Table 5.5. Different preprocessing techniques are used by these systems before classification using the transfer learning approach.

Table 5.5 Comparison with Existing Systems Using Transfer Learning Approach (Prajakta Rathod et al. 2025a)

Author	**Pre-Processing Technique**	**Classification Method**
(Raghu et al. 2020)	Concatenated spectrogram	Transfer learning with pre-trained networks, SVM
(Daoud and Bayoumi 2019)	Raw EEG signal	Transfer learning with DCAE, Bi-LSTM
(Gao et al. 2020)	PSDED	Transfer learning with DCNN
(Abdelhameed and Bayoumi 2018)	Unlabeled raw EEG	Transfer learning with DCAE and Bi-LSTM
(Hu et al. 2021)	Hand-crafted fusion feature	Transfer learning with pre-trained DNNs, HNN
(Qu and Yuan 2019)	CWT	Transfer learning with Alexnet, SVM
(Wang et al. 2019)	MAS	Transfer learning with Resnet152, Inception-v3 and Inception-Resnet-v2
Our Method	Spectrogram image obtained using STFT	Transfer learning with fine-tuning using Resnet18, Densenet121, VGG16 and Squeezenet1_0 + XAI method SHAP for explaining the classification results

Tables 4.3 and 4.4 show that Resnet18 achieved better classification accuracy than other CNN models. Moreover, as shown in Fig. 4.14, the SHAP method produced better results with VGG16 and Squeezenet1_0 than Densenet121 and Resnet18.

5.4.5 SHAP method with Resnet18, Densenet121, VGG16 and identification of top contributing channels

The proposed method identifies epileptic and non-epileptic signals using three CNN models and explains the network prediction using the XAI method SHAP, as shown in Fig. 3.15. Furthermore, the SHAP method from XAI is used to determine the

channels that contribute most significantly to the classification. The experiments are performed on the TUEP dataset, as described in Table 3.2. Multichannel EEG signals are converted into spectrogram images, which are further enhanced by applying filters. CNN models are trained with 30000 images, using a fine-tuning-based transfer learning approach, as shown in Fig. 3.16. This method can be very helpful for EEG technologists as they can identify epileptic and non-epileptic signals. Moreover, they can understand why the network has identified a specific signal as epileptic or non-epileptic and trust the network's decision.

Very limited epilepsy detection systems are reported in the literature, which explain the predictions generated by the model. Therefore, many times, EEG technologists can not understand why the system has arrived at a particular decision. The proposed study makes the epileptic seizure detection system more trustworthy by explaining the generated results. The inclusion of these explanations has made the system more likely to be adopted in clinical practice.

When EEG signals are converted into image form, classification results are better than when directly using them in signal form (Cho and Jang 2020). Therefore, in the proposed method, EEG signals are converted into images by vertically concatenating spectrograms of all 21 channels, as shown in Fig. 3.12 and Fig. 3.13. The visual appearance of images is improved by applying filters like sharpening and contrast enhancement, as shown in Fig. 3.14. Epileptic and non-epileptic signals have different powers and different frequencies at specific channels. The classifier learns this difference in power and frequency and classifies epileptic and non-epileptic signals. In the spectrogram image, low power is indicated by greenish light-yellow colour, and high power is indicated by red colour. After applying filters like sharpening and contrast

enhancement, the red colour is converted to a blackish shade, and the light green colour is converted to a dark green shade. Thus, the difference between epileptic and non-epileptic images appears better to EEG technologists, the intended users of the system.

Epileptic images have high power in lower, middle and high-frequency regions at some specific channels. The SHAP method produced pink marks for these channels and frequency regions, as shown in Fig. 4.17. Similarly, non-epileptic images have low power in the lower, middle and high-frequency regions for some specific channels. SHAP method marked these channels and frequency regions with pink shades, as shown in Fig. 4.18. It shows that, based on the higher and lower power at specific channels in specific frequency ranges, the images are classified into epileptic and non-epileptic classes. Fig. 3.12 and Fig. 3.13 show details of these channels and frequency ranges.

Densenet121 is used to identify the top contributing channels with the help of the relevance matrix generated by the SHAP method. Channels above the median value are identified as top contributing channels, as described in Section 3.2.5.2.3.2. For epileptic and non-epileptic patients, the median of the sum of relevance values of all channels is as shown in Fig. 5.1. Median values for non-epileptic patients are higher than median values for epileptic patients. This discrepancy may have arisen because non-epileptic signals are consistently non-epileptic over their entire duration, whereas epileptic signals often contain only a brief interval of epileptic activity within the full signal length.

The top contributing EEG channels in the classification decision are identified for each patient. A standardised electrode placement technique, the International 10–20 system, is utilised for recording EEG signals. This electrode placement system

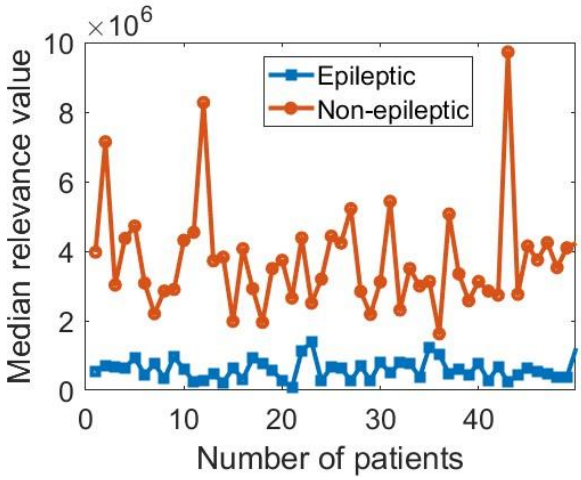


Fig. 5.1 Median plot of the sum of relevance values of all channels as described in Section 3.2.5.2.3.2 (Prajakta Rathod et al. 2025)

attaches each electrode or channel to a specific position on the head. The top contributing EEG channels of epileptic patients possess a seizure pattern in EEG recordings. Therefore, it is convenient for neurologists or EEG technicians to cross-verify the channels showing seizure patterns in the original EEG signals. Moreover, as these channels are attached to specific positions on the head, these channels can identify epileptogenic brain regions, i.e., brain regions from where epileptic signals originate. After that, neurologists can plan treatment accordingly.

During an epileptic seizure, the brain undergoes a series of abnormal neural processes that disrupt its normal functioning. The event begins when neurons become hyperexcitable due to an imbalance between excitatory neurotransmitters. This results in hypersynchronous firing, where large groups of neurons discharge simultaneously, producing abnormal electrical patterns on the EEG. The excessive activity may remain localised in a focal region or spread across brain networks. A failure of inhibitory control mechanisms further amplifies this runaway excitation. As the abnormal discharges spread, clinical symptoms appear, ranging from convulsions and motor jerks

when the motor cortex is involved to visual disturbances or impaired consciousness when other regions are affected. Following the seizure, neurons enter a recovery phase known as the postictal state, during which EEG activity slows and patients often experience confusion, fatigue, or memory lapses. In essence, a seizure represents a breakdown of the brain's finely tuned balance of excitation and inhibition, leading to uncontrolled electrical storms that temporarily impair normal brain function (Wang et al. 2025).

During epileptic episodes, EEG signals display abnormal and characteristic patterns that reflect the underlying hypersynchronous neuronal activity in the brain. The most prominent features include spikes and sharp waves, which are sudden high-amplitude discharges with pointed peaks, and spike-and-wave complexes, where sharp spikes are followed by slow waves, often recurring rhythmically as seen in generalised seizures such as absence epilepsy. In addition, bursts of polyspikes and sustained rhythmic discharges are commonly observed during the ictal phase, frequently starting as low-amplitude, high-frequency activity that evolves into high-amplitude rhythmic spikes or slower theta and delta rhythms as the seizure progresses. This transition is accompanied by a breakdown of normal inhibitory control and an increase in neuronal synchrony, which is directly reflected in EEG recordings. Following the seizure, the postictal phase is usually characterised by generalised slowing, with predominance of delta activity, indicating transient neural exhaustion and recovery. These dynamic EEG patterns provide critical insight into seizure onset, progression, and termination, serving as an essential biomarker for epilepsy diagnosis and monitoring (Li et al. 2020). In the proposed method, these EEG patterns mainly appear in the top contributing channels

for epileptic patient identification. The SHAP method marks these epileptic patterns in the spectrogram images of the EEG.

