Research Article

Using CNN Saliency Maps and EEG Modulation Spectra for Improved and More Interpretable Machine Learning-Based Alzheimer's Disease Diagnosis

Marilia Lopes [ID],[1] **Raymundo Cassani** [ID],[2] and **Tiago H. Falk** [ID][1]

[1]*Institute National de la Recherche Scientifique (INRS-EMT), University of Quebec, Montreal, Canada*
[2]*McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada*

Correspondence should be addressed to Marilia Lopes; marilia.soares@inrs.ca

Biomarkers based on resting-state electroencephalography (EEG) signals have emerged as a promising tool in the study of Alzheimer's disease (AD). Recently, a state-of-the-art biomarker was found based on visual inspection of power modulation spectrograms where three "patches" or regions from the modulation spectrogram were proposed and used for AD diagnostics. Here, we propose the use of deep neural networks, in particular convolutional neural networks (CNNs) combined with saliency maps, trained on power modulation spectrogram inputs to find optimal patches in a data-driven manner. Experiments are conducted on EEG data collected from fifty-four participants, including 20 healthy controls, 19 patients with mild AD, and 15 moderate-to-severe AD patients. Five classification tasks are explored, including the three-class problem, early-stage detection (control vs. mild-AD), and severity level detection (mild vs. moderate-to-severe). Experimental results show the proposed biomarkers outperform the state-of-the-art benchmark across all five tasks, as well as finding complementary modulation spectrogram regions not previously seen via visual inspection. Lastly, experiments are conducted on the proposed biomarkers to test their sensitivity to age, as this is a known confound in AD characterization. Across all five tasks, none of the proposed biomarkers showed a significant relationship with age, thus further highlighting their usefulness for automated AD diagnostics.

1. Introduction

Alzheimer's disease (AD) is a degenerative brain disease and the most common cause of dementia [1, 2]. AD progression leads to the loss of cognitive (e.g., memory, reasoning, and communication) and behavioral functions that will ultimately interfere with the individual's daily life. Diagnosing AD can be challenging, as the disease can initiate patho-physiological processes 20 years before any clinical symptoms appear [3, 4]. As such, improving early diagnostics has become a fundamental element in AD research and therapy [5], especially since there are currently no cures for AD [6–8]. While a definite diagnosis for AD can only be achieved through a postmortem structural examination of the brain, clinical diagnosis currently relies on the use medical history and exams, such as the mini-mental state examination (MMSE) [9] and the clinical dementia rating (CDR) [10]. Moreover, a handful of biomarkers have also been incorporated into the clinical diagnostic process, including structural neuroimaging (e.g., magnetic resonance imaging (MRI)-based measures of hippocampal volumes) [11, 12], blood and urine samples [13], cerebral spinal fluids (e.g., tau and beta-amyloid levels) [14], and genetic risk profiling [15]. More recently, there has also been a push to incorporate biomarkers extracted from electroencephalograms (EEG) into the diagnostic process [16].

Indeed, EEG has proven to be a useful tool in the study of AD, with several advantages over other neuroimaging modalities, including noninvasiveness, lower cost, the possibility to detect early synaptic dysfunction prior to symptoms arising, and the possibility to more easily track the course of the disease [17]. EEGs record from the scalp the

electrical field produced by the interaction of neurons. Most of the published works have relied on the analysis of resting-state EEG (rsEEG) in either eyes-open or eyes-closed conditions, thus making the data collection procedure more comfortable for elderly patients [18]. Over the years, several changes in the EEG have been reported as a function of AD, namely, slowing of the EEG, reduction in signal complexity, perturbations in interelectrode synchrony/coherence, and a neuromodulatory deficit [19], resulting from the reduction of neurotransmitters due to damage to neuro-pathways by the disease. More recently, machine learning tools applied to conventional EEG spectral features have started to show promising results for AD diagnosis (e.g., see [20–22]); in some cases, accuracy around 90% has been reported.

In regards to features, typically rsEEG-based AD studies extract diagnostic features from the five traditional EEG frequency subbands, namely, delta ($\delta = 0.1–4\text{Hz}$), theta ($\theta = 4–8\text{Hz}$), alpha ($\alpha = 8–12\text{Hz}$), beta ($\beta = 12–30\text{Hz}$), and gamma ($\gamma > 30\text{Hz}$). The work in references [23, 24], in turn, proposed the measurement of the spectrotemporal dynamics of each of these subbands and showed that improved diagnostic accuracy could be achieved. The resultant 2-dimensional (frequency vs. frequency) power modulation spectrogram was shown to not only improve accuracy but to also be more robust to artifacts commonly observed with EEGs [25]. The use of these conventional frequency bands, however, may not be optimal for AD analysis, as they have been defined based on visual inspection of EEG signals in healthy subjects [26]. In fact, the work in references [27, 28] showed that nontraditional bands were better for AD diagnostics. Building on these insights, the work in reference [29] used visual inspection to find optimal power modulation spectrogram regions or "patches" that improved diagnostic accuracy relative to conventional measures widely used in AD characterization.

In this article, we build on the work of [29] and propose the use of a data-driven method in order to find the optimal modulation spectrogram patches for AD diagnostics, whereas in [29], visual inspection was used. Visual inspection can lead to the loss of discriminatory information and potential biases, as well as reduce the interpretation of the results. In particular, we propose the use of saliency maps obtained from a convolutional neural network (CNN) trained to detect AD. Saliency maps provide insights into which regions of the input image the CNN is focusing on to make its decisions [30] and could improve the interpretability of the obtained results. While the use of saliency maps has been explored for AD classification based on MRIs [31, 32], it has yet to be explored for EEG data. Given the modulation spectrogram 2-dimensional representation and its improved discrimination relative to the conventional spectrogram, it can be a prime candidate for a saliency map-based biomarker. Experiments described herein show the usefulness of the proposed method not only in detecting AD but also in predicting AD severity level. Finally, the measure is shown to not be affected by participant age, a common confounding factor in AD studies.

The remainder of this paper is organized as follows: Section 2 describes the materials and methods used in our experiments. Section 3 presents the experimental results and discusses them in light of the existing literature, including study limitations. Lastly, conclusions are presented in Section.

2. Materials and Methods

2.1. Participants. We rely on EEG data collected from fifty-four participants recruited from the Behavioral and Cognitive Neurology Unit of the Department of Neurology and the Reference Center for Cognitive Disorders at the Hospital das Clinicas in São Paulo, Brazil. AD diagnosis and identification of severity level (mild-AD or moderate-to-severe AD) were performed by experienced neurologists according to NINCDS-ADRDA criteria [33] and classified based on the Brazilian version of the MMSE [34]. Ethics approval was obtained from the Research Ethics Office, and all participants consented to participate in the study. The data have been divided into three groups: (i) "N" consists of 20 healthy elderly controls, (ii) "AD1" corresponds to 19 mild-AD patients, and (iii) "AD2" to the 15 patients diagnosed with moderate-to-severe AD. In experiments where all AD patients are combined (i.e., AD1 + AD2), this group will be termed "AD". Table 1 shows participant demographic details, including MMSE scores.

Inclusion criteria for the N group included MMSE score ≥ 25 and CDR score = 0, as well as no indication of functional cognitive decline. The AD1 group inclusion criteria, in turn, included MMSE ≤ 24 and 0.5 ≤ CDR ≤ 1; whereas for the AD2 group, they included an MMSE ≤ 20 and CDR score = 2. An additional criterion used for the two AD groups was the presence of functional and cognitive decline over the previous 12 months based on detailed interviews with knowledgeable informants. Exclusion criteria included diabetes mellitus, kidney disease, thyroid disease, alcoholism, liver disease, lung disease, or vitamin B12 deficiency, as these conditions can also cause cognitive decline.

2.2. EEG Acquisition and Preprocessing. For the acquisition of EEG signals, twenty channels were used following the 10–20 international mounting system. EEG signals were recorded with 12 bit resolution and 200 Hz sampling frequency using BrainTech 3.0 instrumentation (EMSA Equipamentos Médicos INC., Brazil). In addition, electrode impedance was kept below $10\,k\Omega$ and attached bi-auricular (A1 and A2) electrodes were used as reference. A resting-state eyes-closed (rsEEG) protocol was followed, and data were recorded for at least eight minutes for each participant.

Based on insights from reference [25], raw EEG was preprocessed with a zero-phase FIR bandpass filter with a bandwidth 0.5–45 Hz, followed by processing with the wavelet-enhanced independent components analysis (wICA) step to remove eye movement and/or muscle artifacts [35, 36].

2.3. Power Modulation Spectrogram. The 2-dimensional power modulation spectrogram representation has been proposed to characterize spectrotemporal changes in the

TABLE 1: Participant demographic details.

Group identifiers	Subjects (female)	Age (years)	Education (years)	MMSE
N	20 (9)	68.0 ± 8.6	10.1 ± 5.5	28.5 ± 1.7
AD1	19 (11)	74.1 ± 5.5	5.6 ± 2.8	19.4 ± 5.3
AD2	15 (9)	75.0 ± 11.8	4.1 ± 3.8	12.8 ± 5.0

EEG due to AD (e.g., see [37]). The signal processing steps involved in the computation of the representation can be seen in Figure 1. First, the EEG time series $x(t)$ is processed by a time-frequency mapping to generate the time-frequency representation $X(t, f)$. This mapping can be a short-time Fourier transform, for example, or a wavelet transform. In order to measure the temporal dynamics of each spectral bin, a second transformation is done on the time-frequency signal, now across the time dimension. This is achieved with a Fourier transform (FT) of the instantaneous power of each spectral bin, resulting in the power modulation spectrogram $X(f, f_{\text{mod}})$, i.e.,

$$X(f, f_{\text{mod}}) = \mathcal{F}_t \{ |X(t, f)|^2 \}, \tag{1}$$

where $\mathcal{F}_t\{\cdot\}$ indicates the Fourier transform over the time dimension and f_{mod} indicates the modulation frequency dimension. The final result is a frequency-modulation frequency spectral representation that describes the second-order periodicities that would not be present in conventional spectral or time-frequency representations [38, 39].

In its original version, frequency bins in this 2-dimensional representation were grouped across the conventional and modulation frequency axes into the five traditional subbands, namely, delta, theta, alpha, beta, and gamma, and each frequency-modulation frequency bin was used as a feature for AD detection (e.g., see [23, 37]). More recently, the work in reference [29] showed that the use of the traditional bands was not optimal for the task at hand, and through visual inspection, three new regions were defined and proposed, termed "patches," as shown in Figure 2. The modulation energy computed from these three patches, as well as their ratios, was proposed as new features for AD diagnosis and severity level prediction. Experimental results showed their superiority to the traditional band-grouping strategy, but suggested that some of the patches could be a result of normal aging. In this work, we build on these patches and propose a data-driven manner to explore if optimal patches can be found that are indicative of AD and not normal aging. Experiments rely on a 45×45 power modulation spectrogram, where each bin corresponds to 1 Hz resolution.

