International Journal of Artificial Intelligence
p-ISSN: 2407-7275, e-ISSN: 2686-3251

Original Research Paper

EEG-based Classifications of Alzheimer's Disease by Using Machine Learning Techniques

Nagarathna C. R[1]

[1] *Department of Artificial Intelligence & Machine Learning, BNM Institute of Technology. Bangalore, India.*

Article History
Received:
03.01.2024

Revised:
25.01.2024

Accepted:
12.02.2024

***Corresponding Author:**
Nagarathna C. R
Email
nagarathna.binu@gmail.com

Abstract: The study has shown how classifiers behave when identifying and categorizing Alzheimer's disease stages. The main characteristics of various frequency bands were fed into the classifier as input. The accuracy of recognition is evaluated using machine learning classifiers. The effort aims to create a novel model that combines "preprocessing, feature extraction, and classification" to identify different stages of disease. The study starts with bands filtering, moves on to feature extraction, which derives several bands from the EEG signals, and then employs KNN, SVM, and MLP algorithms to measure classification performance. "AD detection and classification using machine learning classifiers KNN, SVM, and MLP" is the main focus of this research. Five wavelet band characteristics are used by the built classifiers to categorize different illness phases. These characteristics are computed using DWT, PCA, and ICA, which aid in obtaining wavelet-related knowledge for learning. The proposed machine learning model achieves a classification accuracy of 95% overall.

Keywords: Alzheimer's Disease Classification, EEG Signal Processing, Discrete Wavelet Transform, Machine Learning Classifiers, Support Vector Machine.

International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 12-25, June 2024. **DOI: 10.36079/lamintang.ijai-01101.601**

1. Introduction

EEG analysis can currently be used to identify the characteristics of brain functioning illnesses, although it is not yet clear how these characteristics relate to Alzheimer's Disease (AD). EEG provides non-invasive, cost-effective data with great temporal precision on electrical activity in the brain during neurotransmission as compared to other imaging modalities [1]. Using a Discrete Wavelet Transform (DWT) [2] to decompose the EEG signal into its frequency sub-bands and extract a set of statistical features to represent the distribution of wavelet coefficients, the proposed framework's processing and analysis of EEG were both carried out. Data dimension reduction techniques include Principal Component Analysis (PCA) [3] and Independent Components Analysis (ICA) [4]. These attributes were then sent into a Multi-Layer Perceptron (MLP) and a Support Vector Machine (SVM) that could only produce one of two outcomes: AD or normal control (NC) [5]. To demonstrate the superiority of the classification process, the performance of the process as a result of various approaches is shown and contrasted [6]. These results serve as an illustration of how to train and test an Alzheimer's Detection prediction system using data from specific petit mal epileptic patients [7]. Given the diversity of epilepsy, it is likely that these kinds of technologies will be necessary to tailor intelligent devices for treating epilepsy to each person's neurophysiology before they are put into practical use [8].

2. Literature Review

2.1. Dataset

The EEG signals gathered from the Baskent University Hospital's Neurology Department were taken into consideration for this proposed effort [9]. 19 electrodes (Fp2, Fp1, F8, F7, F4, F3, A2, A1, T4, T3, C4, C4, T6, T5, P4, P3, O2, O1) were used to record the EEG signals. These electrodes were positioned in accordance with the global 10-20 system [10]. Three channels were referred using monopolar montage, and the remaining 15 channels were all referenced using longitudinal bipolar montage. The channel electrode montage is shown in Figure 1 [11].