Fig. 5.2 represents the Receiver Operating Characteristic (ROC) curve for Densenet121, Resnet18 and VGG16 models. The ROC curve is a graphical representation that illustrates the diagnostic ability of a classification model as its discrimination threshold varies. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR), thereby providing a comprehensive view of the trade-off between sensitivity and specificity. In this study, ROC analysis was employed to evaluate and compare the classification performance of three DCNN architectures - Densenet121, Resnet18, and VGG16 - on the test dataset. The ROC curves for all models demonstrated excellent separability between the positive and negative classes, indicating that the networks effectively distinguished target features. Table 5.6 represents the Area Under ROC Curve (AUROC) for Densenet121, Resnet18 and VGG16. Among the models, Densenet121 achieved an outstanding AUROC value of 1.0, representing perfect classification performance with complete separation of classes. Resnet18 followed closely with an AUROC of 0.99, signifying an almost ideal classification ability with minimal false classifications. Similarly, VGG16 achieved an AUROC of 0.9992, indicating an exceptionally high predictive accuracy, nearly equivalent to that of Densenet121. The negligible difference in AUROC values among these architectures underscores their robustness and strong feature extraction capabilities. Densenet121's superior performance can be attributed to its densely connected convolutional layers, which promote feature reuse and mitigate the vanishing gradient problem. In contrast, Resnet18's near-perfect AUROC reflects the effectiveness of residual learning in maintaining gradient flow across deep layers.

VGG16, despite being a relatively older architecture, demonstrated remarkable performance owing to its deep layered structure and uniform convolutional kernel design. Overall, the ROC analysis confirms that all three models achieved near-optimal classification results, with minimal trade-off between sensitivity and specificity. The high AUROC scores across the networks validate the efficiency and reliability of the proposed approach. These findings indicate that the models are highly capable of generalising to unseen data and can serve as dependable tools for accurate classification tasks in the medical domain.

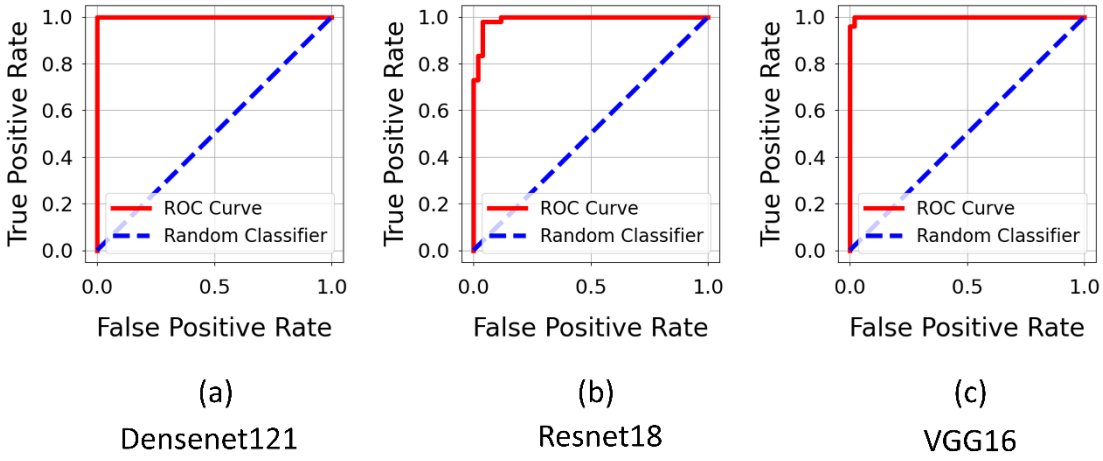


Fig. 5.2 ROC curve for Densenet121, Resnet18 and VGG16

Table 5.6 Area Under the ROC Curve for Densenet121, Resnet18 and VGG16

Network	**Area under the ROC curve**
Densenet121	1.0
Resnet18	0.99
VGG16	0.9992

In the literature, some systems are found to use the same dataset, i.e., the TUEP dataset, to identify epileptic signals. Table 5.7 shows a comparison with such systems.

Table 5.7 Comparison with Existing Systems Using the TUEP Dataset in Terms of Classification Accuracy (Prajakta Rathod et al. 2025)

Author	**Available Dataset Details**	**Pre-processing Technique**	**Classification Method**	**Accuracy**
(Prabin Jose et al. 2021)	TUEP dataset	bandpass filter is used to remove artifacts	deep stacked autoencoder using adaptive rag-Rider optimization algorithm	91.5%,
(Guerrero et al. 2021)	TUEP dataset - 40 patients	signals are transformed into the frequency domain	LR, ANN, SVM and CNN	86.0%
(Sudhamayee et al. 2023)	TUEP dataset - 10 patients	cliques of visibility graphs are analyzed to obtain maximal cliques, which are evaluated to calculate simplicial characterizers	SVM	92.86%
(Huang and Duan 2023)	TUEP dataset – 12000 segments	feature extraction is performed using Graph Convolutional Neural Networks (GCNN)	graph regularized fuzzy broad learning system	92.20%
(Gowtham Reddy N. et al. 2024)	TUEP dataset	DWT is applied to break down the signal into its time-frequency components	a fuzzy pattern tree classifier utilizing the Bonferroni mean operator	86.0%
(Jaffino et al. 2022)	TUEP dataset - 20 patients	weighted one-dimensional local binary pattern	deep Maxout network optimized using Taylor Henry gas solubility method	93.6%
(Tawhid et al. 2022)	TUEP dataset - 187 patients	signals are segmented into 3-sec. time frames	convolutional long short-term memory network	92.17%
(Omari et al. 2023)	TUEP dataset	EEG data is segmented into 2 sec. duration epochs with an overlap of 1 sec., and PCA is applied for dimensionality reduction	LConvNet	97.0%
(Tang et al. 2024)	TUEP dataset	path signature algorithm	Bi-LSTM with an attention mechanism	95.60
Proposed Method	TUEP dataset - 100 patients	a concatenated spectrogram image is obtained using STFT, and the visual representation of the image is enhanced using a filter	transfer learning with fine-tuning using Densenet121, Resnet18 and VGG16. Moreover, the XAI method SHAP is used for explaining the results and for identifying the top contributing channels in the classification	91.29%, 95.00%, 97.55%,

As shown in Table 4.6, VGG16 and Resnet18 achieved an accuracy of 97.55% and 95.00% on the test dataset, respectively, surpassing the performance of existing systems

on the same dataset. The proposed pre-processing technique, as described in Section 3.2.5.1, and training the CNN model with a transfer learning approach played a key role in achieving better accuracy. It shows that the proposed pre-processing method enables efficient extraction of distinguishing features from epileptic and non-epileptic signals, facilitating accurate identification of epileptic patients.

Table 4.6 shows that VGG16 achieved higher classification accuracy than other CNN models. Table 4.8 displays the performance metrics of the three CNN models. A visual comparison of the classification results from the three CNN models is provided in Fig. 4.15 and Fig. 4.16. As shown in Fig. 4.19, Densenet121 produced the best results with the SHAP method among all three CNN models. Table 4.9 shows AOPC values for all three CNN models. Densenet121 scores a higher AOPC value of 355350 than Resnet18 and VGG16.

In this method, the initial stage is the preprocessing of EEG signals, which includes removing artefacts from signals using a band-pass filter, downsampling the signals to reduce processing overhead, and selecting signals from the common 21 channels. Then, the spectrograms of all these 21 channels are concatenated vertically to obtain one image. The visual appearance of images is improved by applying filters. Then, CNN models are used for classification, and the XAI method SHAP is used to explain the network predictions. VGG16 and Resnet18 attained accuracies of 97.55% and 95.00%, respectively, on the test set using the proposed method. This is the highest accuracy reported so far. Also, validation of classification results is provided using the XAI method SHAP, which is not reported in the literature.

The computational complexity of the epilepsy detection system varies depending on the DNN architecture employed - VGG16, Resnet18, and Densenet121.

Among these, VGG16 exhibits the highest computational load due to its deep sequential structure with a large number of parameters (approximately 138 million). Its use of multiple convolutional layers with large kernel sizes results in high memory and processing requirements, making it computationally intensive during both training and inference. In contrast, Resnet18 introduces residual connections that enable efficient gradient flow and reduce the number of parameters to around 11 million, significantly lowering computational cost while maintaining strong classification performance. Densenet121, with its dense connectivity pattern, promotes feature reuse and parameter efficiency, resulting in approximately 8 million parameters and improved computational economy compared to VGG16. All three architectures demonstrate excellent capability for classifying epileptic and non-epileptic EEG spectrogram images; however, Densenet121 offers the best trade-off between accuracy and computational complexity. Thus, while VGG16 demands high computational resources, Resnet18 and Densenet121 provide more practical and efficient alternatives for large-scale or real-time epilepsy detection systems. The preprocessing steps, like band-pass filtering and down-sampling, reduced computational overhead. Signal segmentation and spectrogram generation add some computational overhead. The integration of the SHAP method for interpretability increases computational demand during post-analysis, as it computes feature contribution values for each prediction. Despite this, SHAP analysis was applied selectively to ensure interpretability without excessive runtime cost. Overall, the system achieves an efficient trade-off between accuracy, interpretability, and computational expense, making it suitable for both research and potential clinical applications.

