IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 November 2023, accepted 2 March 2024, date of publication 11 March 2024, date of current version 15 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3375877

METHODS

Toward Transparent AI for Neurological Disorders: A Feature Extraction and Relevance Analysis Framework

MITCHELL D. WOODBRIGHT[1], AHSAN MORSHED[2], (Member, IEEE), MATTHEW BROWNE[3], BIPLOB RAY[2], (Senior Member, IEEE), AND STEVEN MOORE[4]

[1]School of Engineering and Technology, Central Queensland University, Bundaberg, QLD 4670, Australia
[2]School of Engineering and Technology, Central Queensland University, Melbourne, VIC 3000, Australia
[3]School of Medical Health and Applied Sciences, Central Queensland University, Bundaberg, QLD 4670, Australia
[4]School of Engineering and Technology, Central Queensland University, Rockhampton, QLD 4702, Australia

Corresponding author: Mitchell D. Woodbright (m.woodbright@cqu.edu.au)

ABSTRACT The lack of interpretability and transparency in deep learning architectures has raised concerns among professionals in various industries and academia. One of the main concerns is the ability to trust these architectures' without being provided any insight into the decision-making process. Despite these concerns, researchers continue to explore new models and architectures that do not incorporate explainability into their main construct. In the medical industry, it is crucial to provide explanations of any decision, as patient health outcomes can vary according to decisions made. Furthermore, in medical research, incorrectly diagnosed neurological conditions are a high-cost error that contributes significantly to morbidity and mortality. Therefore, the development of new transparent techniques for neurological conditions is critical. This paper presents a novel Autonomous Relevance Technique for an Explainable neurological disease prediction framework called ART-Explain. The proposed technique autonomously extracts features from within the deep learning architecture to create novel visual explanations of the resulting prediction. ART-Explain is an end-to-end autonomous explainable technique designed to present an intuitive and holistic overview of a prediction made by a deep learning classifier. To evaluate the effectiveness of our approach, we benchmark it with other state-of-the-art techniques using three data sets of neurological disorders. The results demonstrate the generalisation capabilities of our technique and its suitability for real-world applications. By providing transparent insights into the decision-making process, ART-Explain can improve end-user trust and enable a better understanding of classification outcomes in the detection of neurological diseases.

INDEX TERMS Alzheimer's disease, brain tumor, deep learning, epilepsy, explainable artificial intelligence, feature extraction.

I. INTRODUCTION

Since the conception of Deep Learning, there has been an ever-growing consensus of concern among academia and industry professionals about whether these systems can be trusted. The lack of adaptation can be attributed to the opaque nature of deep learning. The importance of this issue is particularly salient in the medical section, where accountability and transparency of the decision-making process are paramount [1]. A major rationale for requiring such a high level of transparency comes from the adverse consequences that can occur in a patient when provided with an incorrect diagnosis [2], potentially leading to unwanted or excessive disease progression and adverse treatment outcomes.

One of the most recognised contributors to the increase in mortality and disability of humans worldwide in the medical industry is neurological disorders [3]. When considering deaths alone from epilepsy, brain cancer, and dementia in Australia, it was estimated that in 2019, 1,100 deaths were attributed to epilepsy [4], in 2022, 1,525 deaths related to brain cancer [5]. In 2020 14,500 deaths were

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

associated with dementia [6]. Early detection and diagnosis of neurological diseases are crucial to treating and managing these conditions.

Advancements in Deep Learning (DL) have shown promising results in clinical neuroscience, demonstrating strong potential to assist in diagnosing various intricate neurological disorders that require complex data modalities [7]. The underlying principle for employing DL in healthcare is to lessen the pressure on medical professionals by reducing workloads and minimising the likelihood of human error [8]. Often, DL techniques have demonstrated their ability to outperform humans with disease diagnosis in terms of ease and dependability of results [2]. However, the issue remains of the degree of trust medical professionals are to place in these systems.

The most prominent concern with DL techniques is that they can often learn spurious correlations from artefacts within the data they are trained on that are not related to the underlying clinical decision. [9], [10]. For example, Gautam et al. [11] found that when performing the classification of pneumonia of imbalanced chest X-ray images, they observed that the classifier was using metadata within the X-ray images instead of taking into consideration the location in which the disease is located, that is, in the lungs. Therefore, although DL classifiers may perform well on training and validation datasets, they may still implement incorrect internal logic, whose impact may only be felt after being incorporated into clinical practice. However, if a system can provide some insight into the basis of its decision-making, this can go a long way toward increasing trust in the system for incorporation into clinical practice.

Notably, despite advancements in DL and the increasing agreement amongst researchers and industry professionals about concerns related to relying on opaque methods, researchers are still proposing new techniques that do not provide any insight into how the conclusions are derived. Consequently, to ensure that medical professionals are equipped with the necessary knowledge to make informed decisions when it comes to using deep learning methods for the classification of neurological disorders, we propose a novel eXplainable artificial intelligence (XAI) framework that we have named ART-Explain (Autonomous feature extraction and Relevance Technique for Explainable neurological condition classification). ART-Explain builds on some of the principles of our previous work [8]. More specifically, extracting features autonomously within a convolutional neural network and generating rules to produce an explainable classification. Nevertheless, various limitations of the previous approach have been addressed using significant advancements.

Firstly, the previous technique did not prove any correlation between how the input feature space is related and relevant to the extracted features. At the same time, ART-Explain takes advantage of an XAI heat mapping technique to overcome this issue. Secondly, the previous technique required carefully chosen statistical properties of the extracted features to be defined before rule generation occurs, while the new technique only uses the raw extracted features for rule generation. Lastly, instead of simply providing a list of rules, the new technique converts the rules into a novel graphical representation for easier comprehension and application, which we have nicknamed GET (Graphical Explainable Technique). This ensures an end-user can easily comprehend and justify the technique's decisions.

In addition, the proposed ART-Explain has been applied to multiple modalities of neurological data to showcase its suitability to classify a diverse range of neurological conditions. Namely, electroencephalogram (EEG) signals for epilepsy and magnetic resonance imaging (MRI) images for brain tumours and dementia. A comparative analysis is performed with various modern techniques together with the previously proposed technique [8] to reveal the advantage of ART-Explain over existing techniques.

Our contributions are as follows:

- An autonomous end-to-end explainable classification technique for neurological condition classification (ART-Explain).
- A novel graphical representative (GET) to help facilitate the interpretation of classification rules and how they apply to the extracted features.
- A detailed and extensive comparative analysis of the proposed technique with various techniques.

The remainder of the paper is structured as follows. In Section II, we offer a detailed explanation of all components of ART-Explain. In Section III, we detail the experimental designs, including datasets, data preprocessing, and the evaluation metrics used. The experimental results are discussed in Section IV. We provide a detailed discussion of a single positive record classified by the ART-Explain framework, followed by a comparative analysis. Finally, we conclude the paper in Section V with a summary of the entire paper and a quick discussion of our future work.

II. ART-EXPLAIN FRAMEWORK

The proposed ART-Explain framework is designed to achieve end-to-end explainability without compromising performance. It does so by autonomously extracting features from within a Convolutional Neural Network (CNN), generating a heatmap to map the input space to the extracted features, generating rules from the extracted features to use in classification, and presenting a novel graphical explanation using the rules to classify the neurological condition. Figure 1 displays an overview of this framework. For the remainder of this section, we break down the proposed ART-Explain framework into different components (CNN Architecture, Heatmap Explanations, Extracting Features, Rule-Generation, Novel Graphical Explanation), providing a detailed explanation of each element individually while referring back to Figure 1.