2.4. Convolutional Neural Networks. Deep learning has emerged as a very powerful pattern analysis tool over the last decade. Increases in computational power have allowed for artificial deep neural networks (DNNs) with multiple hidden layers and billions of parameters to be trained within reasonable time frames. This increase in computational capacity has resulted in the emergence of a number of new deep neural network architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and

recursive neural networks (RvNNs), to name a few [40]. Here, we rely on CNNs and saliency maps to extract new biomarkers of AD from EEG modulation spectrograms; hence, a brief description of CNNs is given for the sake of completeness; more details can be found in [41–44].

While feed-forward DNNs multiply the inputs by optimized weights obtained during the training, CNNs include layers that perform convolutions, i.e., the dot product of the convolution kernel with the layer's input matrix. The convolution kernel slides along the input matrix for the layer, thus generating a new feature map that contributes to the input of the subsequent layer. This is followed by other layers such as pooling layers, fully-connected layers, and normalization layers [45]. CNNs have resulted in state-of-the-art image recognition performance [46], as each convolution layer extracts specific features from the image, such as vertical and horizontal lines, color and shape, contrast, exposure variations, or image borders, to name a few. Using sequential convolution layers, each new feature map builds on the properties captured by the previous map. Depending on the number of layers used and the type of data used during training, CNNs may even learn features that take care of preprocessing, detection, and recognition, thus enabling end-to-end systems. When used for biomarker development, their use may also result in regions that could be more robust to EEG artifacts.

While many deep learning models are regarded as "black boxes," providing limited insights on what parts of the input image are being used for discrimination, so-called saliency maps have been proposed to overcome this limitation. This method measures the spatial support of a particular class in each image via a heatmap. Saliency maps have been used for region-of-interest extraction [47], medical imaging [48, 49], robot vision [50], and audio-visual integration [51, 52], in addition to AD diagnosis based on MRI [32]. Saliency maps are obtained by computing the gradient of the output category in relation to the input image. In this way, the impact of how the value of the output category changes in relation to a small change in the pixels of the input image is observed. The highlighted regions in the resulting map indicate that they are important areas for the classification provided, where a small change in that pixel would change the classification relative to other pixels. In essence, saliency maps are constructed by back-projecting the information corresponding to the identified class, thus allowing us to visualize image regions that mostly affect prediction. Here, the keras-vis [53] toolkit was used to extract saliency maps.

It is hypothesized that using saliency maps with EEG power modulation spectrograms will allow for new biomarkers of AD to be developed in a data-driven manner and for comparisons with visually inspected regions to be made. Unlike simple images, however, EEGs are extracted across
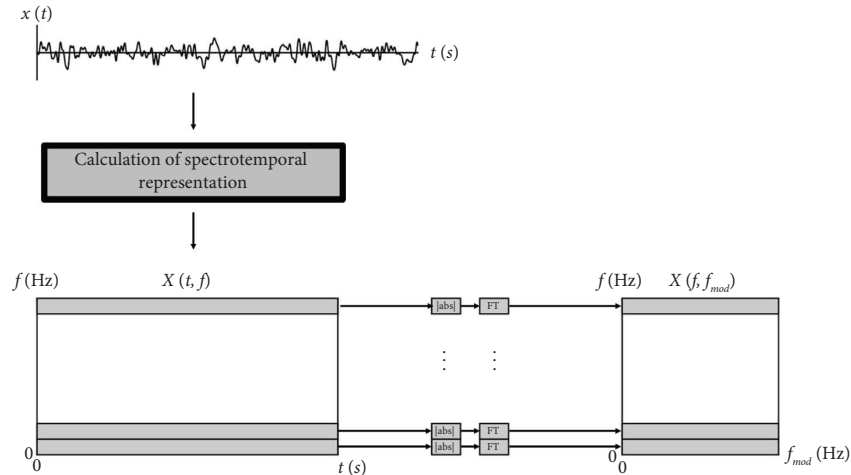
FIGURE 1: Signal processing steps for the computation of the EEG power modulation spectrogram. (Image adapted from [29]).
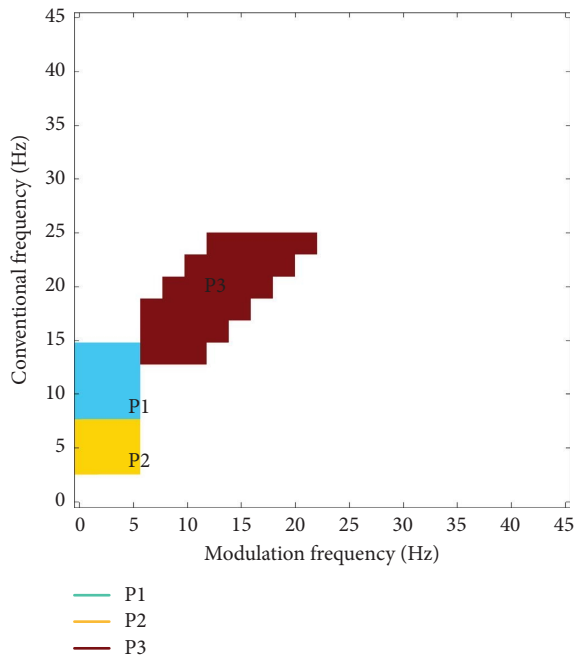

FIGURE 2: Three modulation spectrogram patches proposed in [29].

parameter tuning was performed, and Table 2 presents the tested parameters and final values used.

2.5. Proposed Method. Figure 3 depicts the block diagram of the proposed method. First, EEG signals are segmented and transformed to the power modulation spectrogram domain which is then z-normalized. Segments of 8-second duration with 1-second overlap are taken; a minimum of 460 segments are available per subject. CNNs are then trained on five different classification tasks, namely, Task 1 (T1): N vs. AD1 vs. AD2 (multi-class discrimination); Task 2 (T2): N vs. AD (AD diagnosis); Task 3 (T3): N vs. AD1 (early AD detection); Task 4 (T4): AD1 vs. AD2 (AD progression); and Task 5 (T5): N vs. AD2 (late-stage AD detection). Saliency maps are then extracted for each of the five tasks to indicate regions of importance for each task. To allow for comparisons with the three visually obtained regions [29], a clustering algorithm is applied on the saliency map "islands" to obtain new optimal patches. These patches are then used for classification. In the subsections to follow, more details about the clustering method and final classification steps are given. Algorithm 1 shows an overview of the processing steps involved in the feature extraction and train/testing stages of the proposed method.

2.6. Saliency Map Clustering. After the CNNs are trained, saliency maps are extracted from the last dense layer. Saliency maps from each training input are obtained and averaged over all training samples to obtain one final map. Here, we are interested in using the maps to find optimal regions in the modulation spectrogram for new biomarker development. To this end, we use thresholding and clustering to find the optimal number of patches, in a data-driven manner, for the particular task at hand. To find the optimal clusters per task for this final map, we propose to first take the difference between the average modulation spectrograms of the two groups in a given classification task and use the saliency map as a mask to be applied to this differential spectrogram. When clustering, two parameters

multiple channels, each generating its own image-like modulation spectrogram. As information from multiple channels is taken into account during classification, 20 different saliency maps, one for each EEG channel can be generated. While each of these maps could be used separately, here, for simplicity, we take their overall average and use the aggregated map for analysis; use of individual channel maps is left for a future study. In our experiments, a CNN with two convolution layers was used, having as input a tensor of $45 \times 45 \times 20$ dimensionality. The kernel size is equal to 3×3 and a dropout rate of 85% was used. ReLU activation functions were used for the convolution layers. In the fully-connected layers, in turn, LeakyReLU was used as an activation function. A total of three fully-connected layers were used, including the last classification layer. Hyper-

TABLE 2: CNN hyper-parameter tuning details.

Hyper-parameters	Range explored	Chosen
Kernel_size	(3.3), (5.5), (7.7), (9.9)	(3,3)
Regularizers.l2	$(1e-2), (1e-4)$	$(1e-2)$
Dropout	(65%, 75%, 85%, 90%)	85%
Optimizer	Adam, Nadam, Adagrad, Adamax	Nadam
Learning rate	(0.01, 0.001, 0.0001)	0.0001
Batch_size	4, 8, 32, 64	4
Epochs	20, 30, 40, 50	50
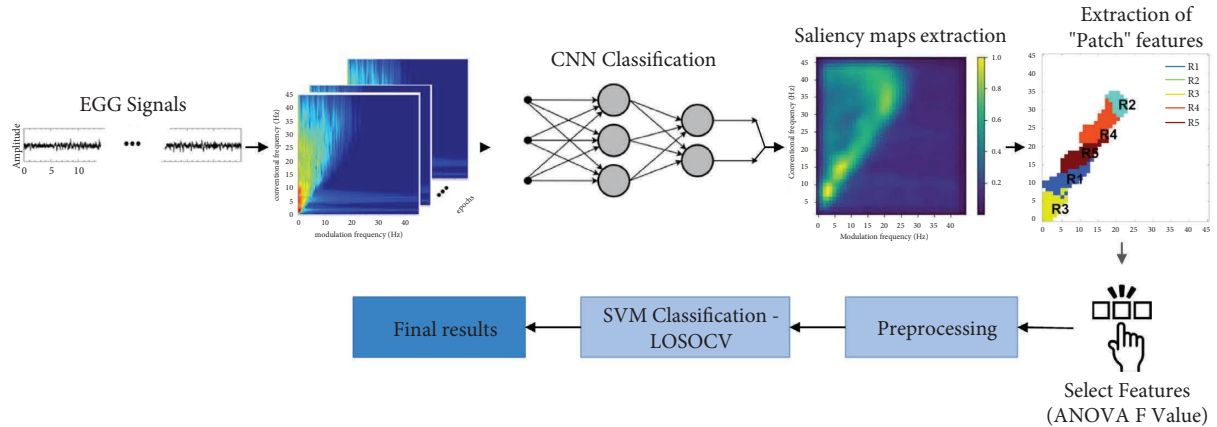
FIGURE 3: Overview of the proposed system architecture.

Step 1: Feature extraction
(1) Compute modulation spectrum from preprocessed EEG $\longrightarrow$ Obtain X (f, f_mod) (equation (1))
(2) Train CNN on five different classification tasks using training and validation partitions of available datasets
(a) N vs. AD1 vs. AD2 (multiclass discrimination) $\longrightarrow$ CNNa
(b) N vs. AD (AD diagnosis) $\longrightarrow$ CNNb
(c) N vs. AD1 (early AD detection) $\longrightarrow$ CNNc
(d) AD1 vs. AD2 (AD progression) $\longrightarrow$ CNNd
(e) N vs. AD2 (late-stage AD detection) $\longrightarrow$ CNNe
(3) Find saliency map of each CNNi ($i = a, \ldots, e$) using validation partition data
(4) Cluster saliency islands using the k-means algorithm into final "patches"
Step 2: Train/testing with feature selection
(1) Compute modulation energy from patches from all electrodes for each CNNi
(2) Apply feature selection based on ANOVA for each CNNi
(3) Run leave-one-subject-out cross-validation using unseen test set data with a SVM classifier for each CNNi
(4) Calculate classifier figures of merit for CNNi

ALGORITHM 1: Summary of methodology steps.

are explored via grid search, namely, the saliency value threshold and the number of clusters. Threshold values from 80–96% were explored, with a hop of 2% and 3–5 clusters were tested. We selected three as the minimum to coincide with the patches proposed in [54] and five to strike a balance with feature dimensionality. Clustering was performed via the k-means algorithm in MATLAB.