The proposed epilepsy detection system exhibits notable advantages over the previously reported systems in terms of computational efficiency and interpretability. Unlike optimisation-intensive approaches such as the adaptive rag-bull rider (Prabin Jose et al. 2021) or Taylor–Henry gas solubility optimisation methods (Jaffino et al. 2022), the proposed CNN–SHAP framework eliminates the need for iterative metaheuristic tuning, thereby reducing training time and computational complexity. The model architecture is lightweight, requiring fewer parameters and lower memory resources compared to deep stacked (Tawhid et al. 2022), hybrid models like CNN–LSTM (Omari et al. 2023) and Bi-LSTM with attention (Tang et al. 2024). This compactness allows faster convergence and real-time applicability, which are critical for clinical EEG analysis. In contrast to complex graph-regularised (Huang and Duan 2023) or fuzzy systems (Gowtham Reddy N. et al. 2024), the proposed method relies on a straightforward convolutional design that ensures reproducibility and stability without compromising feature extraction performance. (Sudhamayee et al. 2023) adopted a simplicial network analysis approach rooted in complex network theory, emphasising topological structures in EEG data rather than deep feature extraction. Their method highlights signal connectivity and dynamical complexity, but lacks the end-to-end automated learning provided by CNNs. The system's explainability through SHAP distinctly enhances its transparency, offering clinicians intuitive insights into feature importance and decision logic, a capability lacking in traditional models (Guerrero et al. 2021). By balancing computational lightness and interpretability, the proposed model stands out as a practical and trustworthy AI solution for automated epilepsy detection.

In the literature, a similar approach of transfer learning is used in (Raghu et al. 2020), to categorise seven different types of seizure signals and non-seizure signals. They used signals with 19 channels obtained from the TUH EEG corpus and converted them into spectrogram images. They achieved 82.85% classification accuracy. In this proposed study, another subset of the TUH EEG corpus is used, and a filter is applied to images obtained from signals. In (Raghu et al. 2020), 83.46 hours of recordings are used to classify 8 classes, whereas in this proposed method, 33.33 hours of recordings are used to classify 2 classes. Moreover, the SHAP method is used to explain network prediction in the proposed method. A comparison of three CNN models - Densenet121, Resnet18 and VGG16 - is performed. It is found that Densenet121 produced the best results with the SHAP method. Therefore, Densenet121 is used with the SHAP method to identify the top contributing channels in classification. VGG16 could achieve an accuracy of 97.55% on the test dataset using the proposed method.

(Shankar et al. 2022) distinguish between five seizure categories and non-seizure signals using a combined CNN and LSTM model. EEG data from the TUH dataset is transformed into image form for experimentation. In (Gao et al. 2020), EEG signal data obtained from the CHB-MIT dataset are classified using transfer learning with DCNN. EEG signals are converted into PSDED and classified into interictal, preictal, and seizure classes. In (Hu et al. 2021), EEG signals from the CHB-MIT and iNeuro datasets are processed to extract features, which are then integrated into image representations. Five pre-trained DNNs and an HNN are used to perform classification. However, in these research studies, the explanation of classification results is missing. The proposed method used a similar approach of obtaining images from EEG signals obtained from the TUEP dataset and using transfer learning with fine-tuning to classify

them. The proposed method also explains the classification results and quantifies the explanations using the AOPC value, and the top contributing channels in classification are also identified.

In (Einizade et al. 2023), a fusion attentive deep multi-view network is proposed with an approach for artefact rejection. Experiments are performed on the TUEP dataset with an accuracy of 86%. They also interpret the brain regions involved in EEG seizures. However, the quantification of interpretability is very limited in (Einizade et al. 2023). In the proposed method, AOPC values are used to quantify the explanability. Moreover, the performance of three CNN models - VGG16, Resnet18 and Densenet121- is compared to identify the network that produces the best explanability.

In (Rashed-Al-Mahfuz et al. 2021), a similar work is performed on the Bonn dataset in which single-channel EEG signals are converted into spectrogram images and classified into seizure, non-seizure and interictal classes. Classification is performed using three methods: building a CNN-based model from the ground up, utilising transfer learning, and fine-tuning with VGG16 and Resnet50 networks. The SHAP method is used to explain classification results. The proposed method differs from that of (Rashed-Al-Mahfuz et al. 2021), by converting 21-channel EEG signals into spectrogram images and enhancing them using filters. In the proposed method, AOPC values are used to quantify explainability and to report a comparative analysis of three CNN models for the SHAP method results. Moreover, the proposed method identifies the top contributing channels in classification using the Densenet121 and the SHAP method.

5.5 Limitations

This research study has the following limitations:

1) Implementation of the XAI method LRP in the Bi-LSTM network is tested on the single-channel Bonn University dataset, which is smaller in size.
2) Some experiments, like testing the implementation of the LRP method in the VGG16 network and comparison of XAI methods, are conducted on the single-channel Bonn University dataset, which is smaller in size. These experiments can be further performed on a bigger and multichannel dataset.
3) In this research study, explainability is produced by the XAI methods on the images rather than the EEG signals. If the explainability is produced directly on EEG signals, then it can provide a better understanding to the neurologists and EEG technicians about the predictions of the network.
4) This system requires EEG technicians and neuroconsultants to go through training to understand the explanations of classification results produced on the EEG signal images.
5) This system lacks real-time EEG data monitoring. This system requires the conversion of EEG signals into images, which increases the processing time of EEG data.

Summary:

This research implements five methods to achieve the objectives outlined in Chapter 2, Section 2.2. The first four methods are applied to the single-channel dataset from Bonn University, while the fifth method is conducted on the multichannel TUH

EEG dataset. The classification accuracy of the networks in distinguishing epileptic from non-epileptic signals is improved through the proposed preprocessing technique. Additionally, a fine-tuning-based transfer learning approach reduces both training time and the amount of data required for model training. To ensure transparency, the XAI methods LRP and SHAP are employed, enabling neurologists to interpret and verify the decisions generated by the networks. The study further identifies the top contributing EEG channels involved in the classification of epileptic and non-epileptic patients. These channels can identify epileptogenic brain regions, i.e., the regions from which epileptic signals originated. Thus, it helps neurologists in treatment planning. Overall, integrating CNN models with XAI methods enhances the trustworthiness and reliability of the system, increasing the likelihood of its acceptance in the medical domain for epileptic seizure detection.

6. Conclusion

This work is focused on developing an XAI-based automated system for epilepsy detection. An automated epilepsy detection system has many benefits, such as early diagnosis, as it analyses longer EEG data in a short time. This system can be more helpful in rural areas with limited access to medical facilities. In this research, DNNs are used to classify epileptic signals through transfer learning with fine-tuning, a method that minimises the required number of training samples and reduces training time. Moreover, the system can leverage the benefits of previous learning of a model to identify curves, edges, and colours in the image.

In this research, five new methods have been developed to achieve the research objectives mentioned in Section 2.2. The first method is LRP in the Bi-LSTM Network. In this method, epileptic signals are classified using the Bi-LSTM network with an accuracy of 87.25%. The classification results are explained using the XAI method LRP. This method worked directly with EEG signals. The literature reports that the classification accuracy improves when EEG signals are converted into images. Therefore, in the second method, i.e., LRP in the VGG16 network, the classification accuracy is improved by converting EEG signals into images and selecting a suitable network that works with images. The VGG16 network is used with transfer learning with a fine-tuning approach to classify epileptic seizure and non-seizure signal images, resulting in a high accuracy of 99.89%. The LRP method is used to explain the classification results.

In the third method, i.e., comparison of the XAI methods SHAP, Guided Backpropagation and LIME, the performance of different XAI methods is compared to explain the classification results of EEG signal images. For this purpose, the most used XAI methods are selected, which are SHAP, guided backpropagation, and LIME. In this method, the VGG16 network achieved a classification accuracy of 98.98%. In the comparative analysis of the three XAI methods, it is found that the SHAP method produced the most insightful explanations for EEG signal classification outcomes. Accordingly, the subsequent phase of this research study, which is the fourth method, i.e., SHAP method with Resnet18, Densenet121, VGG16 and Squeezenet1_0, uses the XAI method SHAP, and compares the performance of four different DNNs for the identification of EEG signals as non-seizure, interictal and epileptic seizures. These four networks are Resnet18, Densenet121, VGG16 and Squeezenet1_0. Literature reports that these four networks achieved good performance in the identification of epileptic seizure signal images. The performance of these four networks is compared, and it is found that Resnet18 reached the best accuracy of 99.7%, outperforming the other three networks. The accuracy achieved by the other three networks, i.e., Densenet121, VGG16 and Squeezenet1_0, is 99.62%, 99.47% and 99.17%, respectively. The VGG16 network produced better results with the XAI method, SHAP, as compared to the other three networks. Moreover, the SHAP method produced better results for the seizure class than the interictal and non-seizure classes.