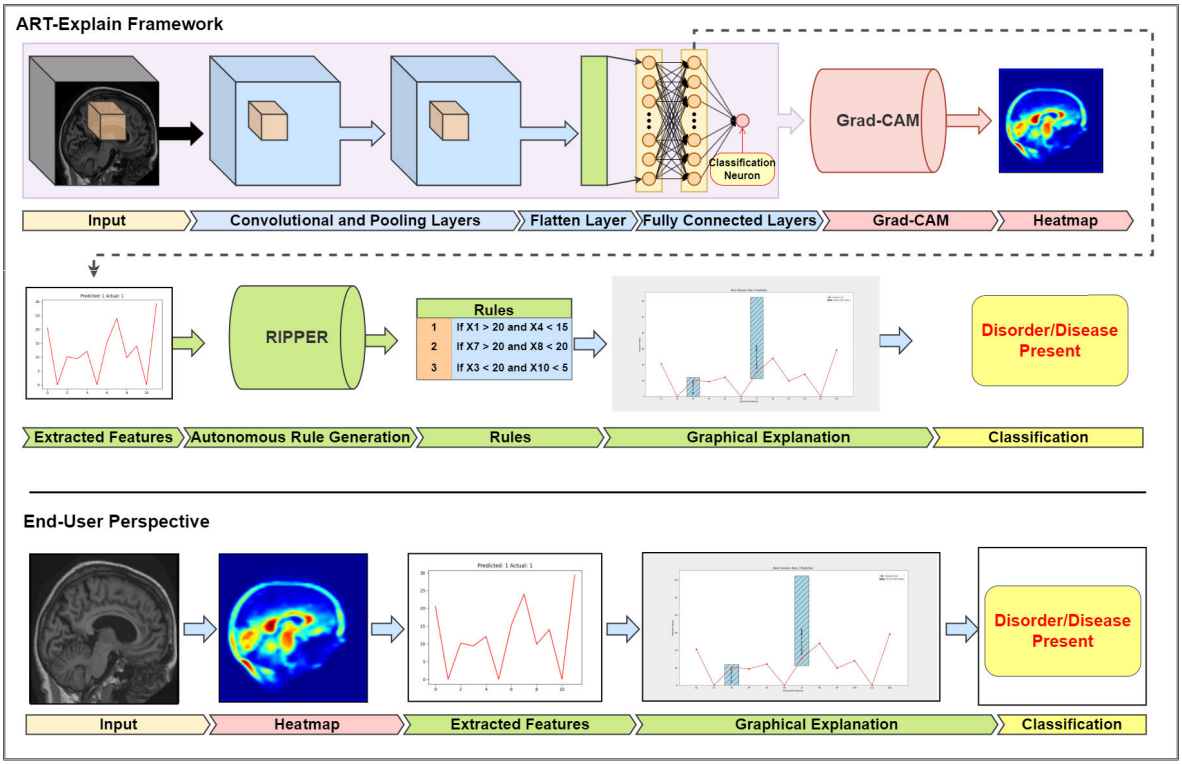

FIGURE 1. ART-Explain framework complete overview.

A. CNN ARCHITECTURE

The foundation of the ART-Explain framework is based on a convolutional neural network (CNN). The choice to use a CNN architecture was two-fold. Firstly, the technique previously proposed by [8] that we are improving uses this architecture. Second, CNNs are the ''most significant'' type of DL network used in literature [12]. Although the overall network architecture varies based on the classification task and input data type (e.g., 2D for image data and 1D for signal channel data), certain commonalities and design choices ensure consistency. More details on each architecture used are displayed in tables in Section III-D: Architectures and Training Details. Each network consists of convolutional layers with ReLU activation functions, followed by a corresponding max pooling layer.

The convolutional layers help extract the most salient features from within the data using the following equation:

$$Conv_i(x) = \text{ReLU}(W_i * x + b_i) \tag{1}$$

For the i^{th} convolutional layer, the weight matrix (often referred to as a kernel) W_i associated with convolutional layer i performs convolution operations over the input data x, which is often taken from the previous layer, and then the i^{th} layer bias term b_i is added to adjust the output operation. The result is passed through the ReLU activation function to produce the output value Z_i. The ReLU activation function ensures non-linearity is introduced into the CNN to ensure it can learn complex relationships within the data. Non-linearity is achieved by taking an output Z_i and producing an activation value A_i by setting the value to 0 if the output value output Z_i is less than 0.

The max-pooling layers, as represented by the equation $Pool_i(x) = \max_{m,n}(x[m, n])$, use the principle of image local correlation to reduce the dimensional of the data [12]. The pooling operation retains the most pertinent features while discarding irrelevant ones. It moves a small window over the input feature space x, selecting the maximum values in each region. The result is a reduced feature map containing only the most prominent features.

After the convolutional and pooling blocks, the data is flattened into a one-dimensional vector before being passed to the fully connected layer. The fully connected layer produces activation values similar to the convolutional layer except for using multiplication instead of convolutional operations. More specifically, using the equation $A_i = \text{ReLU}(\sum_j (w_{ij} * A_j) + b_i)$ the weights are taken from the connection between the current neuron i and neuron from the preceding layer j. To ensure that the explanations are consistent between different data and classification tasks, the fully connected layer of all architectures consists of only one hidden layer with exactly 12 neurons. To maximise explainability, the number of neurons was chosen to restrict the complexity of the revealed graphical explanations.

The classification result is produced using an activation function on the network's last layer.

B. HEATMAP EXPLANATIONS

To produce heatmap explanations that relate the input space to the extracted features, we implement Gradient-weighted Class Activation Mapping (Grad-CAM) [13], [14]. Grad-CAM aims to enhance the transparency of the convolutional neural network's prediction by highlighting the most relevant section (features) within the input space. To achieve this goal, Grad-CAM weights the feature maps produced at the last convolutional layer using the network's gradients to create relevance matrices. The final heatmap explanation is then generated by combining the weighted relevance matrices and stretching it over the network's input. Notably, for our one-dimensional CNN architecture used later, we use the relevance values for each feature and produce a heatmap over a plotted line graph. For more information on Grad-CAM, please refer to this paper [14].

C. FEATURE EXTRACTION

The extraction of features is performed after the network's hyper-parameters are trained. We used the last hidden layer within each network as the extraction point, as this layer contains the most processed information before the output. Furthermore, we ensured a consistent size of 12 neurons in the extraction layer as we believed that any more would make our novel graphical explanations hard to interpret (more details in section II-E).

D. RULE GENERATION

The Repeated Incremental Pruning to Produce Error Reduction algorithm, acronymised as RIPPER, was introduced by Cohen in 1995 [15]. RIPPER was chosen for our rule generation process, as it is known to be great at generating interpretable and compact rule sets in various domains while being immune to biases in imbalanced datasets. It is important to mention that another rule-based classifier can replace RIPPER; however, we chose RIPPER over other techniques for the benefits mentioned previously.

By using a mixture of pre-pruning and post-pruning techniques, RIPPER harnesses the capability to exceed the performance of traditional decision tree algorithms by using their Incremental Reduced Error Pruning (IREP) algorithm. The iterative rule generating of RIPPER can be described as a three-step process involving Growth, Pruning, and Optimization [16], [17]. A high-level overview of the steps is as follows:

1) Growth: In this step, a subset of data must be accurately classified by a rule to progress to the following step. To achieve the desired information gain, conditions are incrementally added to a rule greedily to cover the targeted data, and then it is time to prune.
2) Pruning: This step is performed using a similar approach as decision trees where the information gain criterion is utilised to determine to split the next attribute. More specifically, conditions are greedily removed in decreasing order (opposite to Growth). A rule is immediately pruned if it doesn't reduce entropy in a rule's specificity.
3) Optimization: In this last step, steps 1 and 2 are repeated using a variety of heuristics until there is a decrease in the performance of the classification task.