Four distinct masking approaches are tested to explore their impact on overall accuracy. They vary based on which task the saliency map was obtained from, e.g., more generic, as in Task 2, to more specific, as in Task 3. These four different experiments are detailed as follows:

(i) Experiment 1: This experiment takes the generic N vs. AD task and uses the salience map obtained from the three-class task N vs. AD1 vs. AD2 as a mask.

(ii) Experiment 2: This second experiment is a bit more tuned to the task at hand as it considers individual subclasses directly, while still using the most generic saliency map. In particular, the difference signal is taken for each binary task, namely, N vs. AD1, N vs. AD2, and AD1 vs. AD2. The same saliency map as Experiment 1 is used.

(iii) Experiment 3: This is the most specialized of the experiments as the saliency maps corresponding to

each differential modulation spectrograms are used to find the optimal clusters. For example, saliency maps found for the N vs. AD1, N vs. AD2, and AD1 vs. AD2 tasks are used with differential spectrograms obtained from N vs. AD1, N vs. AD2, and AD1 vs. AD2 classes, respectively.

(iv) Experiment 4: This experiment builds on Experiment 1 and takes the N vs. AD task and uses the salience map obtained from the same N vs. AD mask, which can be seen as the most generic.

2.7. Biomarker Selection and AD Diagnosis.

Once the optimal number of clusters is found, these regions will become candidate patches for each of the five tasks. As in [29], the modulation spectrum power R_i in patch/cluster i is computed as follows:

$$R_i = \iint_{RG_i} = |X(f, f_{\text{mod}})|^2 \, df \, df_{\text{mod}}, \qquad (2)$$

where RG_i corresponds to the new patch found by cluster i. As in [54], the power ratios between different patches are also treated as features and these are computed across all 20 channels.

As can be seen, the number of candidate features/biomarkers can quickly grow with an increasing number of channels and regions, causing potential curse-of-dimensionality issues with the downstream classification tasks. As such, feature selection is needed to reduce the final number of features to a reasonable number. Previous EEG-based works have relied on 24 input features (e.g., see [23, 25]), and we follow this procedure to allow for fair comparisons with prior works. We use 25% of the training set (described in the next subsection) to find the best 24 features using an ANOVA F-value metric computed between each feature and class label.

Previous works have relied on a support vector machine (SVM) classifier to map patch features to a final diagnostic decision. For consistency, we also use an SVM for final AD classification, thus allowing for a more fair comparison with previous work. This also assures that the performance gains achieved are a result of the improved biomarkers and not of an improved classifier (e.g., a CNN itself). SVMs use kernels to map data from two classes into a higher dimension in which a hyperplane can be used to separate the two different classes by a certain margin [55]. Different kernels provide different properties and allow for more complex class partitions to be found. Here, a radial basis function (RBF) kernel with gamma $\gamma = 1/24$, where 24 is the number of features, and $C = 1$ are used as hyper-parameters and tuning is not performed to gauge the benefits of the features themselves, and not on the classifier. Prior to classification, the top-24 new biomarkers are normalized between $[-1, 1]$. Final testing follows a leave-one-subject-out (LOSO) plus bootstrapping procedure using the disjoint test set described below. With the LOSO-plus-bootstrapping setup, for each subject, the classifier is trained with data from $N - 1$ subjects that are randomly sampled and repeated ten times.

2.8. Testing Setup.

The available dataset has to be split to allow for CNN training for biomarker selection as well as for SVM training for final classification. As such, the available data needs to be partitioned in such a way that data leakage does not occur into the final test set. To this end, the data partitioning scheme illustrated in Figure 4 is performed to avoid any data leakage into the disjoint training, validation, and test sets. First, the 460 segments are partitioned into five parts, each with 92 segments. Since the segments are obtained with a 7-second overlap, the last 7 segments of each part are discarded, to avoid information leakage between folds (see Figure 4(a)). Each of the five parts is comprised of 85 segments, for each of the 54 subjects (Figure 4(b)). Finally, a time-dimension shuffle was made to avoid any ordering effects on the data. From the shuffled data, 1/5 of the data was set aside for validation, 1/5 for testing, and 3/5 for training (Figure 4(c)).

2.9. Figures-of-Merit and Benchmark Method.

We use accuracy and $F1$-score as figures-of-merit to gauge the performance of the proposed method and compare against the visual-inspection-based biomarkers from [29]. Accuracy represents the ratio of correct predictions to total predictions, i.e.,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \qquad (3)$$

where TP stands for true positive (target label is correctly predicted), TN for true negative (nontarget label predicted correctly as nontarget), FP for false positive (nontarget label predicted as target), and FN for false negative (target label erroneously predicted as nontarget).

In turn, $F1$-score is given by the weighted average between recall and precision, namely,

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \qquad (4)$$

$F1$-scores are useful for unbalanced datasets, such as the one used here. For completeness, precision identifies how accurately the model predicted the positive classes, i.e.,

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad (5)$$

where recall measures the ratio of the number of true positive events to the sum of true positive and false negative events, i.e.,

$$\text{Recall} = \frac{TP}{TP + FN}. \qquad (6)$$

As for the benchmark, we use the state-of-the-art system described in [29] for comparisons.

2.10. Age-Related Confounds.

Age is a known confounding factor and risk factor for AD [56, 57]. Since the healthy and patient populations in our dataset are not age-matched, we need to be careful not to propose features that are correlates of normal neural deterioration due to aging (e.g., see
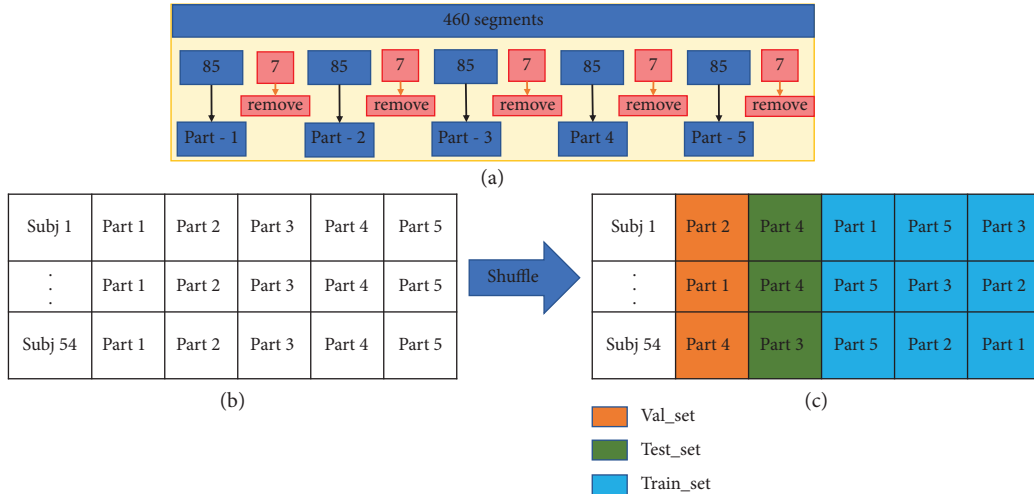
Figure 4: Steps to perform database partitioning to avoid information leakage. (a) First, 460 segments per subject are split into 5 parts of 92 segments. The last seven are discarded to avoid data leakage. (b) Disjoint parts are then combined per subject. (c) Lastly, temporal shuffle is done to avoid ordering effects. 1/5, 1/5, and 3/5 of the shuffled data are partitioned into validation, testing, and training subsets.

[58, 59]), but instead, related to neurodegeneration due to the disease. First, we explore how well the top-24 features can be used to predict age via a linear regression. To test if the top-selected features carry age-related information, we compare them against a random age prediction regressor. With this random linear regressor, we shuffle the age of participants in a random manner. For this task, we combine all the three dataset partitions and split them into train/test using 75% and 25% of the data, respectively. This partition is run 100 times, and the average root mean square error (between true and predicted age) and standard deviations are reported. To assure that the differences with the random regressor are not significant, a t-test is performed with a 99% confidence level ($p > 0.01$); the Python statistical package sciPy.stats was used for this test.

3. Results and Discussion

3.1. CNN Accuracy. First, we explore how well the CNN is behaving on the validation set in order to gauge the effectiveness of the obtained saliency maps for downstream AD detection on the unseen test set. Table 3 reports the results obtained for each of the five classification tasks. We use only accuracy (Acc) and the $F1$-score for this analysis. Overall, the N vs. AD2 (T5) task resulted in the highest accuracy and $F1$ score. This is expected, as the neural changes with moderate-to-severe AD are likely to be the greatest compared to healthy EEG, thus making it easier for the CNN to distinguish them. The accuracies obtained are in line with those reported previously for manually-selected clean EEG segments (e.g., see [23]), hence providing confidence on the potential of the CNN to find discriminatory features for AD classification, thus validating the use of the salience maps for feature extraction and potentially artifact rejection.

3.2. Saliency Maps. With the CNN approach validated, we proceed to investigate the saliency maps generated from the

EEG channels. While each individual channel map could be used for channel-specific feature generation, here, for simplicity, we explore the use of only one general mask for all channels. As such, the average map is used. Figure 5 depicts the average map found for each of the five tasks. At a first glance, similar regions to the patches reported in reference [29] (shown in Figure 2) are seen, but with additional regions also showing importance. Next, we explore the best threshold and number of clusters for each of the four experiments tested.

3.3. Saliency Map Clustering for Patch Detection. Table 4 shows the best thresholds (Th) and number of clusters (C) found for each of the four experiments; henceforth, these combinations are used. The optimal patch regions found based on these parameters can be seen in Figures 6 and 7 for Experiment 1 and Experiment 4, respectively. The plots of the other two experiments are omitted for brevity, but similar regions were found. In each subfigure of these plots, the x-axis corresponds to modulation frequency, the y-axis to conventional frequency, the left-most plot shows the patches found with the optimal threshold, and the right-most plot shows the optimal clusters found. In each image, the found clusters/regions for each task are labelled as "R_i." It is important to emphasize that each task achieved different regions of importance, hence, e.g., R_1 from Task 1 may differ from R_1 from Task 2. As such, when listing the rank of top features in tables to follow, we use a subscript from 1–5 to remind the reader that the region listed is related to its corresponding counterpart seen in Figures 6 and 7.