All the above experiments are conducted on a single-channel dataset from Bonn University. To strengthen the research study with the enhanced dataset, a multi-channel dataset from the TUH is used to perform the next set of experiments in the fifth method, i.e., SHAP method with Resnet18, Densenet121, VGG16 and identification of top

contributing channels. In this method, the signals obtained from the TUH dataset are converted into images by applying STFT. Further, these images are improved for better visual appearance by applying the filters. In the next step, the performance of three DNNs is compared for epileptic signal identification. These three DNNs are Densenet121, Resnet18 and VGG16. It is found that VGG16 achieved the best classification accuracy of 97.55% using the test dataset and 97.1% ± 0.2522 using the validation dataset. Resnet18 and Densenet121 achieved an accuracy of 95.0% and 91.29%, respectively, on the test dataset and 94.005 ± 0.2151 and 89.441 ± 0.3162 on the validation dataset, respectively. Further, the results generated by the XAI method SHAP with three DNNs are also compared. It is found that Densenet121 produced better explanations for the classification results than Resnet18 and VGG16 with the SHAP method. Therefore, Densenet121 is used to identify the top contributing channels in classifying each patient as epileptic or non-epileptic. The TUH EEG data is recorded using the International 10-20 system, where each electrode is positioned at a designated location on the scalp. Major epileptic patterns are present in the channels identified as the top contributing channels for epileptic patients by the Densenet121. These channels are attached to the specific head regions. It means that these channels can identify the epileptogenic brain region from where epileptic signals originate. It helps neurologists diagnose the type of epilepsy and plan treatment accordingly.

In this research study, XAI methods LRP and SHAP have provided a visual explanation for models' decisions to identify a person as an epileptic patient. This transparency in the models' decisions makes the system more trustworthy. It also helps adapt an XAI-based system for epilepsy detection. This system can be used for the initial monitoring of the EEG for epilepsy diagnosis. This system can be used by

neuroconsultants to cross-verify the interpretations and confirm the diagnosis. Moreover, it can be used by EEG technologists to analyse the EEG while preparing reports.

6.1 Future enhancement

This research study can be improved further by incorporating the following enhancements:

1) Top contributing channels determined with the help of the SHAP method can identify the epileptogenic brain regions of epilepsy patients. These channels can be categorised based on their corresponding brain regions, thereby enabling the automatic classification of the epilepsy type. This will help neurologists formulate an appropriate treatment plan.
2) Implementation of the LRP method in the Bi-LSTM network is tested on the single-channel dataset from Bonn University, which has limited data samples. It can be further tested on multichannel datasets with large data samples.
3) Implementation of the LRP method in the VGG16 network is tested on the images obtained from the single-channel Bonn University EEG dataset, which has limited data samples. It can be further tested on multichannel datasets with large data samples.
4) A performance comparison of three XAI techniques—SHAP, guided backpropagation, and LIME—is carried out in this research study on the single-channel dataset from Bonn University. These methods can be further tested on another multichannel dataset. Moreover, other XAI methods like saliency map and Grad CAM can be considered for comparison.

5) This research study compares the performance of three DNNs: Resnet18, Densenet121, and VGG16 for epilepsy detection using a multi-channel EEG dataset from the TUH. Moreover, results produced by the XAI method SHAP with these three networks are also compared, in which the SHAP method produced explanations of classification results on EEG signal images. This research can be further extended to produce explanations directly on EEG signals rather than images. It will enhance the research outcome and help neuroconsultants better understand the XAI method's explanation.
6) This epilepsy detection system can be further enhanced for real-time seizure onset detection, and the system predictions can be explained with the XAI method, SHAP.

List of Publications

Journal publication:

1) **Rathod, P.**, Naik, S. & Bhalodiya, J.M. Epileptic signal classification using convolutional neural network and Shapley additive explainable artificial intelligence method. *Neural Comput & Applic* **37**, 4937–4955 (2025). https://doi.org/10.1007/s00521-024-10915-7

IEEE conference publications:

1) **P. Rathod** and S. Naik, "Review on Epilepsy Detection with Explainable Artificial Intelligence," 2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22), Nagpur, India, 2022, pp. 1-6, doi: 10.1109/ICETET-SIP-2254415.2022.9791595.

2) **P. Rathod**, J. Bhalodiya and S. Naik, "Epilepsy Detection using Bi-LSTM with Explainable Artificial Intelligence," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039816.

3) **P. Rathod**, S. Naik and J. Bhalodiya, "Epilepsy Detection with CNN and Explanation with Layer-wise Relevance Propagation," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10726082.

4) **P. Rathod**, S. Naik and J. Bhalodiya, "Epileptic Signal Classification with CNN Models and Explainable AI," 2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM), Bangkok, Thailand, 2025, pp. 1-7, doi: 10.1109/IMCOM64595.2025.10857540.

5) **P. Rathod**, S. Naik and J. Bhalodiya, "Performance Analysis of Explainable Artificial Intelligence Methods in Epilepsy Detection," 2025 22nd International Learning and Technology Conference (L&T), jeddah, Saudi Arabia, 2025, pp. 321-326, doi: 10.1109/LT64002.2025.10940497.

Bibliography

Abdelhameed, Ahmed M., and Magdy Bayoumi. 2018. "Semi-Supervised Deep Learning System for Epileptic Seizures Onset Prediction." *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, December, 1186–91. https://doi.org/10.1109/ICMLA.2018.00191.

Abedin, Md. Zainal, Sharmin Akther, and Mohammad Shahadat Hossain. 2019. "An Artificial Neural Network Model for Epilepsy Seizure Detection." *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, September, 860–65. https://doi.org/10.1109/ICAEE48663.2019.8975569.

Abhishek S., Sachin Kumar S., Neethu Mohan, and Soman K.p. 2024. "EEG Based Automated Detection of Seizure Using Machine Learning Approach and Traditional Features." *Expert Systems with Applications* 251 (October): 123991. https://doi.org/10.1016/j.eswa.2024.123991.

"Adjust Image Contrast Using Histogram Equalization." n.d. Accessed April 24, 2025. https://in.mathworks.com/help/images/histogram-equalization.html.

Ahmad, Ijaz, Zhenzhen Liu, Lin Li, et al. 2024. "Robust Epileptic Seizure Detection Based on Biomedical Signals Using an Advanced Multi-View Deep Feature Learning Approach." *IEEE Journal of Biomedical and Health Informatics* 28 (10): 5742–54. https://doi.org/10.1109/JBHI.2024.3396130.

Ahmad, Ijaz, Mingxing Zhu, Guanglin Li, Danish Javeed, Prabhat Kumar, and Shixiong Chen. 2024. "A Secure and Interpretable AI for Smart Healthcare System: A Case Study on Epilepsy Diagnosis Using EEG Signals." *IEEE Journal of Biomedical and Health Informatics* 28 (6): 3236–47. https://doi.org/10.1109/JBHI.2024.3366341.

Ahmedt-Aristizabal, David, Tharindu Fernando, Simon Denman, Lars Petersson, Matthew J. Aburn, and Clinton Fookes. 2020. "Neural Memory Networks for Seizure Type Classification." *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, July, 569–75. https://doi.org/10.1109/EMBC44109.2020.9175641.

Aliyu, Ibrahim, Yong Beom Lim, and Chang Gyoon Lim. 2019. "Epilepsy Detection in EEG Signal Using Recurrent Neural Network." *Proceedings of the 2019 3rd*

International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (New York, NY, USA), ISMSI '19, March 23, 50–53. https://doi.org/10.1145/3325773.3325785.

Amrani, Ghita, Amina Adadi, and Mohammed Berrada. 2024. "An Explainable Hybrid DNN Model for Seizure vs. Non-Seizure Classification and Seizure Localization Using Multi-Dimensional EEG Signals." *Biomedical Signal Processing and Control* 95 (September): 106322. https://doi.org/10.1016/j.bspc.2024.106322.

Andrzejak, Ralph G., Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger. 2001. "Indications of Nonlinear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity: Dependence on Recording Region and Brain State." *Physical Review E* 64 (6): 061907. https://doi.org/10.1103/PhysRevE.64.061907.

Antal, Dorin Cristian, Thomas Gabriel Schreiner, Theona Eliza Crihan, Bogdan Emilian Ignat, Victoria San Antonio-Arce, and Iulian Dan Cuciureanu. 2022. "Seizures and Multiple Sclerosis-more than an Epidemiological Association (Review)." *Experimental and Therapeutic Medicine* 24 (5): 689. https://doi.org/10.3892/etm.2022.11625.

Antoniades, Andreas, Loukianos Spyrou, David Martin-Lopez, et al. 2017. "Detection of Interictal Discharges With Convolutional Neural Networks Using Discrete Ordered Multichannel Intracranial EEG." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25 (12): 2285–94. https://doi.org/10.1109/TNSRE.2017.2755770.

Arras, Leila, Grégoire Montavon, Klaus-Robert Müller, and Wojciech Samek. 2017. "Explaining Recurrent Neural Network Predictions in Sentiment Analysis." In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, edited by Alexandra Balahur, Saif M. Mohammad, and Erik van der Goot. Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-5221.

Balestrini, Simona, Alexis Arzimanoglou, Ingmar Blümcke, et al. 2021. "The Aetiologies of Epilepsy." *Epileptic Disorders* 23 (1): 1–16. https://doi.org/10.1684/epd.2021.1255.