The rules produced by the RIPPER algorithm construct explanations in an If-Then format. To illustrate, please observe the following example rule taken from our experimental results (in Section IV: Experimental Results): "If $x6$ is less than or equal to 0.27, $x11$ is less than or equal to 0.81, and $x8$ is greater than or equal to 13.79".

E. NOVEL GRAPHICAL EXPLANATION

In this section, we demonstrate the functionality of our proposed Graphical Explanation Technique (GET) using a toy example of a single rule shown in Figure 2. Notably, this figure is not derived from an actual experiment but a fabricated example produced for illustrative purposes. The proposed GET technique graphs the generated rule conditions over-extracted features to showcase the decision boundaries and the logic behind the outcome. That is, if the rule conditions are satisfied for the classification task. To achieve this, the rule conditions are superimposed as boxes over a red line depicting the extracted features graph boxes. In an aesthetically similar design to a box plot, each box's lower and upper lines represent the decision boundaries. When boundary boxes touch the graph's edge, it indicates that the boundary extends indefinitely in that direction. For example, if a boundary box touches the upper edge of the graph, this implies that any value above the lower decision boundary threshold satisfies the condition. In our toy example, as shown in Figure 2, all points of interest, representing extracted features lying on the decision boundaries, fall within the limits of each decision bound. Therefore, this example satisfies the classification rule, indicating a positive disease detection.

III. EXPERIMENTAL DESIGN

A detailed overview of the study's experimental design is provided in this section, covering the datasets used and any required preprocessing, the training and test strategies employed, and the evaluation metrics used to assess the performance of the components of the Art-Explain. All experiments were implemented in the Python programming language.

A. DATASETS

A description of why the three datasets were chosen and an overview.

1) Bonn University Seizure Dataset [18]: The Bonn University Seizure Dataset is a publicly available seizure data set that contains EEG recordings contains 5 folders (Z, O, N, F, and S), each of which corresponds to a different brain state. Each folder contains one hundred single-channel electroencephalograms (EEG)

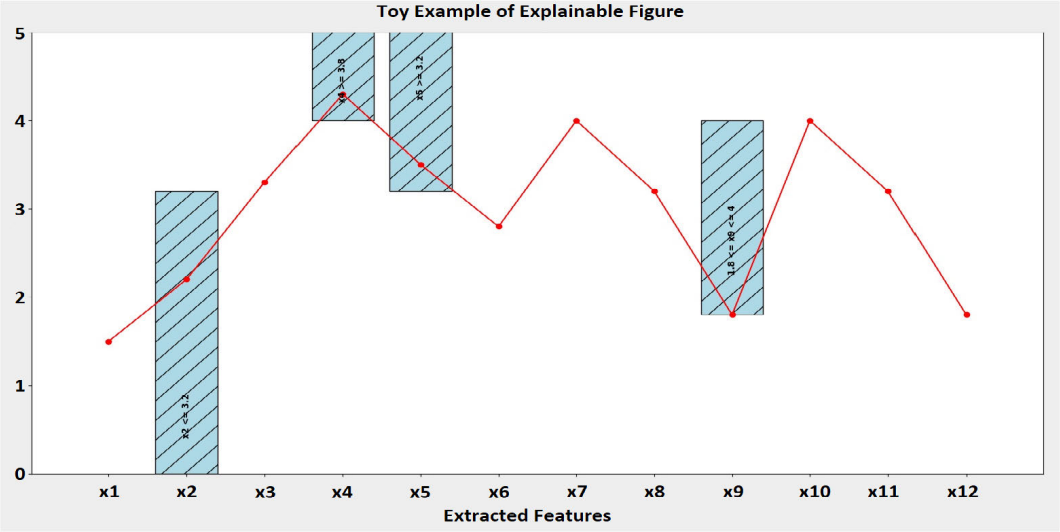

FIGURE 2. Toy example of the Novel Graphical Explanation.

signal files sampled at 173.61 Hz for 23.5 seconds, totalling 500 files(or over 3 hours of recordings). As we are only interested in whether an EEG signal contains the absence and presence of seizure activity, we have combined all the non-seizure-related signals into a single negative class (0) and the seizure as the positive class (1). Additionally, we have divided the recordings into one-second records (discarding the last .5 seconds of each file), randomly shuffled the data to remove any biases, and saved them into a CSV file for each processing. The Bonn University dataset is currently available to download from [19].

2) Kaggle Brain Tumor MRI Dataset [20]: The Brain Tumor MRI dataset is an amalgamation of three popular bench-marking datasets including figshare [21], SARTAJ dataset [22], and Br35H [23]. The dataset contains 7023 human brain MRI images subdivided into four classes: glioma, meningioma, pituitary, and no-tumor. As our research is focused on binary classification, that is, the absence or presence of a tumour, we have joined glioma, meningioma, and pituitary into a single positive class (1) and used the non-tumour data as the negative class (0). The only preprocessing we performed was normalising pixel values between 0 and 1 and resizing the image to 255×55. The Kaggle Brain Tumor MRI Dataset is currently available to download from [24].
3) Alzheimer MRI Preprocessed Dataset [25]: In this study, we use the Alzheimer MRI Preprocessed Dataset from the Kaggle website. The dataset consists of an amalgamation of MRI axial sliced images taken from several homogeneous sources, totalling 6400 MRI images. The data used in this dataset was sourced from various websites, hospitals, and public repositories. The most notable data source is the Alzheimer's Disease Neuroimaging Initiative (ADNI), a well-known source of the highest quality [26]. The dataset's data has been preprocessed and divided into 4 folders corresponding to 4 classes. These include: Class 1 contains 896 Mild Demented images; Class 2 contains 64 Moderate Demented images, Class 3 contains 3200 Non Demented images; and Class 4 contains 2240 very mild demented images. Our research primarily focuses on binary classification, so the dataset has undergone further prepossessing. We have combined all the demented images (Classes 1, 2, and 4) into a single class (1), while the non-demented images are Class 0. The only preprocessing we performed was normalizing pixel values between 0 and 1 and resizing the image to 255×55. The Alzheimer MRI Preprocessed Dataset is currently available to download from [27].

B. DATA PREPARATION

This section discusses any preprocessing performed, including k-fold cross-validation and the 80/20 split used for data preparation. We provide information on how we utilized random sampling to ensure the split accurately represents the overall dataset and discuss the various. Additionally, we provide details on the architectures used and their training details.

1) TRAIN-TEST SPLITTING STRATEGY

In this study, we have chosen to use two of the most common testing strategies, "k-fold cross-validation" and the "80-20 split", to assess the performance of ART-Explain and its generalization capabilities. More specifically, we have chosen these two techniques as they are robust and widely accepted evaluation standards. In the remainder of this subsection, we provide a quick overview of how these techniques operate and the strategies used for each dataset's experiments.

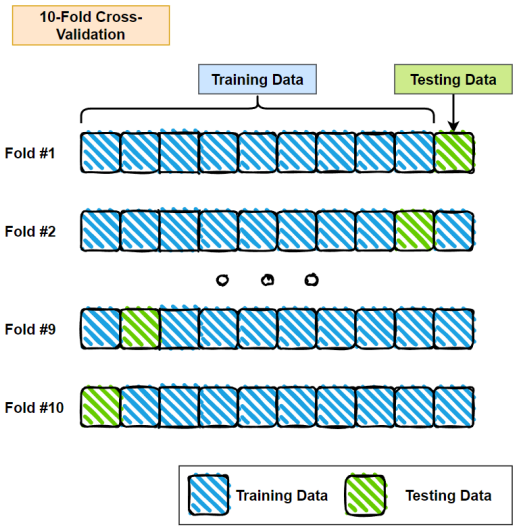

FIGURE 3. 10-Fold Cross-Validation Technique.