As can be seen, the patches found using Experiment 1 settings resemble closer those found via visual inspection in reference [29], especially P2 and P3 seen in Figure 2. Other than the N vs. AD1 task, the obtained regions look fairly alike for all tasks. One region not used before but that was shown to be important in this data-driven analysis method corresponds to patches in lower frequencies, below 5 Hz (delta

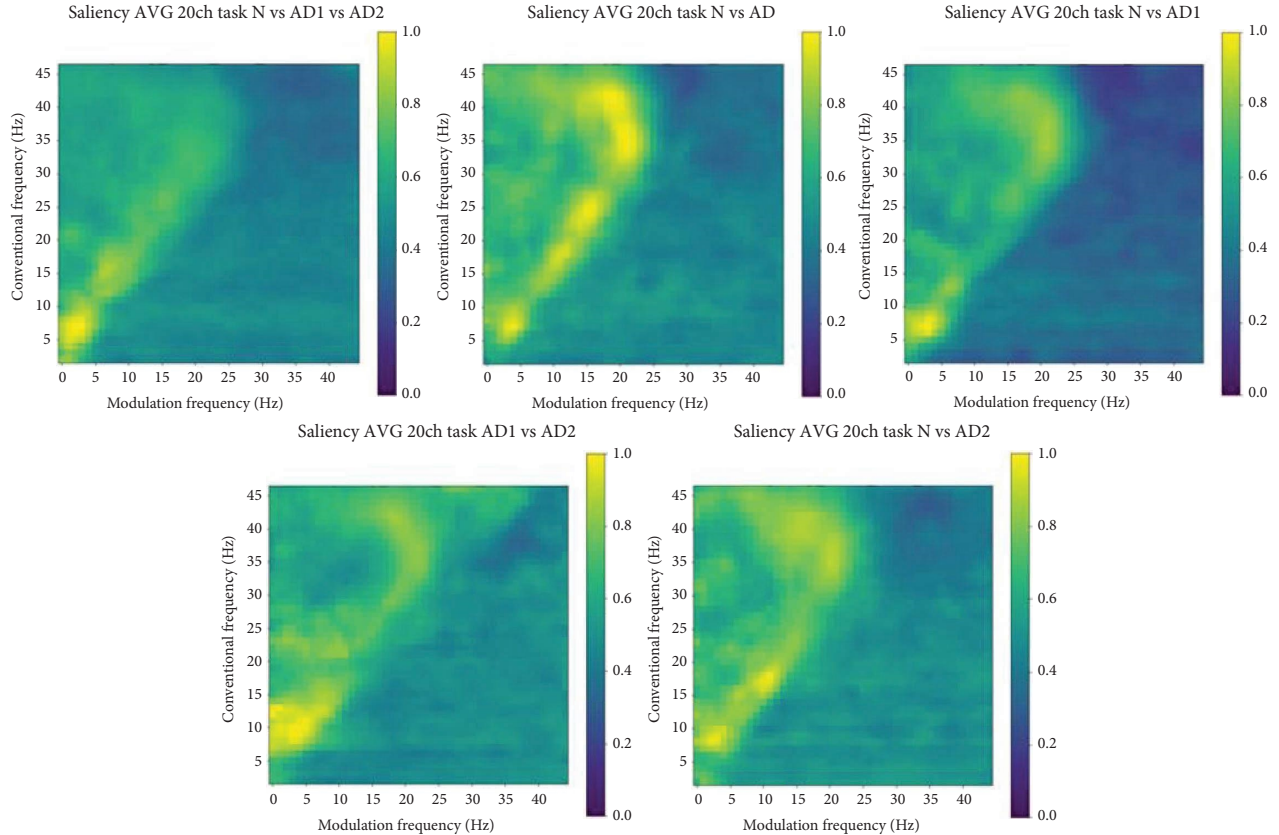| Tasks	Acc (%)	F1 (%)
T1	90.5	90.5
T2	87.3	84.6
T3	91.4	91.0
T4	92.5	91.4
T5	93.0	91.6

FIGURE 5: Average saliency map for each of the five tasks.

and low theta), corroborating some earlier findings from reference [23]. Experiment 4 settings, in turn, showed similar regions to those obtained from visual inspection but also included across all tasks important information extracted from higher frequencies (around the beta and gamma frequency ranges), a finding not seen previously in modulation spectral studies.

3.4. Feature Importance. The optimal patches for each classification task have now been found and they serve as masks for biomarker extraction from each EEG channel. To reduce the number of features being used by the SVM classifier, feature ranking is performed and only the top-24 features are kept. Table 5 lists the top features for each of the five tasks for the Experiment 1 setting. The notation RXoRY is used to indicate the feature corresponding to the ratio of the modulation power in patch RX to the modulation power in patch RY. Subscripts 1–5 indicate that the regions are those corresponding to Tasks 1–5. It can also be seen that for

the majority of the tasks, frontal region electrodes stood out. Interestingly, as reported in reference [25], frontal regions are usually discarded with EEG analysis relying on visual inspection of EEGs, as these areas are often contaminated by eye movement artifacts. Our findings suggest that such important regions may be again incorporated into EEG analysis by utilizing automated artifact removal algorithms, such as wICA, and features that could be inherently more robust to artifacts. For the AD1 vs. AD2 task, most features belong to the frontal and central electrodes, a region known to be affected by the progression of AD [60], likely due to the expansion of the atrophy into the superior parietal and frontal cortex [61]. For the N vs. AD and N vs. AD1 tasks, in turn, features extracted from the temporal brain region stood out, corroborating studies showing atrophy in cortical regions in the temporal regions [62]. In particular, in the N vs. AD1 task, temporal and frontal brain regions stood out; such fronto-temporal regions have been used for early AD diagnosis in the past [63].

TABLE 4: Best combination of threshold (Th) and number of clusters (C) for each of the four experiments.

Tasks	Exp. 1		Exp. 2		Exp. 3		Exp. 4	
	Th (%)	C	Th (%)	C	Th (%)	C	Th (%)	C
T1	92	3	92	3	92	3	92	3
T2	96	4	96	4	96	5	96	5
T3	86	3	84	4	92	4	96	3
T4	94	5	86	5	82	4	86	4
T5	96	3	92	4	94	3	90	4

(a)

(b)
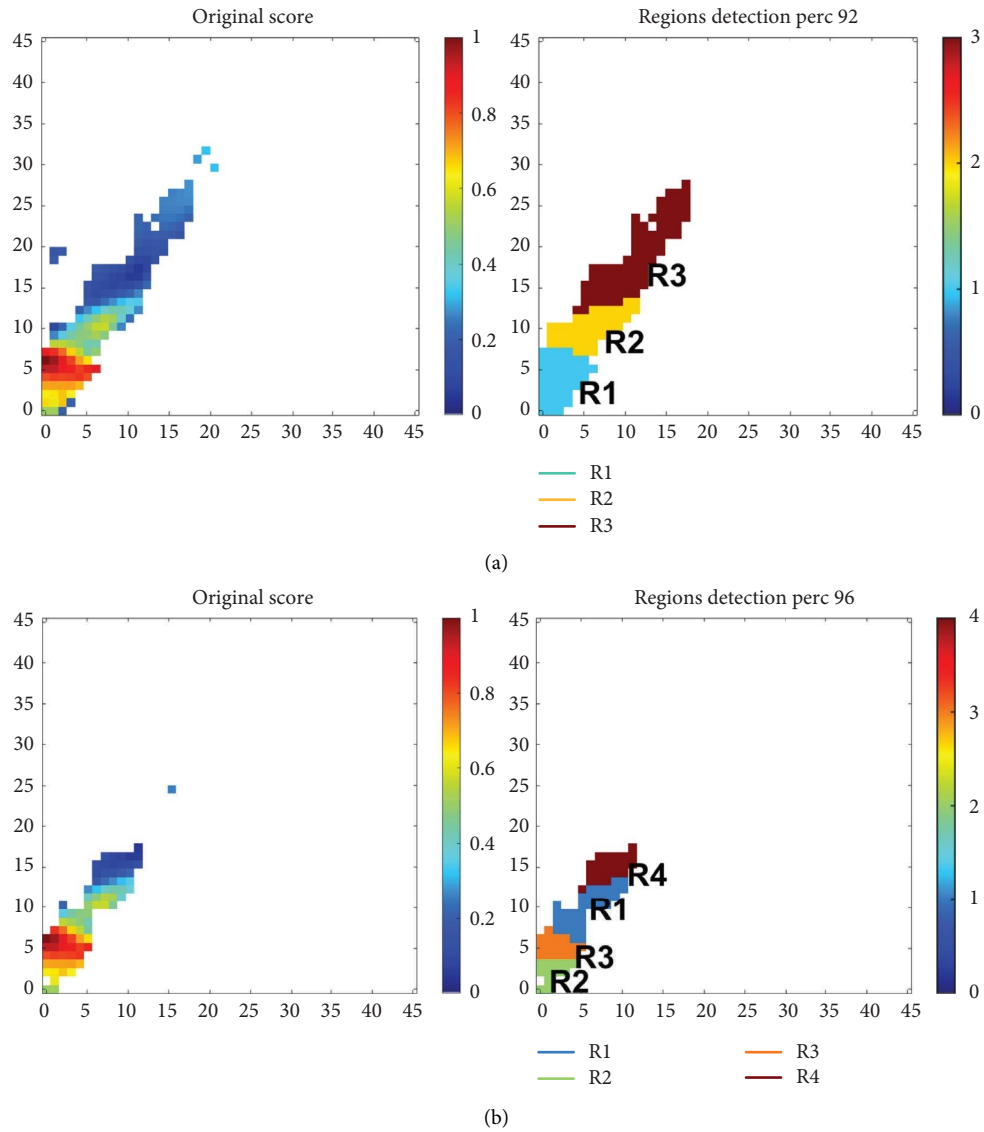
FIGURE 6: Continued.

(c)

(d)

(e)