Banluesombatkul, Nannapas, Pichayoot Ouppaphan, Pitshaporn Leelaarporn, et al. 2021. "MetaSleepLearner: A Pilot Study on Fast Adaptation of Bio-Signals-Based Sleep Stage Classifier to New Individual Subject Using Meta-Learning." *IEEE Journal of Biomedical and Health Informatics* 25 (6): 1949–63. https://doi.org/10.1109/JBHI.2020.3037693.

Bartolini, Emanuele, Anna Rita Ferrari, Simona Fiori, and Stefania Della Vecchia. 2023. "Glycaemic Imbalances in Seizures and Epilepsy of Paediatric Age: A

Literature Review." *Journal of Clinical Medicine* 12 (7): 2580. https://doi.org/10.3390/jcm12072580.

Bekbalanova, Marzhan, Aliya Zhunis, and Zhasdauren Duisebekov. 2019. "Epileptic Seizure Prediction in EEG Signals Using EMD and DWT." *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*, December, 1–4. https://doi.org/10.1109/ICECCO48375.2019.9043270.

Bhatia, Mandip Singh, Saurabh C. Sharda, Gaurav Yadav, Sahil Mehta, Ritu Attri, and Neeraj Singla. 2022. "Etiology of New-Onset Seizures in Adult Patients of Different Age Groups Presenting to the Emergency Department in North India and Their Outcomes." *Journal of Family Medicine and Primary Care* 11 (11): 7129–35. https://doi.org/10.4103/jfmpc.jfmpc_730_22.

Bouazizi, Samar, and Hela Ltifi. 2024. "Enhancing Accuracy and Interpretability in EEG-Based Medical Decision Making Using an Explainable Ensemble Learning Framework Application for Stroke Prediction." *Decision Support Systems* 178 (March): 114126. https://doi.org/10.1016/j.dss.2023.114126.

Bozzi, Yuri, Simona Casarosa, and Matteo Caleo. 2012. "Epilepsy as a Neurodevelopmental Disorder." *Frontiers in Psychiatry* 3 (March): 19. https://doi.org/10.3389/fpsyt.2012.00019.

Cao, Jiuwen, Jiahua Zhu, Wenbin Hu, and Anton Kummert. 2020. "Epileptic Signal Classification With Deep EEG Features by Stacked CNNs." *IEEE Transactions on Cognitive and Developmental Systems* 12 (4): 709–22. https://doi.org/10.1109/TCDS.2019.2936441.

Chapatwala, Neeta, Chirag N. Paunwala, and Poojan Dalal. 2022. "An Explainable AI Approach towards Epileptic Seizure Detection." *2022 IEEE 19th India Council International Conference (INDICON)*, November, 1–6. https://doi.org/10.1109/INDICON56171.2022.10039982.

Chen, Xin, Yuanjie Zheng, Yi Niu, and Chengiiang Li. 2020. "Epilepsy Classification for Mining Deeper Relationships between EEG Channels Based on GCN." *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, July, 701–6. https://doi.org/10.1109/CVIDL51233.2020.00050.

Cho, Kyung-Ok, and Hyun-Jong Jang. 2020. "Comparison of Different Input Modalities and Network Structures for Deep Learning-Based Seizure Detection." *Scientific Reports* 10 (1): 122. https://doi.org/10.1038/s41598-019-56958-y.

Choi, Gwangho, Chulkyun Park, Junkyung Kim, et al. 2019. "A Novel Multi-Scale 3D CNN with Deep Neural Network for Epileptic Seizure Detection." *2019 IEEE International Conference on Consumer Electronics (ICCE)*, January, 1–2. https://doi.org/10.1109/ICCE.2019.8661969.

"Contrast Enhancement Techniques." n.d. Accessed April 22, 2025. https://in.mathworks.com/help/images/contrast-enhancement-techniques.html.

Daoud, Hisham, and Magdy A. Bayoumi. 2019. "Efficient Epileptic Seizure Prediction Based on Deep Learning." *IEEE Transactions on Biomedical Circuits and Systems* 13 (5): 804–13. https://doi.org/10.1109/TBCAS.2019.2929053.

Das, Arun, and Paul Rad. 2020. "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey." arXiv:2006.11371. Preprint, arXiv, June 23. https://doi.org/10.48550/arXiv.2006.11371.

Dell'Aquila, Jason Tyler, and Varun Soti. 2022. "Sleep Deprivation: A Risk for Epileptic Seizures." *Sleep Science* 15 (2): 245–49. https://doi.org/10.5935/1984-0063.20220046.

Ding, Yulan, and Wenshan Zhao. 2024. "Channel Selection for Seizure Detection Based on Explainable AI With Shapley Values." *IEEE Sensors Journal* 24 (16): 26126–35. https://doi.org/10.1109/JSEN.2024.3422388.

Einizade, Aref, Samaneh Nasiri, Mohsen Mozafari, Sepideh Hajipour Sardouie, and Gari D. Clifford. 2023. "Explainable Automated Seizure Detection Using Attentive Deep Multi-View Networks." *Biomedical Signal Processing and Control* 79 (January): 104076. https://doi.org/10.1016/j.bspc.2022.104076.

Emmady, Prabhu D., and Arayamparambil C. Anilkumar. 2025. "EEG Abnormal Waveforms." In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK557655/.

Emoryy, Shawn. 2025. "LVIS Corporation - LVIS Corporation Launches NeuroMatch Pro: AI-Powered Seizure and Spike Detection for Enhanced EEG Diagnostics." *LVIS Corporation*, February 6. https://lviscorp.com/en/lvis-corporation-launches-neuromatch-pro-ai-powered-seizure-and-spike-detection-for-enhanced-eeg-diagnostics/.

Espinosa-Jovel, Camilo, Rafael Toledano, Ángel Aledo-Serrano, Irene García-Morales, and Antonio Gil-Nagel. 2018a. "Epidemiological Profile of Epilepsy in Low Income Populations." *Seizure - European Journal of Epilepsy* 56 (March): 67–72. https://doi.org/10.1016/j.seizure.2018.02.002.

Espinosa-Jovel, Camilo, Rafael Toledano, Ángel Aledo-Serrano, Irene García-Morales, and Antonio Gil-Nagel. 2018b. "Epidemiological Profile of Epilepsy in Low Income Populations." *Seizure* 56 (March): 67–72. https://doi.org/10.1016/j.seizure.2018.02.002.

Faiq Ahmad Khan, Zainab Umar, Alireza Jolfaei, and Muhammad Tariq. 2024. "Explainable Fuzzy Deep Learning for Prediction of Epileptic Seizures Using EEG." *IEEE Transactions on Fuzzy Systems* 32 (10): 5428–37. https://doi.org/10.1109/TFUZZ.2024.3434709.

Fisher, Robert S., Carlos Acevedo, Alexis Arzimanoglou, et al. 2014. “ILAE Official Report: A Practical Clinical Definition of Epilepsy.” *Epilepsia* 55 (4): 475–82. https://doi.org/10.1111/epi.12550.

Gailus, Björn, Hannah Naundorf, Lisa Welzel, et al. 2021. “Long-Term Outcome in a Noninvasive Rat Model of Birth Asphyxia with Neonatal Seizures: Cognitive Impairment, Anxiety, Epilepsy, and Structural Brain Alterations.” *Epilepsia* 62 (11): 2826–44. https://doi.org/10.1111/epi.17050.

Galovic, Marian, Carolina Ferreira-Atuesta, Laura Abraira, et al. 2021. “Seizures and Epilepsy After Stroke: Epidemiology, Biomarkers and Management.” *Drugs & Aging* 38 (4): 285–99. https://doi.org/10.1007/s40266-021-00837-7.

Gao, Yunyuan, Bo Gao, Qiang Chen, Jia Liu, and Yingchun Zhang. 2020. “Deep Convolutional Neural Network-Based Epileptic Electroencephalogram (EEG) Signal Classification.” *Frontiers in Neurology* 11 (May). https://doi.org/10.3389/fneur.2020.00375.

“Global, Regional, and National Burden of Epilepsy, 1990–2021: A Systematic Analysis for the Global Burden of Disease Study 2021.” 2025. *The Lancet. Public Health* 10 (3): e203–27. https://doi.org/10.1016/S2468-2667(24)00302-5.

Gowtham Reddy N, Pragya Gupta, Debashree Guha, Manjunatha Mahadevappa, and Debjani Chakraborty. 2023. “Ordered Weighted Aggregation Operators Based Statistical Features for Seizure Classification Using EEG.” *2023 IEEE 20th India Council International Conference (INDICON)*, December, 293–98. https://doi.org/10.1109/INDICON59947.2023.10440939.

Gowtham Reddy N., Swati Rani Hait, Debashree Guha, and Manjunatha Mahadevappa. 2024. “Classification of Epileptic EEG Signals with the Utilization of Bonferroni Mean Based Fuzzy Pattern Tree.” *Expert Systems with Applications* 239 (April): 122424. https://doi.org/10.1016/j.eswa.2023.122424.

Guerrero, Maria Camila, Juan Sebastián Parada, and Helbert Eduardo Espitia. 2021. “EEG Signal Analysis Using Classification Techniques: Logistic Regression, Artificial Neural Networks, Support Vector Machines, and Convolutional Neural Networks.” *Heliyon* 7 (6): e07258. https://doi.org/10.1016/j.heliyon.2021.e07258.