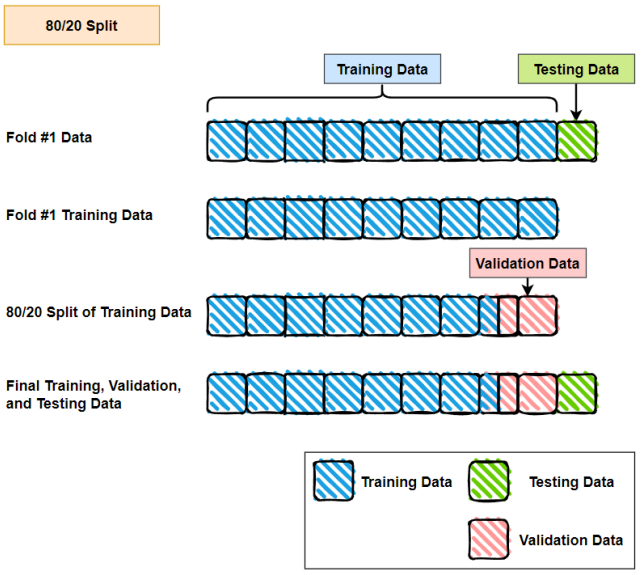

FIGURE 4. 80/20 Data Split.

To perform the 10-fold cross-validation, we split the entire dataset into ten equal pieces. Then, for exactly ten iterations (also known as folds), nine pieces of data are used for the training dataset, and the 10 piece is held out as the testing data set. During each fold, the testing data piece is exchanged to ensure that all pieces of data are eventually used for testing and that no pieces are tested more than once. By repeating this process for all ten pieces, we can robustly estimate ART-Explain's performance across the data. This approach helps assess the model's ability to generalize and perform consistently on unseen data, avoiding biases that could arise from other train-test strategies. We have illustrated this technique in Figure 3.

The 80-20 split is a very simple, easy-to-understand concept yet useful training and testing strategy. As the name suggests, the data is split into 80% training and 20% testing. As our technique uses a Convolutional Neural Network, the 80/20 split is always performed during each fold of any performed 10-fold cross-validation. The training data is subdivided into 80 percent and 20 percent validation data during each fold. To aid in comprehending this strategy, we have provided a visual aid of this strategy in Figure 4.

In the Bonn seizure experimentation, the dataset is split using 10-fold cross-validation to obtain training and testing data to train the model. Additionally, to gain validation data for the CNN model, we subdivided each fold's training data into 80% training and 20% validation data. The subdivision is illustrated in Figure 4. After the network is trained, the best-performing model produces the final results, including the heatmaps, extracted features, rules, etc. The training and testing data are used to train and evaluate the RIPPER technique.

In the brain tumor experimentation, the dataset is split into 80% training data and 20% testing data using the 80/20 strategy. Additionally, the training data is subdivided again using the 80-20 strategy to obtain the final training and validation data. The resulting trained model and training and testing data are then employed to obtain the final results, including the heatmaps, extracted features, rules, etc.

In the Alzheimer's experimentation, the dataset has already been partitioned into training and testing datasets. Due to the use of CNN architectures, we employed the 80/20 strategy on the training dataset to obtain a validation dataset. Like the brain tumor experimentation, the resulting trained model and training and testing data are then employed to obtain the final results, including the heatmaps, extracted features, rules, etc.

C. EVALUATION METRICS

The evaluation metrics used in this research include Accuracy, Sensitivity, and Specificity. We have excluded other commonly used evaluation metrics, such as F1-Score, for simplicity of interpretation and analysis of results. To calculate and understand various evaluation metrics defined in the remainder of this section, we define four key components derived from a confusion matrix. These components are as follows:

- True Positives (TP): The number of positive cases correctly classified as positive by the technique.
- True Negatives (TN): The number of negative cases correctly classified as negative by the technique.
- False Positives (FP): The number of negative cases incorrectly predicted as positive by the technique.
- False Negatives (FN): The number of positive cases incorrectly predicted as negative by the technique.

1) ACCURACY

Accuracy measures the degree to which a technique makes correct predictions. That is, it measures the proportion of correctly classified cases (both positive and negative) against all predictions being made. Accuracy is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

2) SENSITIVITY

Sensitivity (also known as Recall) measures the technique's ability to classify positive cases when only considering positive cases correctly. That is, the calculation only considers records of the positive class. Sensitivity is calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

3) SPECIFICITY

Specificity (also called True Negative Rate) measures the technique's ability to classify negative cases when only considering negative cases correctly. That is, the calculation only considers records of the negative class. Specificity is calculated as follows:

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

D. ARCHITECTURES AND TRAINING DETAILS

Our research utilises two specialised convolutional neural network architectures: one for one-dimensional (1D) data and the other for two-dimensional (2D) data. The 1D is employed on the Bonn dataset, while the 2D is used for the brain tumour and Alzheimer's datasets. Detailed information about these architectures and training processes can be found in Tables 1 and 2. It's important to mention that we implemented an early stopping condition to prevent over-fitting.

TABLE 1. 1-Dimensional Convolutional Neural Network.

Layer Type	Number of Filters	Kernel Size	Activation Function
Conv2D	32	(3, 3)	ReLU
MaxPooling2D	-	(2, 2)	-
Conv2D	64	(3, 3)	ReLU
MaxPooling2D	-	(2, 2)	-
Conv2D	128	(3, 3)	ReLU
MaxPooling2D	-	(2, 2)	-
Flatten	-	-	-
Dense	12	-	ReLU
Dense	2	-	Softmax
Optimizer	Adam (Learning Rate: 0.001)		
Loss Function	Categorical Crossentropy		
Metrics	Accuracy		
Early Stopping	Monitor: val_loss, Patience: 20		
Batch Size	32		
Number of Epochs	300		

TABLE 2. 2-Dimensional Convolutional Neural Network.

Layer Type	Number of Filters	Kernel Size	Activation Function
Conv1D	32	3	ReLU
MaxPooling1D	-	2	-
Conv1D	64	3	ReLU
MaxPooling1D	-	2	-
Conv1D	128	3	ReLU
MaxPooling1D	-	2	-
Flatten	-	-	-
Dense	12	-	ReLU
Dense	1	-	Sigmoid
Optimizer	Adam (Learning Rate: 0.0001)		
Loss Function	Binary Crossentropy		
Metrics	Accuracy, val_accuracy		
Early Stopping	Monitor: val_loss, Patience: 20		
Batch Size	32		
Number of Epochs	200		

IV. EXPERIMENTAL RESULTS

The experimental section discusses the explainability results from each dataset's experiments and concludes with their respective comparative analysis. The explainability results and comparative analysis provide an insightful evaluation of ART-Explain, highlight potential advantages and disadvantages of the state-of-the-art techniques, and discuss any potential for further development.

A. FEATURE EXTRACTION, EXPLAINABLE RESULTS, AND DISCUSSION

In this section, we will delve into the explainable results for each experiment from their corresponding datasets. That is, we will discuss the results from feeding a single positive class record through the trained model until the production of the final explainable graph from our novel GET method. We will commence with the Bonn University Seizure dataset experiments and outcomes and finish with the results from the Alzheimer's dataset. This section is discussed in a manner that assumes the model is trained and the rules have been created. That is, the positive record is unseen data and is the only record needing to be processed.

1) BONN DATASET RESULTS

Beginning with the initial phase, the trained network was presented with a single seizure EEG signal (record) as seen in Figure 5. The record was then passed through the network for processing to transform the EEG signal into a different representation (extracted features).

Before feature extraction, a heatmap showcasing the most relevant features used by the network is produced to showcase which areas of the signal are used during feature extraction. The heatmap generated is displayed in Figure 6.