Figure 6: Optimal regions and clusters found for each of the five tasks using experiment 1 settings. In each subplot, the y-axis corresponds to conventional frequency (unit: Hz) and the x-axis to modulation frequency (unit: Hz). (a) N vs. AD1 vs. AD2, (b) N vs. AD, (c) N vs. AD1, (d) AD1 vs. AD2, and (e) N vs. AD2.
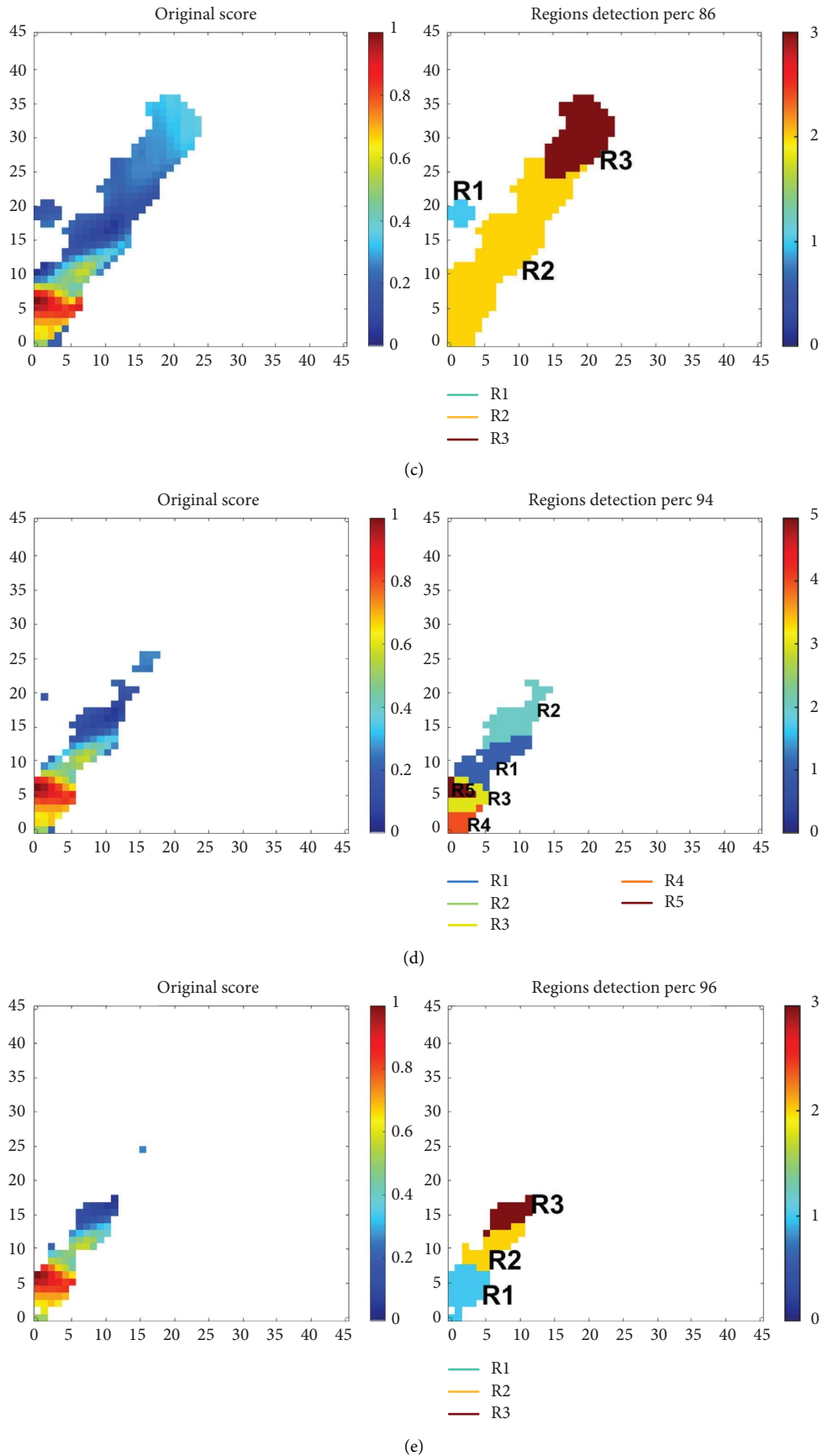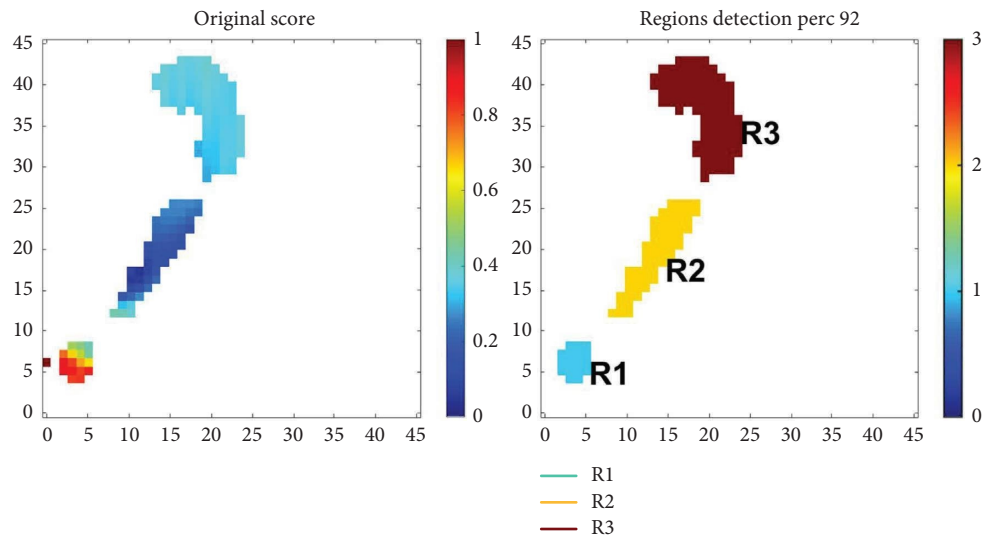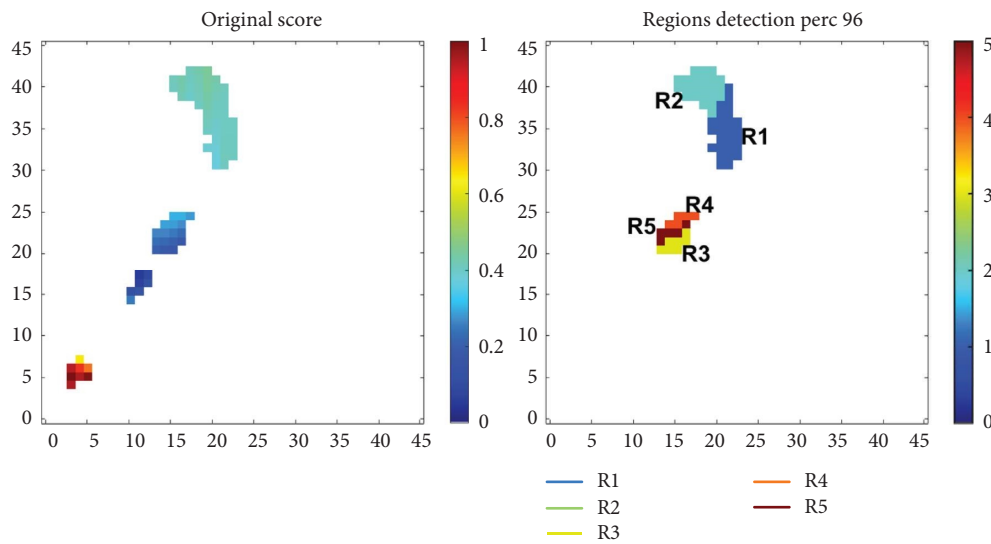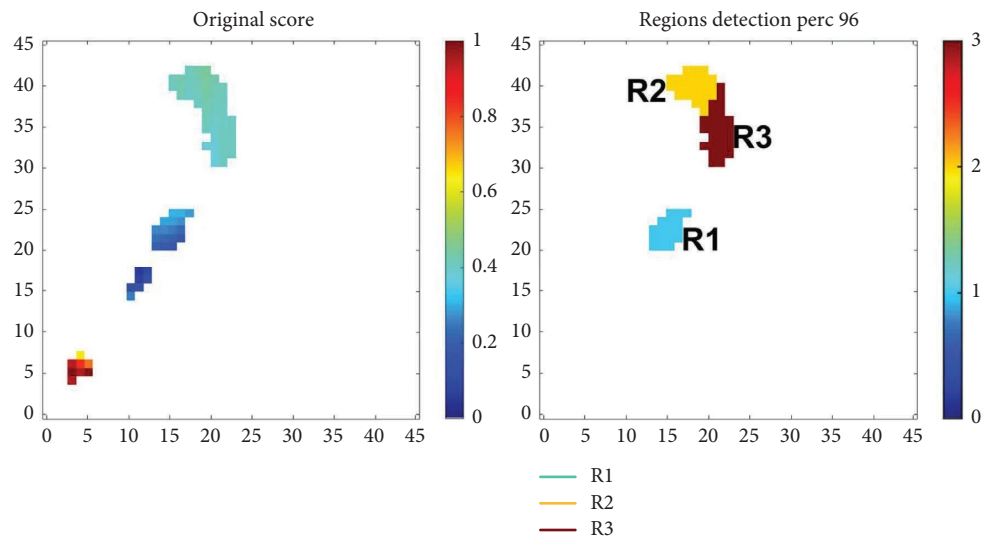
(a)
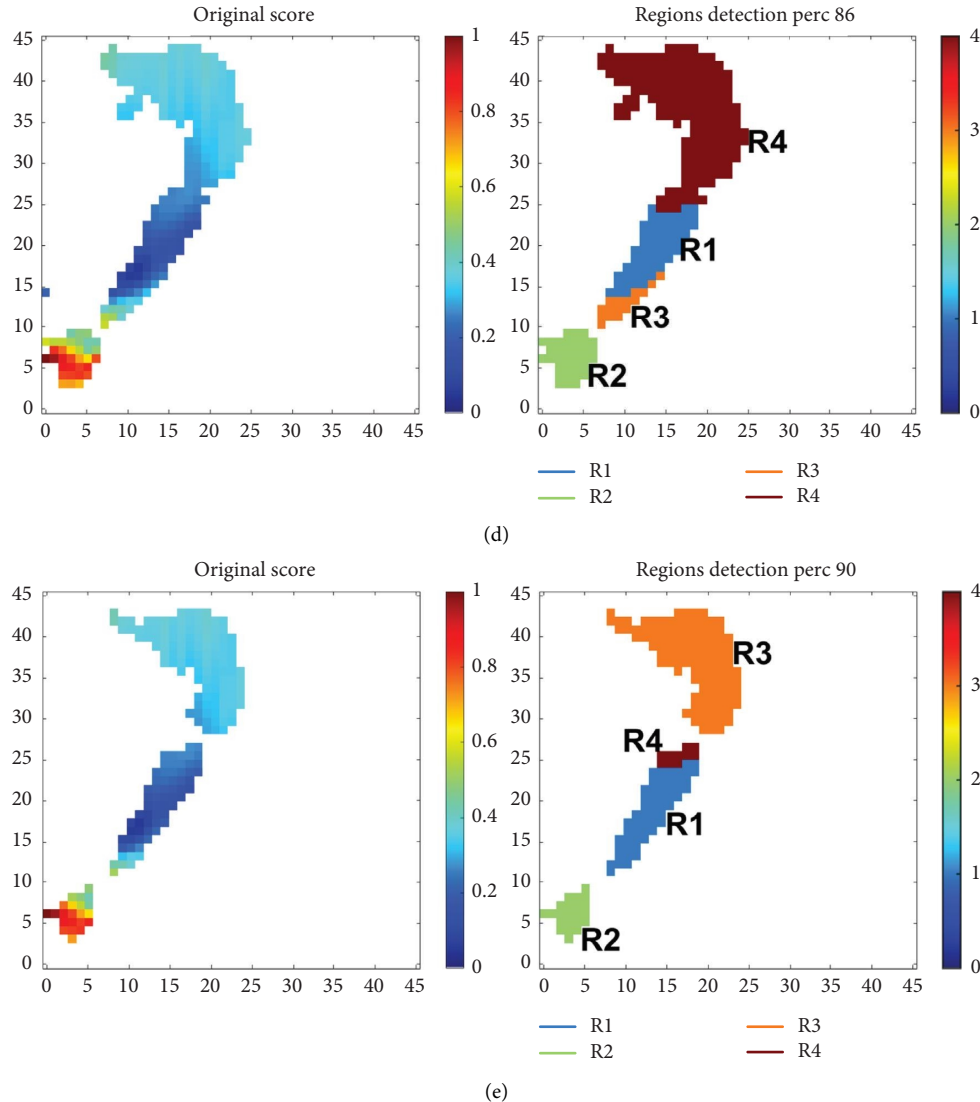

(b)

(c)

Figure 7: Continued.