Holzinger, Andreas, Anna Saranti, Christoph Molnar, Przemyslaw Biecek, and Wojciech Samek. 2022. “Explainable AI Methods - A Brief Overview.” In *xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*, edited by Andreas Holzinger, Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller, and Wojciech Samek. Springer International Publishing. https://doi.org/10.1007/978-3-031-04083-2_2.

Hu, Dinghan, Jiuwen Cao, Xiaoping Lai, Yaomin Wang, Shuang Wang, and Yao Ding. 2021. “Epileptic State Classification by Fusing Hand-Crafted and Deep Learning EEG Features.” *IEEE Transactions on Circuits and Systems II: Express Briefs* 68 (4): 1542–46. https://doi.org/10.1109/TCSII.2020.3031399.

Hu, Xinmei, Shasha Yuan, Fangzhou Xu, Yan Leng, Kejiang Yuan, and Qi Yuan. 2020. “Scalp EEG Classification Using Deep Bi-LSTM Network for Seizure Detection.” *Computers in Biology and Medicine* 124 (September): 103919. https://doi.org/10.1016/j.compbiomed.2020.103919.

Huang, Zixuan, and Junwei Duan. 2023. “GFBLS: Graph-Regularized Fuzzy Broad Learning System for Detection of Interictal Epileptic Discharges.” *Engineering Applications of Artificial Intelligence* 125 (October): 106763. https://doi.org/10.1016/j.engappai.2023.106763.

Huff, J. Stephen, Forshing Lui, and Najib I. Murr. 2025. “Psychogenic Nonepileptic Seizures.” In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK441871/.

Ijaz Ahmad, Chen Yao, Lin Li, et al. 2024. “An Efficient Feature Selection and Explainable Classification Method for EEG-Based Epileptic Seizure Detection.” *Journal of Information Security and Applications* 80 (February): 103654. https://doi.org/10.1016/j.jisa.2023.103654.

“Imsharpen.” n.d. Accessed April 22, 2025. https://in.mathworks.com/help/images/ref/imsharpen.html.

Islam, Shaikh Mahmudul, and Himadri Shekhar Mondal. 2019. “Image Enhancement Based Medical Image Analysis.” *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, July, 1–5. https://doi.org/10.1109/ICCCNT45670.2019.8944910.

Jaffino, G., M. Sundaram, and J. Prabin Jose. 2022. “Weighted 1D-Local Binary Pattern Features and Taylor-Henry Gas Solubility Optimization Based Deep Maxout Network for Discovering Epileptic Seizure Using EEG.” *Digital Signal Processing* 122 (April): 103349. https://doi.org/10.1016/j.dsp.2021.103349.

Jemal, Imene, Neila Mezghani, Lina Abou-Abbas, and Amar Mitiche. 2022. “An Interpretable Deep Learning Classifier for Epileptic Seizure Prediction Using EEG Data.” *IEEE Access* 10: 60141–50. https://doi.org/10.1109/ACCESS.2022.3176367.

Jhaveri, Dhanisha J., Aileen McGonigal, Christel Becker, et al. 2023. “Stress and Epilepsy: Towards Understanding of Neurobiological Mechanisms for Better Management.” *eNeuro* 10 (11): ENEURO.0200-23.2023. https://doi.org/10.1523/ENEURO.0200-23.2023.

Jose, J Prabin, M. Sundaram, and G. Jaffino. 2020. “FPGA Implementation of Epileptic Seizure Detection Using ELM Classifier.” *2020 Sixth International Conference*

on Bio Signals, Images, and Instrumentation (ICBSII), February, 1–5. https://doi.org/10.1109/ICBSII49132.2020.9167598.

Khan, Faiq Ahmad, Zainab Umar, Alireza Jolfaei, and Muhammad Tariq. 2024. "Explainable AI for Epileptic Seizure Detection in Internet of Medical Things." *Digital Communications and Networks*, ahead of print, August 22. https://doi.org/10.1016/j.dcan.2024.08.013.

"Lab2rgb." n.d. Accessed April 22, 2025. https://in.mathworks.com/help/images/ref/lab2rgb.html.

Lee, Sang Kun. 2014. "Old versus New: Why Do We Need New Antiepileptic Drugs?" *Journal of Epilepsy Research* 4 (2): 39–44. https://doi.org/10.14581/jer.14010.

Li, Gang, and Muhammad Adeel Khan. 2019. "Deep Learning on VR-Induced Attention." *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, December, 163–1633. https://doi.org/10.1109/AIVR46125.2019.00033.

Li, Pengcheng, Mio Yokoyama, Daiki Okamoto, Hironori Nakatani, and Tohru Yagi. 2024. "Resting-State EEG Features Modulated by Depressive State in Healthy Individuals: Insights from Theta PSD, Theta-Beta Ratio, Frontal-Parietal PLV, and sLORETA." *Frontiers in Human Neuroscience* 18 (August). https://doi.org/10.3389/fnhum.2024.1384330.

Li, Qiong, Jianbo Gao, Ziwen Zhang, Qi Huang, Yuan Wu, and Bo Xu. 2020. "Distinguishing Epileptiform Discharges From Normal Electroencephalograms Using Adaptive Fractal and Network Analysis: A Clinical Perspective." *Frontiers in Physiology* 11: 828. https://doi.org/10.3389/fphys.2020.00828.

Liang, Jian, Rui Lu, Changshui Zhang, and Fei Wang. 2016. "Predicting Seizures from Electroencephalography Recordings: A Knowledge Transfer Strategy." *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, October, 184–91. https://doi.org/10.1109/ICHI.2016.27.

Liang, Weixia, Haijun Pei, Qingling Cai, and Yonghua Wang. 2020. "Scalp EEG Epileptogenic Zone Recognition and Localization Based on Long-Term Recurrent Convolutional Network." *Neurocomputing* 396 (July): 569–76. https://doi.org/10.1016/j.neucom.2018.10.108.

Linardatos, Pantelis, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. 2020. "Explainable AI: A Review of Machine Learning Interpretability Methods." *Entropy (Basel, Switzerland)* 23 (1): 18. https://doi.org/10.3390/e23010018.

Liu, Jianguo, and Blake Woodson. 2019. "Deep Learning Classification for Epilepsy Detection Using a Single Channel Electroencephalography (EEG)." *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies* (New York, NY, USA), ICDLT '19, July 5, 23–26. https://doi.org/10.1145/3342999.3343008.

Lundberg, Scott, and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” arXiv:1705.07874. Preprint, arXiv, November 25. https://doi.org/10.48550/arXiv.1705.07874.

Mathur, Priyanka, and Vijay Kumar Chakka. 2020. “Graph Signal Processing of EEG Signals for Detection of Epilepsy.” *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, February, 839–43. https://doi.org/10.1109/SPIN48934.2020.9070326.

Mathur, Priyanka, Vijay Kumar Chakka, and Shaik Basheeruddin Shah. 2021. “Ramanujan Periodic Subspace Based Epileptic EEG Signals Classification.” *IEEE Sensors Letters* 5 (7): 1–4. https://doi.org/10.1109/LSENS.2021.3086755.

Montavon, Grégoire, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. 2019. “Layer-Wise Relevance Propagation: An Overview.” In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, edited by Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller. Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_10.

Mridha, Krishna, MD. Mezbah Uddin, Ebisa Leta Desisa, Asefa Teganu Feyisa, Segni Worku Guta, and Wolde Assefa Debele. 2023. “EEG-Based Epilepsy Seizure Classification Using Explainable Machine Learning Algorithms.” *2023 IEEE Fifth International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, September, 1–7. https://doi.org/10.1109/ICAECC59324.2023.10560233.

Munakomi, Sunil, and Joe M. Das. 2025. “Epilepsy Surgery.” In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK562151/.

Myers, Patrick, Kristin M. Gunnarsdottir, Adam Li, et al. 2025. “Diagnosing Epilepsy with Normal Interictal EEG Using Dynamic Network Models.” *Annals of Neurology* 97 (5): 907–18. https://doi.org/10.1002/ana.27168.

Nardone, Raffaele, Francesco Brigo, and Eugen Trinka. 2016. “Acute Symptomatic Seizures Caused by Electrolyte Disturbances.” *Journal of Clinical Neurology (Seoul, Korea)* 12 (1): 21–33. https://doi.org/10.3988/jcn.2016.12.1.21.

Nayak, Chetan S., and Arayamparambil C. Anilkumar. 2025. “EEG Normal Waveforms.” In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK539805/.

“Neuro Event Labs - Seizures. Clearly.” n.d. Accessed April 22, 2025. https://neuroeventlabs.com/.

Olokodana, Ibrahim L., Saraju P. Mohanty, and Elias Kougianos. 2020. “Distributed Kriging-Bootstrapped DNN Model for Fast, Accurate Seizure Detection from EEG Signals.” *2020 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July, 264–69. https://doi.org/10.1109/ISVLSI49217.2020.00055.