During the feature extraction phase, the trained model extracted complex, highly processed representations of the EEG signal from within its hidden layer. The 12 neuron activation values within the fully connected layer were extracted as presented in Figure 7.

The previously completed training of the RIPPER rule generation provided the following rules presented in Table 3.

The extracted feature values were then programmatically compared to the rules provided in the previous table to produce the final GET diagram. That is, using the extracted features and the corresponding rule, a graphical explanation is produced using our proposed GET technique to showcase the reason behind the classification outcome. The final rule explanation is displayed in Figure 8.

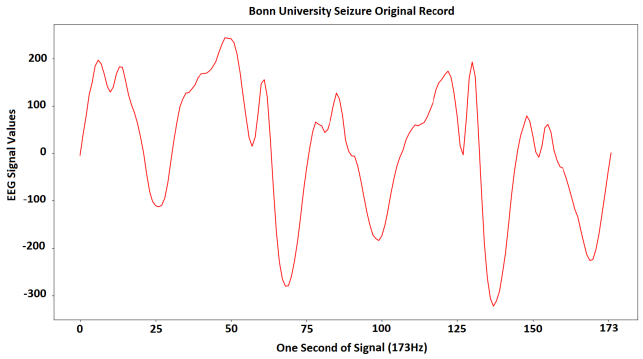

FIGURE 5. Bonn University Seizure Original Record: One-second signal at 173 Hz (or 173 features).

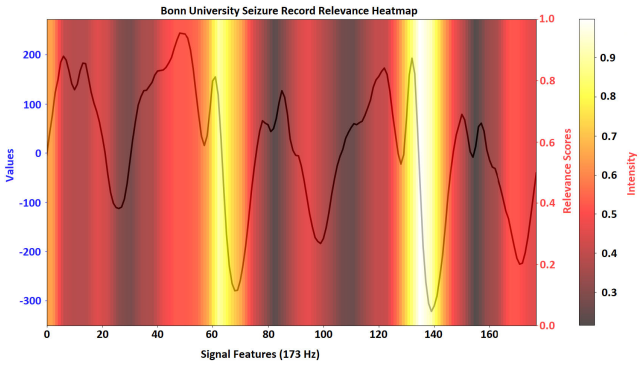

FIGURE 6. Bonn University Seizure Record Relevance Heatmap.

2) BRAIN TUMOR DATASET RESULTS

Beginning with the initial phase, the trained network was presented with a brain tumor MRI image (record) as seen in Figure 9. The record was then passed through the network for processing to transform the raw MRI image into a different representation (extracted features).

Before feature extraction, a heatmap showcasing the most relevant features used by the network is produced to showcase which areas of the MRI are used during feature extraction. The heatmap generated is displayed in Figure 10. Notably, the skull and tumor are considered relevant areas of the input MRI image. This indicates that a technique such as skull stripping may have increased the model's performance.

During the feature extraction phase, the trained model extracted complex, highly processed representations of the brain tumor MRI image from within its hidden layer. The 12 neuron activation values within the fully connected layer were extracted as presented in Figure 11.

The previously completed training of the RIPPER rule generation provided the following rules presented in Table 4.

Like the previous Bonn experiment, the extracted feature values were then programmatically compared to the rules in the previous table to produce the final GET diagram. That is, using the extracted features and the corresponding rule, a graphical explanation is produced using our proposed GET technique to showcase the reason behind the classification outcome. The final rule explanation is displayed in Figure 12.

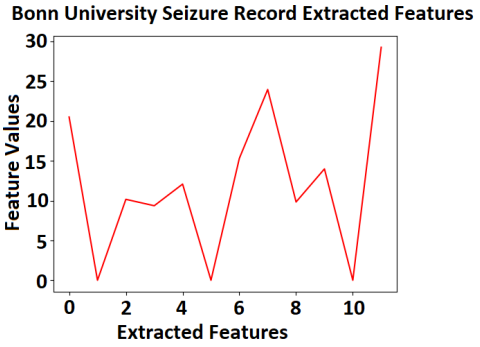

FIGURE 7. Bonn University Seizure Record Extracted Features.

TABLE 3. Bonn University Seizures Rules Table.

Rule Number	Rules
1	If x3 is less than or equal to 11.92, and x7 is greater than or equal to 62.45.
2	If x7 is between 11.19 and 62.45, and x3 is less than or equal to 11.92.
3	If x7 is between 11.19 and 62.45, and x12 is less than or equal to 8.09.
4	If x8 is between 17.83 and 85.68, and x5 is less than or equal to 1.23.
5	If x8 is greater than or equal to 85.68.
6	If x8 is between 17.83 and 85.68, and x4 is less than or equal to 4.93, and x7 is between 11.19 and 62.45.
7	If x8 is between 17.83 and 85.68, and x9 is greater than or equal to 1.28.
8	If x8 is between 17.83 and 85.68, and x10 is between 4.41 and 31.94.

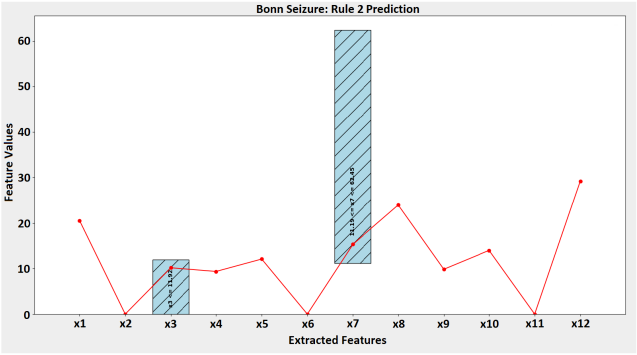

FIGURE 8. Bonn University Dataset Final Explanation Seizure Prediction.

3) ALZHEIMER'S DATASET RESULTS

The Alzheimer's experimental results use the same sequence as the previously discussed brain tumor dataset results. Beginning with the initial phase, the trained network was presented with a positive Alzheimer's disease MRI image (record) as seen in Figure 13. The MRI image was passed through the network for processing to transform the EEG signal into a different representation.

Before feature extraction, a heatmap showcasing the most relevant features used by the network is produced to showcase which areas of the Alzheimer's MRI image are used during

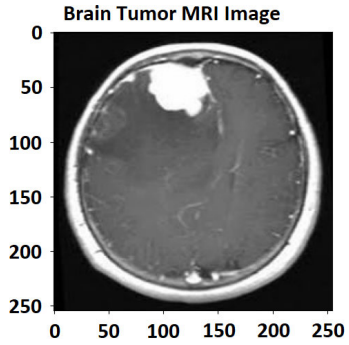

FIGURE 9. Brain Tumor MRI Image.

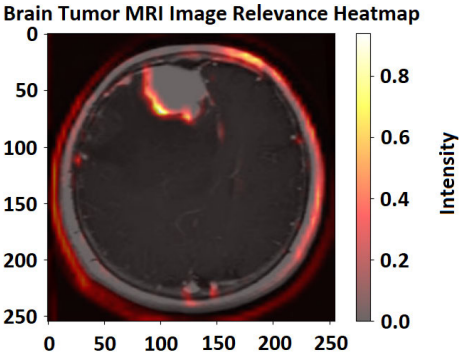

FIGURE 10. Brain Tumor MRI Relevance Heatmap.

feature extraction. The heatmap generated is displayed in Figure 14.

During the feature extraction phase, the trained model extracted complex, highly processed representations of the MRI image from within its hidden layer. The 12 neuron activation values within the fully connected layer were extracted as presented in Figure 15.

The previously completed training of the RIPPER rule generation provided the following rules presented in Table 5.