Figure 7: Optimal regions and clusters found for each of the five tasks on the experiment 4 settings. In each subplot, the *y*-axis corresponds to conventional frequency (unit: Hz) and the *x*-axis to modulation frequency (unit: Hz). (a) *N* vs. AD1 vs. AD2, (b) *N* vs. AD, (c) *N* vs. AD1, (d) AD1 vs. AD2, and (e) *N* vs. AD2.

As for the Experiment 4 setting, Table 6 shows the top-24 features selected. Unlike the Experiment 1 settings, high-frequency range features (both in conventional and modulation frequency domains) appear in the top-5 features for three of the five tasks, namely, *N* vs. AD, *N* vs. AD1, and AD1 vs. AD2. This suggests that such frequency ranges could be useful for early diagnosis and AD progression assessment, hence, providing new insights not previously achieved via visual inspection. Some research has shown the importance of gamma band activity in AD [64], though this region is often corrupted by muscle artifacts. The findings obtained here suggest that the temporal dynamics of the beta and gamma bands may provide some discriminatory information and that a more generic mask may capture the dynamics of a band that is typically discarded due to artifacts, hence further emphasizing the robustness of the new biomarkers. In addition, some of the regions found belong to the beta-m-

alpha and gamma-m-alpha frequencies. These frequencies have been shown in [65] to correlate with cerebral hemodynamics information conveyed by functional near-infrared spectroscopy in regions related to impaired blood flow in AD [66, 67], thus providing some additional interpretability to the selected biomarkers.

With these settings, it can be seen that most of the features were derived from electrodes over the temporal, parietal, and occipital regions, thus in agreement with [62]. An occipital brain region was most important for two of the five tasks (*N* vs. AD and *N* vs. AD1) and was the second most important region in the AD1 vs. AD2 task, hence corroborating classical findings from reference [68]. Unlike these classical studies, however, where occipital changes were found mostly in lower frequency ranges, here we observe them to be extracted from higher ranges around beta conventional frequency and gamma modulation

Table 5: Top 24 features selected using the experiment 1 settings.

Rankings	N vs. AD1 vs. AD2	N vs. AD	N vs. AD1	AD1 vs. AD2	N vs. AD2
1	$R3\text{-}C4_1$	$R2\text{-}F7_2$	$R2\text{-}F7_3$	$R2\text{-}Oz_4$	$R3\text{-}C4_5$
2	$R3\text{-}P3_1$	$R3\text{-}F7_2$	$R1oR2\text{-}F7_3$	$R2\text{-}Fz_4$	$R3\text{-}P3_5$
3	$R3\text{-}Fz_1$	$R2\text{-}C4_2$	$R1oR3\text{-}Oz_3$	$R2\text{-}F4_4$	$R3\text{-}C3_5$
4	$R3\text{-}Oz_1$	$R1\text{-}C4_2$	$R1oR3\text{-}Pz_3$	$R2\text{-}P3_4$	$R3\text{-}F3_5$
5	$R3\text{-}F3_1$	$R3\text{-}C4_2$	$R1oR3\text{-}O2_3$	$R3oR2\text{-}C3_4$	$R3\text{-}P4_5$
6	$R3\text{-}C3_1$	$R1\text{-}F7_2$	$R2\text{-}C4_3$	$R2\text{-}F3_4$	$R3\text{-}T4_5$
7	$R3\text{-}F4_1$	$R2\text{-}T4_2$	$R1oR3\text{-}O1_3$	$R4oR2\text{-}C3_4$	$R2\text{-}C4_5$
8	$R3\text{-}P4_1$	$R4\text{-}C4_2$	$R2\text{-}T3_3$	$R2\text{-}C3_4$	$R3\text{-}F4_5$
9	$R3\text{-}Cz_1$	$R2\text{-}F3_2$	$R1oR2\text{-}T4_3$	$R2\text{-}O2_4$	$R2\text{-}F3_5$
10	$R3\text{-}T4_1$	$R2\text{-}T3_2$	$R1oR2\text{-}C4_3$	$R2\text{-}Pz_4$	$R3\text{-}Oz_5$
11	$R3\text{-}O1_1$	$R3\text{-}T4_2$	$R1oR2\text{-}C3_3$	$R5oR2\text{-}C3_4$	$R3\text{-}Fz_5$
12	$R3\text{-}O2_1$	$R3\text{-}F3_2$	$R1oR2\text{-}F3_3$	$R2\text{-}O1_4$	$R2\text{-}P3_5$
13	$R3\text{-}Pz_1$	$R2\text{-}T6_2$	$R2\text{-}T6_3$	$R3oR1\text{-}C3_4$	$R2\text{-}T4_5$
14	$R2\text{-}C4_1$	$R1\text{-}T4_2$	$R3\text{-}F7_3$	$R2\text{-}Cz_4$	$R1\text{-}F3_5$
15	$R1\text{-}F7_1$	$R2oR3\text{-}F8_2$	$R1oR2\text{-}T3_3$	$R1oR2\text{-}C3_4$	$R3\text{-}T5_5$
16	$R2\text{-}F3_1$	$R3\text{-}T3_2$	$R2\text{-}T4_3$	$R1oR2\text{-}F7_4$	$R3\text{-}O1_5$
17	$R2\text{-}fz_1$	$R4\text{-}F7_2$	$R1oR3\text{-}Fp1_3$	$R5oR3\text{-}C3_4$	$R1\text{-}C4_5$
18	$R1\text{-}C4_1$	$R2\text{-}C3_2$	$R1oR2\text{-}T6_3$	$R2\text{-}P4_4$	$R2\text{-}Fz_5$
19	$R1\text{-}F3_1$	$R4\text{-}T4_2$	$R1oR3\text{-}P4_3$	$R3oR2\text{-}F7_4$	$R3\text{-}T6_5$
20	$R2\text{-}P3_1$	$R2oR3\text{-}F7_2$	$R3oR2\text{-}T3_3$	$R4oR1\text{-}C3_4$	$R2\text{-}C3_5$
21	$R1\text{-}T4_1$	$R1\text{-}F3_2$	$R3oR2\text{-}F7_3$	$R5oR4\text{-}C3_4$	$R3\text{-}Cz_5$
22	$R2\text{-}T4_1$	$R3\text{-}T6_2$	$R1oR2\text{-}Fp1_3$	$R2\text{-}C4_4$	$R1\text{-}T4_5$
23	$R3\text{-}T5_1$	$R2\text{-}P3_2$	$R1oR2\text{-}T5_3$	$R1\text{-}Oz_4$	$R2\text{-}F4_5$
24	$R2\text{-}F4_1$	$R2oR3\text{-}T4_2$	$R1oR2\text{-}F8_3$	$R5oR3\text{-}T5_4$	$R2\text{-}P4_5$
Number of features per brain region—experiment 1 settings					
Frontal	8	9	8	5	7
Central	5	5	3	11	6
Temporal	4	9	8	1	5
Parietal	4	1	2	3	4
Occipital	3	0	3	4	2

frequencies, hence a finding also not previously found via visual inspection. For space limitations, we omit the features selected via the other two experiments, but they were similar to those found in Experiment 1 and 4.

3.5. Overall Classification Accuracy. Lastly, using the top-24 features, Table 7 shows the final accuracy and F1 scores obtained for each of the four experiments and five classification tasks. The results achieved with the benchmark, i.e., the visually obtained power modulation spectrogram patch features from reference [29], are also added for comparisons. As can be seen, the proposed features outperformed the state-of-the-art benchmark on all five tasks by a substantial margin, with the exception of the N vs. AD1 task where only a subtle improvement was obtained with the experiment 4 settings. Overall, these results suggest that making the mask generic (i.e., settings in experiments 1 and 4) can lead to improved accuracy on unseen data. This was the case for four out of the five tasks, with the exception being the N vs. AD2 task, where experiment 2 settings resulted in the best results. Ultimately, having very specialized maps (e.g., as in experiment 3 settings) did not lead to accuracy levels that could not be achieved with other less specialized settings.

3.6. Age-Related Confounds. Ultimately, it is important to gauge if the proposed features are indeed measuring neural changes due to AD and not solely due to normal aging. Such evaluation is particularly important in settings where age matching is not possible between groups, as is the case herein. Previous studies have shown a direct link between normal aging and changes in EEG powers and frequencies (e.g., see [58, 59]), thus we explore a linear mapping between the proposed feature and age and test if the obtained results correspond to those obtained from a random mapping function. Table 8 shows the root mean square errors (between estimated and true age) obtained for the Experiment 1 and Experiment 4 settings. None of the tests showed significant differences from chance at a 99% confidence level, thus suggesting that the proposed features are not capturing normal aging-related changes in the EEG and are indeed capturing neurodegenerative insights. Similar findings were found for the other two experiment settings and are omitted here for the sake of brevity.

3.7. Study Limitations and Future Work. The results reported in this study were performed on a limited sample size of 54 participants. While the results are promising, they need to be validated on a larger dataset. Open-source EEG datasets for AD are not widely available; thus, future work should focus on creating open-source datasets. The recent international push to enable EEG as a clinical biomarker [16] may enable more widespread collection that will help push the research forward. Furthermore, as is widely known, the machine

TABLE 6: Top 24 features selected using experiment 4 settings.