Omari, Swaleh, Michael Kimwele, Akeem Olowolayemo, and Dennis Kaburu. 2023. *Enhancing EEG Signals Classification Using LSTM-CNN Architecture*. https://doi.org/10.22541/au.168737045.52367778/v1.

Park, Eu Gene, and Il Han Yoo. 2025. "Risk of Epilepsy in Pediatric Patients with Febrile Seizures: Insights from Nationwide Registry Data in Korea." *Seizure* 127 (April): 29–35. https://doi.org/10.1016/j.seizure.2025.01.006.

Peker, Musa, Baha Sen, and Dursun Delen. 2016. "A Novel Method for Automated Diagnosis of Epilepsy Using Complex-Valued Classifiers." *IEEE Journal of Biomedical and Health Informatics* 20 (1): 108–18. https://doi.org/10.1109/JBHI.2014.2387795.

Pocevičiūtė, Milda, Gabriel Eilertsen, and Claes Lundström. 2020. "Survey of XAI in Digital Pathology." In *Artificial Intelligence and Machine Learning for Digital Pathology: State-of-the-Art and Future Challenges*, edited by Andreas Holzinger, Randy Goebel, Michael Mengel, and Heimo Müller. Springer International Publishing. https://doi.org/10.1007/978-3-030-50402-1_4.

Prabin Jose, J., M. Sundaram, and G. Jaffino. 2021. "Adaptive Rag-Bull Rider: A Modified Self-Adaptive Optimization Algorithm for Epileptic Seizure Detection with Deep Stacked Autoencoder Using Electroencephalogram." *Biomedical Signal Processing and Control* 64 (February): 102322. https://doi.org/10.1016/j.bspc.2020.102322.

Prajakta Rathod, Jayendra Bhalodiya, and Shefali Naik. 2022. "Epilepsy Detection Using Bi-LSTM with Explainable Artificial Intelligence." *2022 IEEE 19th India Council International Conference (INDICON)*, November, 1–6. https://doi.org/10.1109/INDICON56171.2022.10039816.

Prajakta Rathod and Shefali Naik. 2022. "Review on Epilepsy Detection with Explainable Artificial Intelligence." *2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22)*, April, 1–6. https://doi.org/10.1109/ICETET-SIP-2254415.2022.9791595.

Prajakta Rathod, Shefali Naik, and Jayendra Bhalodiya. 2024. "Epilepsy Detection with CNN and Explanation with Layer-Wise Relevance Propagation." *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, June, 1–6. https://doi.org/10.1109/ICCCNT61001.2024.10726082.

Prajakta Rathod, Shefali Naik, and Jayendra M. Bhalodiya. 2025. "Epileptic Signal Classification Using Convolutional Neural Network and Shapley Additive Explainable Artificial Intelligence Method." *Neural Computing and Applications* 37 (6): 4937–55. https://doi.org/10.1007/s00521-024-10915-7.

Prajakta Rathod, Shefali Naik, and Jayendra Bhalodiya. 2025a. "Epileptic Signal Classification with CNN Models and Explainable AI." *2025 19th International*

Conference on Ubiquitous Information Management and Communication (IMCOM), January, 1–7. https://doi.org/10.1109/IMCOM64595.2025.10857540.

Prajakta Rathod, Shefali Naik, and Jayendra Bhalodiya. 2025b. "Performance Analysis of Explainable Artificial Intelligence Methods in Epilepsy Detection." *2025 22nd International Learning and Technology Conference (L&T)* 22 (January): 321–26. https://doi.org/10.1109/LT64002.2025.10940497.

Psaizi. n.d. "Home." Accessed April 22, 2025. https://psaizi.com/.

Qu, Guiguo, and Qi Yuan. 2019. "Epileptogenic Region Detection Based on Deep CNN with Transfer Learning." *2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT)*, October, 73–77. https://doi.org/10.1109/ICAIT.2019.8935898.

Raghu, S., Natarajan Sriraam, Yasin Temel, Shyam Vasudeva Rao, and Pieter L. Kubben. 2020. "EEG Based Multi-Class Seizure Type Classification Using Convolutional Neural Network and Transfer Learning." *Neural Networks* 124 (April): 202–12. https://doi.org/10.1016/j.neunet.2020.01.017.

Rashed-Al-Mahfuz, Md., Mohammad Ali Moni, Shahadat Uddin, Salem A. Alyami, Matthew A. Summers, and Valsamma Eapen. 2021. "A Deep Convolutional Neural Network Method to Detect Seizures and Characteristic Frequencies Using Epileptic Electroencephalogram (EEG) Data." *IEEE Journal of Translational Engineering in Health and Medicine* 9: 1–12. https://doi.org/10.1109/JTEHM.2021.3050925.

Rashid, Adib Bin, and MD Ashfakul Karim Kausik. 2024. "AI Revolutionizing Industries Worldwide: A Comprehensive Overview of Its Diverse Applications." *Hybrid Advances* 7 (December): 100277. https://doi.org/10.1016/j.hybadv.2024.100277.

Rayi, Appaji, and Najib I. Murr. 2025. "Electroencephalogram." In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK563295/.

"Rgb2lab." n.d. Accessed April 22, 2025. https://in.mathworks.com/help/images/ref/rgb2lab.html.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA), KDD '16, August 13, 1135–44. https://doi.org/10.1145/2939672.2939778.

Ricci, S., F. Vigevano, M. Manfredi, and D. G. Kasteleijn-Nolst Trenité. 1998. "Epilepsy Provoked by Television and Video Games: Safety of 100-Hz Screens." *Neurology* 50 (3): 790–93. https://doi.org/10.1212/wnl.50.3.790.

Roohi-Azizi, Mahtab, Leila Azimi, Soomaayeh Heysieattalab, and Meysam Aamidfar. 2017. “Changes of the Brain’s Bioelectrical Activity in Cognition, Consciousness, and Some Mental Disorders.” *Medical Journal of the Islamic Republic of Iran* 31 (September): 53. https://doi.org/10.14196/mjiri.31.53.

Salem, Nema, Hebatullah Malik, and Asmaa Shams. 2019. “Medical Image Enhancement Based on Histogram Algorithms.” *Procedia Computer Science*, 16th Learning and Technology Conference 2019Artificial Intelligence and Machine Learning: Embedding the Intelligence, vol. 163 (January): 300–311. https://doi.org/10.1016/j.procs.2019.12.112.

Samek, Wojciech, Alexander Binder, Grégoire Montavon, Sebastian Lapuschkin, and Klaus-Robert Müller. 2017. “Evaluating the Visualization of What a Deep Neural Network Has Learned.” *IEEE Transactions on Neural Networks and Learning Systems* 28 (11): 2660–73. https://doi.org/10.1109/TNNLS.2016.2599820.

Samek, Wojciech, Grégoire Montavon, Sebastian Lapuschkin, Christopher J. Anders, and Klaus-Robert Müller. 2021. “Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications.” *Proceedings of the IEEE* 109 (3): 247–78. https://doi.org/10.1109/JPROC.2021.3060483.

San-Segundo, Rubén, Manuel Gil-Martín, Luis Fernando D’Haro-Enríquez, and José Manuel Pardo. 2019. “Classification of Epileptic EEG Recordings Using Signal Transforms and Convolutional Neural Networks.” *Computers in Biology and Medicine* 109 (June): 148–58. https://doi.org/10.1016/j.compbiomed.2019.04.031.

Shankar, Anand, Samarendra Dandapat, and Shovan Barma. 2022. “Seizure Types Classification by Generating Input Images With In-Depth Features From Decomposed EEG Signals for Deep Learning Pipeline.” *IEEE Journal of Biomedical and Health Informatics* 26 (10): 4903–12. https://doi.org/10.1109/JBHI.2022.3159531.

Shanmugam, Shalini, and Selvathi Dharmar. 2023. “A CNN-LSTM Hybrid Network for Automatic Seizure Detection in EEG Signals.” *Neural Computing and Applications* 35 (28): 20605–17. https://doi.org/10.1007/s00521-023-08832-2.

Sharmila, A., and P. Geethanjali. 2016. “DWT Based Detection of Epileptic Seizure From EEG Signals Using Naive Bayes and K-NN Classifiers.” *IEEE Access* 4: 7716–27. https://doi.org/10.1109/ACCESS.2016.2585661.

Shoeb, Ali. 2010. “CHB-MIT Scalp EEG Database.” Physionet.org. https://doi.org/10.13026/C2K01R.

Shoeb, Ali, and John Guttag. 2010. “Application of Machine Learning to Epileptic Seizure Detection.” *Proceedings of the 27th International Conference on International Conference on Machine Learning* (Madison, WI, USA), ICML’10, June 21, 975–82. https://dl.acm.org/doi/10.5555/3104322.3104446.

Shrikumar, Avanti, Peyton Greenside, and Anshul Kundaje. 2019. “Learning Important Features Through Propagating Activation Differences.” arXiv:1704.02685. Preprint, arXiv, October 12. https://doi.org/10.48550/arXiv.1704.02685.