The extracted feature values were then programmatically compared to the rules provided in the previous table to produce the final GET diagram. That is, using the extracted features and the corresponding rule, a graphical explanation is produced using our proposed GET technique to showcase the reason behind the classification outcome. The final rule explanation is displayed in Figure 16.

B. COMPARATIVE ANALYSIS

This section presents a comparative analysis of the ART-Explain technique with various recently proposed high-quality techniques. Three datasets with varying modalities of data are used to evaluate the techniques' performance. The remainder of this section is divided into subsections for each dataset's comparative analysis and finished with an overall discussion on the outcomes and future direction.

1) BONN UNIVERSITY COMPARATIVE ANALYSIS

To better understand the techniques we are comparing ART-Explain to, we start with a quick discussion on each technique before presenting the comparative analysis of the results.

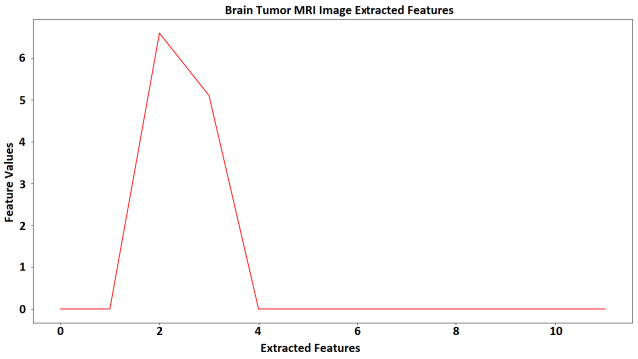

FIGURE 11. Brain Tumor MRI Image Extracted Features.

TABLE 4. Brain Tumor Rules Table.

Rule Number	Rules
1	If x2 is less than or equal to 1.08.
2	If x2 is between 1.08 and 7.41, and x3 is less than or equal to 0.53 and x4 is between 9.55 and 10.5.
3	If x2 is between 1.08 and 7.41, and x4 is between 5.08 and 8.22.
4	If x2 is between 1.08 and 7.41, and x4 is between 8.22 and 9.55.
5	If x4 is between 10.5 and 11.47.
6	If x4 is between 12.51 and 13.88.
5	If x4 is between 11.47 and 12.51.

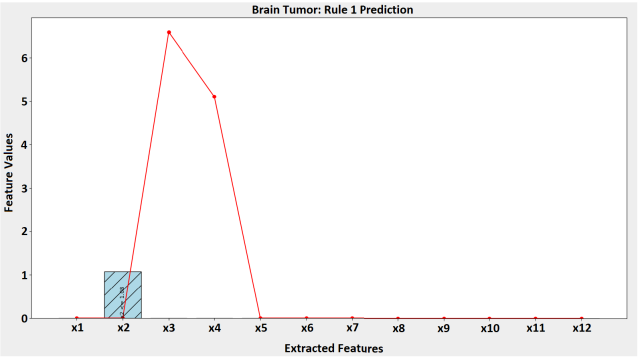

FIGURE 12. Brain Tumor Prediction Final Explanation.

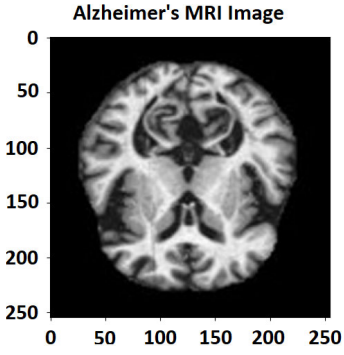

FIGURE 13. Alzheimer's MRI Image.

In 2016, Peker et al. [28] proposed combining features extracted from Dual-Tree Complex Wavelet Transform (DTCWT) applied at various levels and statistical features

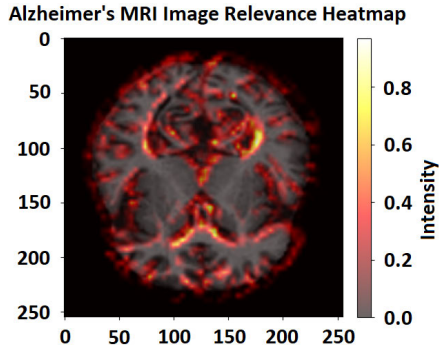

FIGURE 14. Alzheimer's MRI Image Relevance Heatmap.

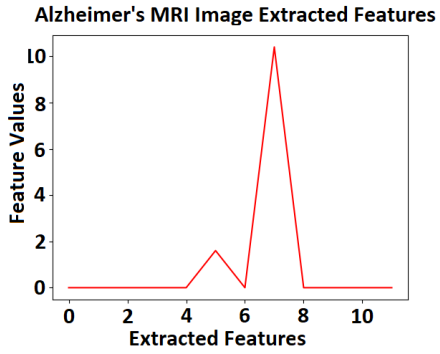

FIGURE 15. Alzheimer's MRI Image Extracted Features.

TABLE 5. Alzheimer's MRI Rules Table.

Rule Number	Rules
1	If x6 is less than or equal to 0.27, x11 is less than or equal to 0.81, and x8 is greater than or equal to 13.79.
2	If x8 is between 11.68 and 13.79.
3	If x8 is between 10.18 and 11.68.
4	If x8 is between 8.54 and 10.18.
5	If x8 is between 5.49 and 8.54, and x11 is less than or equal to 0.81.
6	If x8 is greater than or equal to 13.79.
7	If x8 is between 5.49 and 8.54, and x11 is between 0.81 and 4.08.
8	If x6 is less than or equal to 0.27, x11 is less than or equal to 0.81, and x8 is between 2.7 and 5.49.
9	If x8 is between 5.49 and 8.54, x11 is between 4.08 and 9.42, and x6 is less than or equal to 0.27.
10	If x8 is between 2.7 and 5.49, x6 is between 0.27 and 1.8, and x11 is less than or equal to 0.81.

obtained from the obtained complex-valued feature vector. The extracted features are then fed into a Complex-Valued Artificial Neural Network (CVANN) to produce the final classification. During their experimentation, their proposed DTCWT+CVANN-2 achieved the best performance concerning binary seizure classification, achieving an accuracy of 99.15%, sensitivity of 100%, and specificity of 97.89%.

In the next year, Wan et al. [29] proposed performing seizure detection from extracted time, frequency, time-frequency, and non-linear analysis features. Multiple classifiers performed the classification task. Their technique extracted frequency domain (FFT) and Non-linear analysis

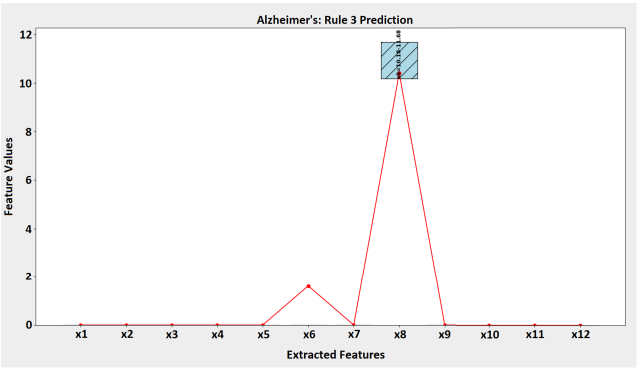

FIGURE 16. Alzheimer's MRI Final Prediction Explanation.

(IMF) features from denoised signals. Additionally, they used wavelet decomposition and reconstruction to extract the time domain, time-frequency domain, and non-linear analysis features (Entropy) from the sub-bands of the EEG signal. Then, using a combination of component analysis (PCA) and analysis of variance (ANOVA) to eliminate non-useful features, the resulting features were then passed to multiple different classifiers to detect seizures. The best-performing classifier was the Support Vector Machine (SVM), with an accuracy of 99.25%, sensitivity of 97.98%, and specificity of 99.56%.