Rankings	N vs. AD1 vs. AD2	N vs. AD	N vs. AD1	AD1 vs. AD2	N vs. AD2
1	R2-Oz$_1$	R4oR3-O1$_2$	R1-Fp2$_3$	R1-Fz$_4$	R4-C3$_5$
2	R2-Fz$_1$	R4-F7$_2$	R1-Pz$_3$	R1-Oz$_4$	R4-P3$_5$
3	R2-P3$_1$	R4oR3-Pz$_2$	R1-Fz$_3$	R1-F4$_4$	R4-T5$_5$
4	R2-P4$_1$	R4oR2-C4$_2$	R2oR3-Oz$_3$	R1oR4-Fz$_4$	R4-P4$_5$
5	R2-C4$_1$	R4oR3-Oz$_2$	R1-O2$_3$	R1-P3$_4$	R4-O1$_5$
6	R2-F4$_1$	R4oR2-T6$_2$	R1oR3-Pz$_3$	R1-F3$_4$	R4-C4$_5$
7	R2-F3$_1$	R4-C4$_2$	R2oR3-O1$_3$	R1-O2$_4$	R4oR3-C3$_5$
8	R2-O2$_1$	R4oR2-T5$_2$	R1-Cz$_3$	R3-Oz$_4$	R4-F3$_5$
9	R2-Pz$_1$	R4oR2-T3$_2$	R2oR3-O2$_3$	R3-Fz$_4$	R4-Oz$_5$
10	R2oR3-Fz$_1$	R4-T4$_2$	R2oR3-P4$_3$	R1-Pz$_4$	R4oR3-F3$_5$
11	R2oR3-F3$_1$	R2oR1-O1$_2$	R1-O3$_3$	R1oR4-Oz$_4$	R4-T4$_5$
12	R2-C3$_1$	R4oR2-O1$_2$	R2oR3-Pz$_3$	R1-Cz$_4$	R4-T6$_5$
13	R2-O1$_1$	R4oR2-P4$_2$	R1oR3-Oz$_3$	R1-C3$_4$	R4-F7$_5$
14	R2-Cz$_1$	R4oR2-T4$_2$	R2-O2$_3$	R1-O1$_4$	R4-O2$_5$
15	R2oR3-C3$_1$	R4oR3-F4$_2$	R1oR3-O2$_3$	R1oR4-F3$_4$	R4-Pz$_5$
16	R2oR3-Oz$_1$	R2oR1-O2$_2$	R1oR3-Fp1$_3$	R1oR4-F4$_4$	R4-F4$_5$
17	R2oR3-C4$_1$	R5oR2-C4$_2$	R1oR3-Fp2$_3$	R1oR4-C3$_4$	R4oR3-T4$_5$
18	R2oR3-F4$_1$	R4oR3-O2$_2$	R2-T5$_3$	R1-P4$_4$	R4oR3-C4$_5$
19	R2oR3-P3$_1$	R4oR3-Fp1$_2$	R1oR3-Fz$_3$	R1oR4-P3$_4$	R4oR3-O1$_5$
20	R2-T4$_1$	R4oR5-O1$_2$	R2-O1$_3$	R3oR4-Fz$_4$	R4oR3-P3$_5$
21	R1oR3-F3$_1$	R4-T6$_2$	R2-P4$_3$	R1oR4-O1$_4$	R4-T3$_5$
22	R2oR3-cz$_1$	R4oR2-P3$_2$	R1oR3-T3$_3$	R3oR2-C3$_4$	R4-Fz$_5$
23	R2oR3-O1$_1$	R4oR3-P4$_2$	R1oR2-T6$_3$	R1-C4$_4$	R4oR3-F4$_5$
24	R1-C4$_1$	R4-P3$_2$	R2-Oz$_3$	R1oR4-Pz$_4$	R4-Cz$_5$
Number of features per brain region—experiment 4 settings					
Frontal	7	3	5	8	6
Central	7	3	1	5	5
Temporal	1	6	3	0	5
Parietal	4	5	5	5	4
Occipital	5	7	10	6	4

TABLE 7: Performance comparison achieved with best threshold-cluster settings and top-24 features. Bold values indicate the highest accuracy achieved for a given classification task.

Tasks	Benchmark [29]		Proposed (exp. 1)		Proposed (exp. 2)		Proposed (exp. 3)		Proposed (exp. 4)	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
T1	50 ± 3	50 ± 3	**54 ± 2**	**55 ± 2**	54 ± 2	55 ± 2	54 ± 2	55 ± 2	47 ± 2	49 ± 2
T2	70 ± 1	65 ± 1	60 ± 2	49 ± 3	60 ± 2	49 ± 3	**71 ± 2**	**61 ± 2**	71 ± 2	61 ± 2
T3	64 ± 3	63 ± 3	48 ± 2	48 ± 2	46 ± 2	46 ± 2	46 ± 4	46 ± 4	**65 ± 2**	**64 ± 2**
T4	73 ± 2	71 ± 2	**83 ± 2**	**83 ± 2**	82 ± 0	82 ± 0	83 ± 1	83 ± 1	79 ± 1	79 ± 1
T5	73 ± 2	72 ± 2	86 ± 3	86 ± 3	**89 ± 3**	**89 ± 3**	85 ± 3	84 ± 3	82 ± 2	81 ± 2

Bold values indicate the highest accuracy achieved for a given classification task.

TABLE 8: Age prediction accuracy comparison against a random regression model for experiment 1 and experiment 4 settings.

Tasks	Exp. 1		Exp. 4	
	Nonrandom	Random	Nonrandom	Random
T1	10.33 ± 2.02	9.94 ± 1.55	9.82 ± 1.65	9.71 ± 1.75
T2	9.83 ± 1.85	9.74 ± 1.45	9.40 ± 1.69	9.48 ± 1.65
T3	8.05 ± 1.20	8.55 ± 1.55	7.81 ± 1.18	8.19 ± 1.36
T4	9.74 ± 2.77	9.30 ± 2.42	9.70 ± 2.64	9.26 ± 2.63
T5	12.19 ± 2.63	11.97 ± 2.14	11.28 ± 1.82	11.63 ± 2.46

learning algorithms are sensitive to hyper-parameter tuning. Since the data available for this study was limited, not much effort was spent on parameter optimization; hence, the accuracy results reported may not be the highest achievable with the proposed feature set. Once more data become available, future work could explore the overall impact of hyperparameter tuning on classification accuracy. Moreover, in this study, saliency maps per channel were obtained, but an average map was used for simplified feature importance selection. Once more data becomes available, per-

channel optimized feature patches may be used, resulting in improved accuracy. It is known that degeneration due to AD can spread to different brain regions as the disease progresses, hence it is expected that different regional saliency maps will be useful for the different tasks explored herein. This channel-aware optimization may also lead to an overall system that can rely on a subset of the EEG channels, hence, enabling the creation of a portable, low-density, and low-cost solution, as in [69], that could help tackle AD worldwide.

Moreover, here we explored AD1 vs. AD2 discrimination, hence, obtaining insights on how disease severity can change the EEG patterns used in biomarker development. Ultimately, future work should explore neural changes seen in longitudinal studies where data from one patient's progression is monitored, thus truly leading to disease progression insights. For example, it is known that roughly a quarter of patients with mild cognitive impairment (MCI) will progress to AD within a 4-year window. Understanding the neural signature changes between MCI patients that do progress to AD and those that do not could provide useful clues not only for disease progression but also for the risk associated with developing AD at very early stages. Lastly, recent studies have shown that EEG combined with MRI could lead to useful insights for disease severity level prediction [70]. On that study, the authors showed the complementarity of the visually obtained patch features with anatomical features extracted from MRIs to predict the MMSE scores of the patients. Future work should explore the benefits of combining the proposed features with MRI ones to quantify the gains that can be achieved.

4. Conclusions

In this article, we have proposed the use of a convolutional neural network (CNN) combined with saliency maps, trained on an image-like power modulation spectrogram of eyes-closed resting-state EEGs, to find new biomarkers of Alzheimer's disease. The goal was to explore if a data-driven biomarker selection method could provide insights complementary to those obtained via visual inspection. In particular, we explored biomarkers for five classification tasks: healthy (N) vs. mild-AD (AD1) vs. moderate-to-severe AD (AD2), (2) N vs. AD (combined AD1 and AD2), (3) N vs. AD1, (4) AD1 vs. AD2, and (5) N vs. AD2. The biomarkers found were extracted for each of the available EEG channels and reduced to the top 24 features via feature selection before being input to a support vector machine for final classification. The most important brain regions found coincided with those widely reported in the AD literature, and most importantly, the power modulation spectrogram patches complemented those found previously via visual inspection. Overall, the proposed method outperformed the benchmark on all five classification tasks and by a large margin. To assure the newly-proposed features were not measuring EEG changes due to normal aging, results were compared to a random age prediction classifier and no significant differences were found.

References

[1] R. S. Wilson, E. Segawa, P. A. Boyle, S. E. Anagnos, L. P. Hizel, and D. A. Bennett, "The natural history of cognitive decline in Alzheimer's disease," *Psychology and Aging*, vol. 27, no. 4, pp. 1008–1017, 2012.

[2] W. W. Barker, C. A. Luis, A. Kashuba et al., "Relative frequencies of Alzheimer disease, lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank," *Alzheimer Disease and Associated Disorders*, vol. 16, no. 4, pp. 203–212, 2002.

[3] R. A. Sperling, P. S. Aisen, L. A. Beckett et al., "Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's and Dementia*, vol. 7, no. 3, pp. 280–292, 2011.

[4] R. Sperling, E. Mormino, and K. Johnson, "The evolution of preclinical Alzheimer's disease: implications for prevention trials," *Neuron*, vol. 84, no. 3, pp. 608–622, 2014.

[5] M. Prince, R. Bryce, and C. Ferri, *World Alzheimer Report 2011: The Benefits of Early Diagnosis and Intervention*, Alzheimer's Disease International, London, 2018.

[6] H. Rania M, E. Ammar, I. Mohamed et al., "Alzheimer's disease and its current treatments; Is there a possibility for a cure?" *Open Journal of Chemistry*, vol. 5, no. 1, pp. 013–019, 2019.

[7] A. Singh, P. Deshpande, and N. Gogia, "Exploring the efficacy of natural products in alleviating Alzheimer's disease," *Neural regeneration research*, vol. 14, no. 8, p. 1321, 2019.

[8] K. G. Yiannopoulou and S. G. Papageorgiou, "Current and future treatments in Alzheimer disease: an update," *Journal of Central Nervous System Disease*, vol. 12, Article ID 117957352090739, 2020.

[9] A. J. Mitchell, "A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and

mild cognitive impairment," *Journal of Psychiatric Research*, vol. 43, no. 4, pp. 411–431, 2009.

[10] J. C. Morris, "Current vision and scoring rules the clinical dementia rating (CDR)," *Neurology*, vol. 43, pp. 2412–2414, 1993.

[11] G. Chetelat and J.-C. Baron, "Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging," *Neuro-Image*, vol. 18, no. 2, pp. 525–541, 2003.

[12] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural mri in Alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, pp. 67–77, 2010.

[13] R. Mayeux and N. Schupf, "Blood-based biomarkers for Alzheimer's disease: plasma Aβ40 and Aβ42, and genetic variantsβ40 and a β42, and genetic variants," *Neurobiology of Aging*, vol. 32, pp. S10–S19, 2011.

[14] A. Caroli and G. Frisoni, "The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort," *Neurobiology of Aging*, vol. 31, no. 8, pp. 1263–1274, 2010.

[15] C. Van Cauwenberghe, C. Van Broeckhoven, and K. Sleegers, "The genetic landscape of Alzheimer disease: clinical implications and perspectives," *Genetics in Medicine*, vol. 18, no. 5, pp. 421–430, 2016.

[16] P. M. Rossini, R. Di Iorio, F. Vecchio et al., "Early diagnosis of Alzheimer's disease: the role of biomarkers including advanced EEG signal analysis. report from the IFCN-sponsored panel of experts," *Clinical Neurophysiology*, vol. 131, no. 6, pp. 1287–1310, 2020.

[17] F. Maestú, P. Cuesta, O. Hasan, A. Fernandéz, M. Funke, and P. E. Schulz, "The importance of the validation of M/EEG with current biomarkers in Alzheimer's disease," *Frontiers in Human Neuroscience*, vol. 13, p. 17, 2019.

[18] F. Vecchio, C. Babiloni, R. Lizio et al., "Resting state cortical EEG rhythms in Alzheimer's disease: toward EEG markers for clinical applications: a review," *Supplements to Clinical Neurophysiology*, vol. 62, pp. 223–236, 2013.