Sidulova, Mariia, Nina Nehme, and Chung Huyk Park. 2021. “Towards Explainable Image Analysis for Alzheimer’s Disease and Mild Cognitive Impairment Diagnosis.” *2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, October, 1–6. https://doi.org/10.1109/AIPR52630.2021.9762082.

Singh, Amitojdeep, Sourya Sengupta, and Vasudevan Lakshminarayanan. 2020. “Explainable Deep Learning Models in Medical Image Analysis.” *Journal of Imaging* 6 (6): 6. https://doi.org/10.3390/jimaging6060052.

Sood, Vidhi, Deepak Kumar, Vijay Anant Athavale, and Suresh Chand Gupta. 2020. “Epilepsy Seizure Detection by Using Bayesian Optimize Bi-LSTM Model.” *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, November, 291–96. https://doi.org/10.1109/PDGC50313.2020.9315779.

Soufineyestani, Mahsa, Dale Dowling, and Arshia Khan. 2020. “Electroencephalography (EEG) Technology Applications and Available Devices.” *Applied Sciences* 10 (21): 21. https://doi.org/10.3390/app10217453.

Springenberg, Jost Tobias, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. 2015. “Striving for Simplicity: The All Convolutional Net.” arXiv:1412.6806. Preprint, arXiv, April 13. https://doi.org/10.48550/arXiv.1412.6806.

Stafstrom, Carl E., and Lionel Carmant. 2015. “Seizures and Epilepsy: An Overview for Neuroscientists.” *Cold Spring Harbor Perspectives in Medicine* 5 (6): a022426. https://doi.org/10.1101/cshperspect.a022426.

Sudhamayee, K., M. Gopal Krishna, and P. Manimaran. 2023. “Simplicial Network Analysis on EEG Signals.” *Physica A: Statistical Mechanics and Its Applications* 630 (November): 129230. https://doi.org/10.1016/j.physa.2023.129230.

Tang, Yixuan, Qianyi Wu, Haifeng Mao, and Lihua Guo. 2024. “Epileptic Seizure Detection Based on Path Signature and Bi-LSTM Network With Attention Mechanism.” *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 32: 304–13. https://doi.org/10.1109/TNSRE.2024.3350074.

Taqi, Arwa M., Fadwa Al-Azzo, M. Mariofanna, and Jassim M. Al-Saadi. 2017. “Classification and Discrimination of Focal and Non-Focal EEG Signals Based on Deep Neural Network.” *2017 International Conference on Current Research in Computer Science and Information Technology (ICCIT)*, April, 86–92. https://doi.org/10.1109/CRCSIT.2017.7965539.

Tawhid, Md. Nurul Ahad, Siuly Siuly, and Tianning Li. 2022. “A Convolutional Long Short-Term Memory-Based Neural Network for Epilepsy Detection From

EEG." *IEEE Transactions on Instrumentation and Measurement* 71: 1–11. https://doi.org/10.1109/TIM.2022.3217515.

Thara D.k., PremaSudha B.g, and Fan Xiong. 2019. "Auto-Detection of Epileptic Seizure Events Using Deep Neural Network with Different Feature Scaling Techniques." *Pattern Recognition Letters* 128 (December): 544–50. https://doi.org/10.1016/j.patrec.2019.10.029.

Torse, Dattaprasad A., Rajashri Khanai, and Veena V. Desai. 2019. "Classification of Epileptic Seizures Using Recurrence Plots and Machine Learning Techniques." *2019 International Conference on Communication and Signal Processing (ICCSP)*, April, 0611–15. https://doi.org/10.1109/ICCSP.2019.8697989.

Truong, Nhan Duy, and Omid Kavehei. 2019. "Low Precision Electroencephalogram for Seizure Detection with Convolutional Neural Network." *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, March, 299–301. https://doi.org/10.1109/AICAS.2019.8771569.

Vargas, Dionathan Luan de, Jefferson Tales Oliva, Marcelo Teixeira, Dalcimar Casanova, and João Luís Garcia Rosa. 2023. "Feature Extraction and Selection from Electroencephalogram Signals for Epileptic Seizure Diagnosis." *Neural Computing and Applications* 35 (16): 12195–219. https://doi.org/10.1007/s00521-023-08350-1.

Velden, Bas H. M. van der, Hugo J. Kuijf, Kenneth G. A. Gilhuijs, and Max A. Viergever. 2022. "Explainable Artificial Intelligence (XAI) in Deep Learning-Based Medical Image Analysis." *Medical Image Analysis* 79 (July): 102470. https://doi.org/10.1016/j.media.2022.102470.

Veloso, L., J. McHugh, E. von Weltin, S. Lopez, I. Obeid, and J. Picone. 2017. "Big Data Resources for EEGs: Enabling Deep Learning Research." *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, December, 1–3. https://doi.org/10.1109/SPMB.2017.8257044.

Wang, Gang, Dong Wang, Changwang Du, et al. 2020. "Seizure Prediction Using Directed Transfer Function and Convolution Neural Network on Intracranial EEG." *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society* 28 (12): 2711–20. https://doi.org/10.1109/TNSRE.2020.3035836.

Wang, Jiaqi, Suhui Kuang, Zhirong Wei, and Shuli Liang. 2025. "Research Progress of Connexins in Epileptogensis." *Acta Epileptologica* 7 (1): 14. https://doi.org/10.1186/s42494-025-00203-9.

Wang, Lina, Weining Xue, Yang Li, et al. 2017. "Automatic Epileptic Seizure Detection in EEG Signals Using Multi-Domain Feature Extraction and Nonlinear Analysis." *Entropy* 19 (6): 6. https://doi.org/10.3390/e19060222.

Wang, Yaomin, Jiuwen Cao, Jianzhong Wang, Dinghan Hu, and Muqing Deng. 2019. "Epileptic Signal Classification with Deep Transfer Learning Feature on Mean Amplitude Spectrum." *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July, 2392–95. https://doi.org/10.1109/EMBC.2019.8857082.

Wang, Yueming, Kang Lin, Yu Qi, et al. 2018. "Estimating Brain Connectivity With Varying-Length Time Lags Using a Recurrent Neural Network." *IEEE Transactions on Biomedical Engineering* 65 (9): 1953–63. https://doi.org/10.1109/TBME.2018.2842769.

Wojewodka, Gabriella, Alison McKinlay, and Leone Ridsdale. 2021. "Best Care for Older People with Epilepsy: A Scoping Review." *Seizure* 85 (February): 70–89. https://doi.org/10.1016/j.seizure.2020.12.007.

Wolfe, Caitlin E., David M. Wood, Alison Dines, et al. 2019. "Seizures as a Complication of Recreational Drug Use: Analysis of the Euro-DEN Plus Data-Set." *Neurotoxicology* 73 (July): 183–87. https://doi.org/10.1016/j.neuro.2019.04.003.

Wong, Sheng, and Levin Kuhlmann. 2020. "Computationally Efficient Epileptic Seizure Prediction Based on Extremely Randomised Trees." *Proceedings of the Australasian Computer Science Week Multiconference* (New York, NY, USA), ACSW '20, February 4, 1–3. https://doi.org/10.1145/3373017.3373058.

Wu, Haiyue, Aihua Huang, and John W. Sutherland. 2022. "Layer-Wise Relevance Propagation for Interpreting LSTM-RNN Decisions in Predictive Maintenance." *The International Journal of Advanced Manufacturing Technology* 118 (3): 963–78. https://doi.org/10.1007/s00170-021-07911-9.

Xiaofei Ye, Panpan Hu, Bin Yang, et al. 2024. "Using Scalp EEG to Predict Seizure Recurrence and Electrical Status Epilepticus in Children with Idiopathic Focal Epilepsy." *Seizure* 118 (May): 8–16. https://doi.org/10.1016/j.seizure.2024.03.013.

Xie, Lixiao, Zhaohong Deng, Peng Xu, Kup-Sze Choi, and Shitong Wang. 2019. "Generalized Hidden-Mapping Transductive Transfer Learning for Recognition of Epileptic Electroencephalogram Signals." *IEEE Transactions on Cybernetics* 49 (6): 2200–2214. https://doi.org/10.1109/TCYB.2018.2821764.

Xu, Shan, Yan Piao, Enjia Li, and Yue Wang. 2021. "Epileptic Seizure Detection Based on Multidimensional Feature Extraction." *2021 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*, July, 157–60. https://doi.org/10.1109/MLISE54096.2021.00035.

Yang, Christopher C. 2022. "Explainable Artificial Intelligence for Predictive Modeling in Healthcare." *Journal of Healthcare Informatics Research* 6 (2): 228–39. https://doi.org/10.1007/s41666-022-00114-1.

Zöllner, Johann Philipp, Friedhelm C. Schmitt, Felix Rosenow, et al. 2021. “Seizures and Epilepsy in Patients with Ischaemic Stroke.” *Neurological Research and Practice* 3 (1): 63. https://doi.org/10.1186/s42466-021-00161-w.