In 2018, Hussain et al. [30] proposed extracting statistical and morphological features from the raw EEG signals before being classified by either a Logistic Regressive (LR) or Decision Tree (DT) classifier. Logistic Regressive achieved an accuracy of 99.48%, a sensitivity of 99.38%, and a specificity of 99.58%, while the DT attained an accuracy of 99.12%, sensitivity of 99.12%, and specificity of 99.12%. In the same year, Mursalin et al. [31] proposed extracting time and frequency domain features and Entropy-base features before being passed to a Random Forrest (RF) classifier. They used DWT from the EEG signal and time-domain and entropy features from the raw signal; they combined these and selected features based on their novel approach, Improved Correlation-based Feature Selection (ICFS). The RF produced the final seizure classification task, achieving an accuracy of 97.4%, a sensitivity of 97.4%, and a specificity of 97.5%. Also, in the same year, Acharya et al. [32] proposed using a 13-layer deep convolutional neural network for seizure detection. The idea is to use the original EEG signal only using Z-score normalization with a mean of zero and a standard deviation of 1. They achieved an accuracy of 88.67%, sensitivity of 95.00%, and specificity of 90.00%.

In 2019, Sriraam et al. [33] proposed using discrete wavelet transformation on the raw EEG signal, extracting 11 features from the wavelet coefficients in each level. Classifying seizures is then performed by a support vector machine. The best results were achieved using a quadratic kernel, achieving an accuracy of 99.18%, sensitivity of 98.73%, and specificity of 99.62%. During the same calendar year, Zeng et al. [34] proposed classifying seizures by

TABLE 6. Comparative analysis of proposed technique with state-of-the-art techniques for UCI ML.

Technique	Accuracy [%]	Sensitivity [%]	Specificity [%]
DTCWT + CVANN-2 [28]	99.15%	100%	97.89%
DWT, multi-domain feature extraction, and nonlinear analysis [29]	99.25%	97.98%	99.56%
Stat and morph features + DT [30]	99.12%	99.12	99.12%
Stat and morph features + LR [30]	99.48%	99.38%	99.58%
Time, Freq, and Entropy features + ICFS + RF [31]	97.40%	97.4%	97.5%
13-layer deep CNN [32]	88.67%	95.00%	90.00%
DWT + quadratic kernel SVM [33]	99.18%	98.90%	99.62
PRC + ITD + DWT + PSR + NN [34]	98.00%	94.00%	99.00%
3100 Extracted Features from CNN + Stat Properties + DT [8]	98.17%	96.29%	98.49%
ART-Explain	99.30	97.84	99.67

deconstructing raw EEG into multiple Proper Rotation Components (PRCs) using time-scale decomposition (ITD). Different frequency bands of the PRCs are then extracted using discrete wavelet transform (DWT). Euclidean distances and non-linear dynamics are then computed using Phase Space Reconstruction (PSR). The final feature vectors are then used within a neural network (NN) to classify seizures, achieving an accuracy of 98%, a sensitivity of 94.00%, and a specificity of 99.00%.

In 2021, in our previous work by Woodbright, Verma, and Haidar, we proposed using a feature extraction rule-based seizure detection technique. The technique extracts 3200 features from within one of the hidden layers of a custom-made one-dimensional neural network. Then using carefully chosen statistical properties of the extracted features, we use a decision tree classifier to generate the rules used for seizure classification. This technique achieved an accuracy of 98.65%, a sensitivity of 96.29%, and a specificity of 99.25%.

The results of our proposed ART-Explain and the comparative techniques are presented in Table 6. Notably, as can be observed from our results, ART-Explain reaches a high level of performance while providing a higher accuracy overall comparative methods except for [30]. However, Hussain et al. [30]'s technique only outperforms our technique by a small margin; however, their technique requires the careful selection of statistical or morphological features that require expert knowledge. The results indicate that ART-Explain is a highly suitable replacement for the comparative methods and achieves state-of-the-art performance in seizure detection.

2) BRAIN TUMOR COMPARATIVE ANALYSIS

In this comparative analysis, we assess the performance of the proposed ART-Explain by evaluating its performance against well-known state-of-the-art deep learning techniques. The selected comparative techniques include VGG-16 [35], VGG-19 [35], ResNet-50 [36], DenseNet121 [37], and Xception [38]. Note, that the comparative techniques were trained in the same manner as ART-Explain except for using a single hidden layer with 1024 neurons. Each of the deep learning techniques was selected as they have demonstrated exceptional results for various applications in research. Additionally, these techniques are widely recognized and respected in deep learning.

TABLE 7. Comparative analysis of proposed technique with state-of-the-art techniques for the brain tumor dataset.

Technique	Accuracy [%]	Sensitivity [%]	Specificity [%]
VGG16 [35]	97.18%	95.92%	100.0%
VGG19 [35]	98.09%	99.12%	95.80%
ResNet50 [36]	97.48%	99.78%	92.35%
DenseNet121 [37]	97.48%	96.69%	99.26%
Xception [38]	97.18%	98.90%	93.34%
ART-Explain	98.32%	98.34%	98.27%

TABLE 8. Comparative analysis of proposed technique with state-of-the-art techniques for the Alzheimer dataset.

Technique	Accuracy [%]	Sensitivity [%]	Specificity [%]
VGG16	81.48%	64.30%	98.30%
VGG19	88.05%	89.57%	86.55%
ResNet50	85.94%	89.26%	82.69%
DenseNet121	73.67%	48.50%	98.30%
Xception	65.16%	44.24%	85.63%
ART-Explain	94.61%	92.42%	96.75%

The results of our comparative analysis can be observed in Table 7. Notably, despite a slight margin, ART-Explain outperforms all comparative methods in terms of accuracy. These results indicate that ART-Explain is a suitable technique for brain tumor classification as it outperforms the state-of-the-art deep learning approaches at the same task.

3) ALZHEIMER'S COMPARATIVE ANALYSIS

In this comparative analysis, we assess the performance of the proposed ART-Explain by evaluating its performance against well-known state-of-the-art deep learning techniques. The selected relative techniques are the same in the brain tumour comparative analysis in Section IV-B2, trained similarly. Each of the deep learning techniques was selected as they have demonstrated exceptional results for various applications in research. Additionally, these techniques are widely recognised and respected in deep learning.

The results of our comparative analysis can be observed in Table 8. Notably, ART-Explain outperforms all techniques by a decent margin in accuracy. These results indicate that ART-Explain is a suitable technique for Alzheimer's

disease classification as it outperforms the state-of-the-art deep learning approaches at the same task.

V. CONCLUSION AND FUTURE DIRECTION

This paper presented the proposed novel Autonomous Relevance Technique for an Explainable neurological disease prediction framework, ART-Explain. The proposed technique was benchmarked against state-of-the-art and well-respected deep learning techniques on diverse categories of neurological data and conditions, attaining exceptional results. The results of extensive experimentations indicate that ART-Explain could enhance end-user trust and enable a better understanding of the classification outcomes in neurological condition detection.

In the future, we would like to explore the production of a novel XAI heatmapping technique to improve the spatial and temporal resolution of the resulting explainable heatmap. Furthermore, we would like to adapt the technique for multi-modal deep learning tasks involving Parkinson's Disease.

VI. STATEMENT

During the preparation of this work the author(s) used GrammarlyGo and/or ChatGPT in order to enhance the representation of information and improve the quality of the content. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

REFERENCES

[1] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.

[2] J. Chaki and M. Woźniak, "Deep learning for neurodegenerative disorder (2016 to 2022): A systematic review," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104223, doi: 10.1016/j.bspc.2022.104223.

[3] V. L. Feigin, "Global, regional, and national burden of neurological disorders, 1990–2016: A systematic analysis for the global burden of disease study 2016," *Lancet Neurol.*, vol. 18, no. 5, pp. 459–480, May 2019, doi: 10.1016/s1474-4422(18)30499-x.