[19] R. Cassani, M. Estarellas, R. San-Martin, F. J. Fraga, and T. H. Falk, "Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment," *Disease Markers*, vol. 2018, Article ID 5174815, 2018.

[20] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, S. Luzzi, and C. Turchetti, "Eeg-based alzheimer's disease recognition using robust-pca and lstm recurrent neural network," *Sensors*, vol. 22, no. 10, p. 3696, 2022.

[21] K. AlSharabi, Y. Bin Salamah, A. M. Abdurraqeeb, M. Aljalal, and F. A. Alturki, "Eeg signal processing for alzheimer's disorders using discrete wavelet transform and machine learning approaches," *IEEE Access*, vol. 10, Article ID 89781, 89797 pages, 2022.

[22] D. Pirrone, E. Weitschek, P. Di Paolo, S. De Salvo, and M. C. De Cola, "Eeg signal processing and supervised machine learning to early diagnose alzheimer's disease," *Applied Sciences*, vol. 12, no. 11, p. 5413, 2022.

[23] T. H. Falk, F. J. Fraga, L. Trambaiolli, and R. Anghinah, "EEG amplitude modulation analysis for semi-automated diagnosis of Alzheimer's disease," *EURASIP Journal on Applied Signal Processing*, vol. 2012, no. 1, Article ID 192, 199 pages, 2012.

[24] F. J. Fraga, T. H. Falk, P. A. M. Kanda, and R. Anghinah, "Characterizing Alzheimer's disease severity via resting-awake EEG amplitude modulation analysis," *PLoS One*, vol. 8, no. 8, Article ID e72240, 2013.

[25] R. Cassani, T. H. Falk, F. J. Fraga, P. A. M. Kanda, and R. Anghinah, "The effects of automated artifact removal algorithms on electroencephalography-based Alzheimer's disease diagnosis," *Frontiers in Aging Neuroscience*, vol. 6, p. 55, 2014.

[26] P. L. Nunez, R. Srinivasan, and D. Chen, *Electric fields of the Brain: The Neurophysics of EEG*, Oxford University Press, USA, 2006.

[27] M. Elgendi, F. Vialatte, A. Cichocki, C. Latchoumane, J. Jeong, and J. Dauwels, "Optimization of EEG frequency bands for improved diagnosis of Alzheimer disease," in *Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6087–6091, IEEE, Boston, MA, USA, August 2011.

[28] E. Gallego-Jutglà, J. Solé-Casals, F.-B. Vialatte, M. Elgendi, A. Cichocki, and J. Dauwels, "A hybrid feature selection approach for the early diagnosis of Alzheimer's disease," *Journal of Neural Engineering*, vol. 12, no. 1, Article ID 016018, 2015.

[29] R. Cassani and T. H. Falk, "Alzheimer's disease diagnosis and severity level detection based on electroencephalography modulation spectral "patch" features," *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 1982–1993, 2020.

[30] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," 2013, https://arxiv.org/abs/1312.6034.

[31] A. Pulido, A. Rueda, and E. Romero, "Classification of Alzheimer's disease using regional saliency maps from brain MR volumes," *Medical Imaging 2013: Computer-Aided Diagnosis*, vol. 8670, Article ID 86700R, 2013.

[32] O. Ben-Ahmed, F. Lecellier, M. Paccalin, and C. Fernandez-Maloigne, "Multi-view visual saliency-based MRI classification for Alzheimer's disease diagnosis," in *Proceedings of the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, IEEE, Montreal, QC, Canada, November 2017.

[33] G. Mckhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of alzheimer's disease: report of the NINCDS-ADRDA work group∗ under the auspices of department of health and human services task force on alzheimer's disease," *Neurology*, vol. 34, no. 7, pp. 939–944, 1984.

[34] S. M. Brucki, R. Nitrini, P. Caramelli, P. H. Bertolucci, and I. H. Okamoto, "Sugestões para o uso do mini-exame do estado mental no Brasil," *Arquivos de neuro-psiquiatria*, vol. 61, no. 3B, pp. 777–781, 2003.

[35] N. P. Castellanos and V. A. Makarov, "Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis," *Journal of Neuroscience Methods*, vol. 158, no. 2, pp. 300–312, 2006.

[36] M. T. Akhtar, W. Mitsuhashi, and C. J. James, "Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data," *Signal Processing*, vol. 92, no. 2, pp. 401–416, 2012.

[37] L. R. Trambaiolli, T. H. Falk, F. J. Fraga, R. Anghinah, and A. C. Lorena, "EEG spectro-temporal modulation energy: a new feature for automated diagnosis of Alzheimer's disease," in *Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3828–3831, IEEE, Boston, MA, USA, August 2011.

[38] L. Atlas and S. A. Shamma, "Joint acoustic and modulation frequency," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 7, Article ID 310290, 310298 pages, 2003.

[39] R. Cassani and T. H. Falk, *Spectrotemporal Modeling of Biomedical Signals: Theoretical Foundation and Applications, Reference Module In Biomedical Sciences*, Elsevier, Amsterdam, The Netherlands, 2018.

[40] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach*, O'Reilly Media, Inc, Sebastopol, California, 2017.

[41] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems*, pp. 253–256, IEEE, Paris, France, May 2010.

[42] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," 2015, https://arxiv.org/abs/1511.08458.

[43] J. Koushik, "Understanding Convolutional Neural Networks," 2016, https://arxiv.org/abs/1605.09081.

[44] S. J. Plathottam, H. Salehfar, and P. Ranganathan, "Convolutional Neural Networks (CNNs) for power system big data analysis," in *Proceedings of the 2017 North American Power Symposium (NAPS)*, pp. 1–6, IEEE, Morgantown, WV, USA, September 2017.

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT press, Cambridge, Massachusetts, 2016.

[46] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[47] L. Brocki and N. C. Chung, "Concept saliency maps to visualize relevant features in deep generative models," in *Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1771–1778, IEEE, Boca Raton, FL, USA, December 2019.

[48] S. Banerjee, S. Mitra, B. U. Shankar, and Y. Hayashi, "A novel GBM saliency detection model using multi-channel MRI," *PLoS One*, vol. 11, no. 1, Article ID e0146388, 2016.

[49] S. Mitra, S. Banerjee, and Y. Hayashi, "Volumetric brain tumour detection from MRI using visual saliency," *PLoS One*, vol. 12, no. 11, Article ID e0187209, 2017.

[50] B. Rasolzadeh, M. Björkman, K. Huebner, and D. Kragic, "An active vision system for detecting, fixating and manipulating objects in the real world," *The International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 133–154, 2010.

[51] S. Ramenahalli, "A biologically motivated, proto-object-based audiovisual saliency model," *A&I*, vol. 1, no. 4, pp. 487–509, 2020.

[52] S. Ramenahalli, D. R. Mendat, S. Dura-Bernal, E. Culurciello, E. Niebur, and A. Andreou, "Audio-visual saliency map: overview, basic models and hardware implementation," in *Proceedings of the 2013 47th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, IEEE, Baltimore, MD, USA, March 2013.

[53] R. Henderson and R. Rothe, "Picasso: A Modular Framework for Visualizing the Learning Process of Neural Network Image Classifiers," 2017, https://arxiv.org/abs/1705.05627.

[54] R. Cassani and T. H. Falk, "Automated alzheimer's disease diagnosis using a low-density EEG layout and new features based on the power of modulation spectral "patches"," in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 1259–1263, IEEE, Bari, Italy, October 2019.

[55] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[56] R. Guerreiro and J. Bras, "The age factor in Alzheimer's disease," *Genome Medicine*, vol. 7, no. 1, 2015.

[57] C. Sala Frigerio, L. Wolfs, N. Fattorelli et al., "The major risk factors for alzheimer's disease: age, sex, and genes modulate the microglia response to a β plaquesβplaques," *Cell Reports*, vol. 27, no. 4, pp. 1293–1306.e6, 2019.

[58] J. Polich, "EEG and ERP assessment of normal aging," *Electroencephalography and Clinical Neurophysiology: Evoked Potentials Section*, vol. 104, no. 3, pp. 244–256, 1997.

[59] J. Polich, "On the relationship between EEG and P300: individual differences, aging, and ultradian rhythms," *International Journal of Psychophysiology*, vol. 26, no. 1-3, pp. 299–317, 1997.

[60] S. T. DeKosky and S. W. Scheff, "Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity," *Annals of Neurology*, vol. 27, no. 5, pp. 457–464, 1990.

[61] B. C. Dickerson, A. Bakkour, D. H. Salat et al., "The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild ad dementia and is detectable in asymptomatic amyloid-positive individuals," *Cerebral Cortex*, vol. 19, no. 3, pp. 497–510, 2009.

[62] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, 2017.

[63] M. G. Knyazeva, M. Jalili, A. Brioschi et al., "Topography of EEG multivariate phase synchronization in early Alzheimer's disease," *Neurobiology of Aging*, vol. 31, no. 7, pp. 1132–1144, 2010.

[64] J. A. van Deursen, E. F. P. M. Vuurman, F. R. J. Verhey, V. H. J. M. van Kranen-Mastenbroek, and W. J. Riedel, "Increased EEG gamma band activity in Alzheimer's disease and mild cognitive impairment," *Journal of Neural Transmission*, vol. 115, no. 9, pp. 1301–1311, 2008.

[65] L. R. Trambaiolli, R. Cassani, and T. H. Falk, "Eeg spectrotemporal amplitude modulation as a measurement of cortical hemodynamics: an eeg-fnirs study," in *Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3481–3484, IEEE, Montreal, QC, Canada, July 2020.

[66] K. Kisler, A. R. Nelson, A. Montagne, and B. V. Zlokovic, "Cerebral blood flow regulation and neurovascular dysfunction in alzheimer disease," *Nature Reviews Neuroscience*, vol. 18, no. 7, pp. 419–434, 2017.

[67] J. C. de la Torre, "Cerebral hemodynamics and vascular risk factors: setting the stage for alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 32, no. 3, pp. 553–567, 2012.

[68] M. Penttilä, J. V. Partanen, H. Soininen, and P. Riekkinen, "Quantitative analysis of occipital EEG in different stages of Alzheimer's disease," *Electroencephalography and Clinical Neurophysiology*, vol. 60, no. 1, pp. 1–6, 1985.

[69] R. Cassani, T. H. Falk, F. J. Fraga, M. Cecchi, D. K. Moore, and R. Anghinah, "Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices," *Biomedical Signal Processing and Control*, vol. 33, pp. 261–271, 2017.

[70] B. De Jesus Junior, R. Cassani, W. J. McGeown, M. Cecchi, K. Fadem, and T. H. Falk, "Multimodal prediction of alzheimer's disease severity level based on resting-state EEG and structural MRI," *Frontiers in Human Neuroscience*, vol. 495, 2020.

[71] M. K. Soares Lopes, "On the use of saliency maps for improved modulation spectral patch features for alzheimer's disease diagnosis," Ph.D. dissertation, Maîtrise en télécommunications, New York City, 2021.