[4] Australian Institute of Health Welfare. *Epilepsy in Australia*. AIHW, Australian Government. Accessed: Mar. 7, 2024. [Online]. Available: https://www.aihw.gov.au/reports/chronic-disease/epilepsy-in-australia/contents/deaths-due-to-epilepsy

[5] Cancer Australia. *Brain Cancer in Australia Statistics*. Cancer Australia, Australian Government. Accessed: Mar. 7, 2024. [Online]. Available: https://www.canceraustralia.gov.au/cancer-types/brain-cancer/statistics

[6] Australian Institute of Health Welfare. *Dementia in Australia*. Australian Institute of Health and Welfare. Accessed: Mar. 7, 2024. [Online]. Available: https://www.aihw.gov.au/reports/dementia/dementia-in-aus

[7] A. A.-A. Valliani, D. Ranti, and E. K. Oermann, "Deep learning and neurology: A systematic review," *Neurol. Therapy*, vol. 8, no. 2, pp. 351–365, Aug. 2019, doi: 10.1007/s40120-019-00153-8.

[8] M. Woodbright, B. Verma, and A. Haidar, "Autonomous deep feature extraction based method for epileptic EEG brain seizure classification," *Neurocomputing*, vol. 444, pp. 30–37, Jul. 2021, doi: 10.1016/j.neucom.2021.02.052.

[9] F. Pahde, M. Dreyer, W. Samek, and S. Lapuschkin, "Reveal to revise: An explainable AI life cycle for iterative bias correction of deep models," 2023, *arXiv:2303.12641*.

[10] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?" in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[11] S. Gautam, M. M.-C. Höhne, S. Hansen, R. Jenssen, and M. Kampffmeyer, "Demonstrating the risk of imbalanced datasets in chest X-ray image-based diagnostics (don't short) by prototypical relevance propagation," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9761651

[12] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[13] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626. [Online]. Available: https://ieeexplore.ieee.org/document/8237336

[14] R. R Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?" 2016, *arXiv:1611.07450*.

[15] W. W. Cohen, "Fast effective rule induction," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 115–123. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/B9781558603776500232

[16] S. Sonvane. *The RIPPER Algorithm*. Medium. Accessed: Mar. 7, 2024. [Online]. Available: https://medium.com/swlh/the-ripper-algorithm-a5eebbe3661d

[17] I. Moscovitz. *How to Perform Explainable Machine Learning Classification–Without Any Trees*. Medium. Accessed: Mar. 7, 2024. [Online]. Available: https://towardsdatascience.com/how-to-perform-explainable-machine-learning-classification-without-any-trees-873db4192c68

[18] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, Nov. 2001, Art. no. 061907.

[19] *EEG Data Download*. UKB. Accessed: Mar. 7, 2024. [Online]. Available: https://www.ukbonn.de/epileptologie/arbeitsgruppen/ag-lehnertz-neurophysik/downloads/

[20] M. Nickparvar. *Brain Tumor MRI Dataset*. Kaggle. Accessed: Mar. 7, 2024. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

[21] J. Cheng. *Brain Tumor Dataset*. Figshare. Accessed: Mar. 7, 2024. [Online]. Available: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427

[22] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan. *Brain Tumor Classification (MRI)*. Kaggle. Accessed: Mar. 7, 2024. [Online]. Available: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri

[23] A. Hamada. *Br35H: Brain Tumor Detection 2020*. Kaggle. Accessed: Mar. 7, 2024. [Online]. Available: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection?select=no

[24] *Brain Tumor MRI Dataset*. Kaggle. Accessed: Mar. 7, 2024. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/data

[25] S. Kumar and S. Shastri. *Alzheimer MRI Preprocessed Dataset*. Kaggle. Accessed: Mar. 7, 2024. [Online]. Available: https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset

[26] *Alzheimer's Disease Neuroimaging Initiative*. ADNI. Accessed: Mar. 7, 2024. [Online]. Available: https://adni.loni.usc.edu/

[27] *Alzheimer MRI Preprocessed Dataset*. Kaggle. Accessed: Mar. 7, 2024. [Online]. Available: https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset

[28] M. Peker, B. Sen, and D. Delen, "A novel method for automated diagnosis of epilepsy using complex-valued classifiers," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 108–118, Jan. 2016, doi: 10.1109/JBHI.2014.2387795.

[29] L. Wang, W. Xue, Y. Li, M. Luo, J. Huang, W. Cui, and C. Huang, "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, p. 222, May 2017, doi: 10.3390/e19060222.

[30] M. S. Hussain, D. M. Sarfraz, and S. Rukhsar, "Epileptic seizure detection using temporal based measures in EEG signal," in *Proc. 3rd Int. Conf. Commun. Electron. Syst. (ICCES)*, Oct. 2018, pp. 743–748. [Online]. Available: https://ieeexplore.ieee.org/document/8723966

[31] M. Mursalin, Y. Zhang, Y. Chen, and N. V. Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier," *Neurocomputing*, vol. 241, pp. 204–214, Jun. 2017, doi: 10.1016/j.neucom.2017.02.053.

[32] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018, doi: 10.1016/j.compbiomed.2017.09.017.

[33] N. Sriraam, S. Raghu, Y. Temel, S. Vasudevarao, A. S. Hedge, J. Mahendra, and P. L. Kubben, "Automated detection of epileptic seizures using DWT based features and SVM classifier," in *Proc. 2nd Int. Conf. Signal Process. Commun. (ICSPC)*, Mar. 2019, pp. 263–266. [Online]. Available: https://ieeexplore.ieee.org/document/8976611

[34] W. Zeng, M. Li, C. Yuan, Q. Wang, F. Liu, and Y. Wang, "Identification of epileptic seizures in EEG signals using time-scale decomposition (ITD), discrete wavelet transform (DWT), phase space reconstruction (PSR) and neural networks," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 3059–3088, Aug. 2019, doi: 10.1007/s10462-019-09755-y.

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

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[37] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html

[38] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf

MITCHELL D. WOODBRIGHT received the bachelor's degree (Hons.) in computing from Charles Sturt University, Australia, in 2019. He is currently pursuing the Ph.D. degree with the School of Engineering and Technology, Central Queensland University, Australia. His current research interests include artificial intelligence, deep learning, explainable artificial intelligence, and machine learning, with a strong interest in medical applications.

AHSAN MORSHED (Member, IEEE) received the Ph.D. degree in computer science, in 2010. He was a Researcher with the FAO (UN), CSIRO, RMIT, and the Swinburne University of Technology. He is currently a Lecturer with Central Queensland University. He has published more than 50 scientific articles; his Google citation is 1380. His research interests include data science, machine learning, deep learning, AI, and data integration. He is a member of ACS.

MATTHEW BROWNE received the Ph.D. degree in computational psychophysiology, in 2002. He was a Researcher with the Fraunhofer Gesellschaft, Japan, CSIRO, and Griffith University. He is currently a Professor with Central Queensland University, specializing in applied statistics, mathematical psychology, and addiction.

BIPLOB RAY (Senior Member, IEEE) received the Ph.D. degree in information technology from Deakin University, Melbourne, Australia, in 2015. He is currently a Senior Lecturer with Central Queensland University, with a background mix of research, academic, and industry experience. His research interests include secure communication protocols of cyber-physical systems, artificial intelligence (AI), and the Internet of Things (IoT).

STEVEN MOORE received the Ph.D. degree in biomedical engineering, in 1996. He was a NASA Researcher with the Human Aerospace Laboratory, The Mount Sinai Hospital, New York, NY, USA, before joining Central Queensland University, Australia, as a Professor.