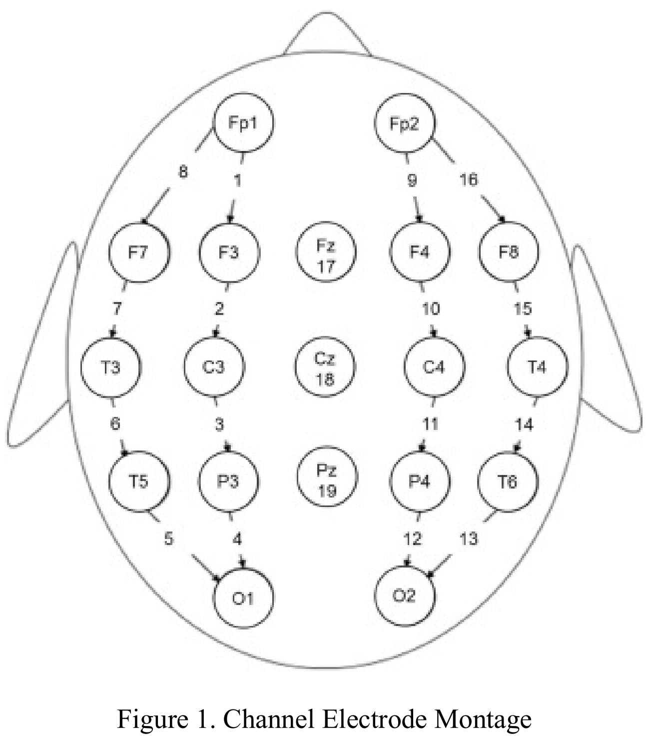

Figure 1. Channel Electrode Montage

2.2. Dataset Preprocessing

EEG signals are routinely corrupted with different artifacts while being recorded [12]. Motion, muscular, ocular, and cardiac artifacts are the most typical kinds of artifacts. Thus, the first step in our

International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 12-25, June 2024. **DOI: 10.36079/lamintang.ijai-01101.601**

proposed approach is to apply a bandpass filter to eliminate the noise [13]. Frequencies above 60 Hz are frequently labeled as noise and eliminated. In order to reduce the noise, a bandpass filter and an IIR filter are used in preprocessing [14]. Networks called filters are used to handle signals in a frequency-dependent way. A bandpass filter is an electrical component or circuit that discriminates against signals at other frequencies while permitting signals between two designated frequencies to pass [15].

For a Butterworth filter of order N with Ω_c = 1 rad/s, the poles are given by:

$$p'ak = -sin(\theta) + jcos(\theta) \tag{1}$$

where,

$$\theta = \frac{(2k-1)}{\pi 2N}, k = 1:N \tag{2}$$

Here, in order to distinguish the lowpass prototype poles from the bandpass poles that have yet to be calculated, we place a prime superscript on p. Find the analog bandpass filter's corresponding frequencies given the upper and lower -3 dB frequencies of the digital bandpass filter. We define bwHz as the -3 dB bandwidth in Hz and fcenter as the center frequency in Hz.

The discrete frequencies at -3 dB are then:

$$f1 = fcenter - bwHz/2 \tag{3}$$

$$f2 = fcenter + bwHz/2 \tag{4}$$

As previously, we will pre-warp the analog frequencies to account for the bilinear transform's nonlinearity:

$$F1 = \frac{fs}{\pi} tan(\pi f1fs) \tag{5}$$

$$F2 = \frac{fs}{\pi} tan(\pi f2fs) \tag{6}$$

Two further quantities need to be defined:

$BWHz = F2 - F1$ Hz is the pre-warped -3 dB bandwidth, and

$F0 = \sqrt{F1F1}$ is the geometric mean of F1 and F2.

Analog bandpass poles are created by converting the analog lowpass poles. We obtain two bandpass poles for each lowpass pole pa'.

$$P_a = 2\pi [\frac{BW_{Hz}}{2F0} p'_a \pm j\sqrt{1 - (\frac{BW_{Hz} p'^2_a}{2F0})}] \tag{7}$$

Utilize the bilinear transform to move the poles from the s-plane to the z-plane. The only difference is that there are 2N poles rather than N poles, much like with the IIR lowpass:

$$P_k = \frac{1+p_{ak}/2f_s}{1-p_{ak}/2f_s} \quad k = 1 \; to \; 2N \tag{8}$$

N zeros at z=-1 and N zeros at z=+1 must be added. Now we can write H(z) as follows:

$$H(z) = K \frac{(z+1)^N (z-1)^N}{(z-p_1)(z-p_2)(z-p_3)\ldots} \tag{9}$$

The N zeros at -1 and N zeros at +1 are represented as a vector in bp_synth:

q= [-ones(1,N) ones(1,N)] (10)

Polynomials with the coefficients a and bn can be created by converting poles and zeros. We obtain polynomials in z-n if we multiply the numerator and denominator of equation 1 by z2N and then expand the numerator and denominator.

$$H(z) = K(z + 1\,\frac{b_0+b_1z^{-1}+\cdots b_{2N}z^{-2N}}{1+a_1z^{-1}\cdots a_{2N}z^{-2N}}) \tag{11}$$

The input noise signal and output of bandpass filter are shown in Figure 2 and Figure 3 respectively.

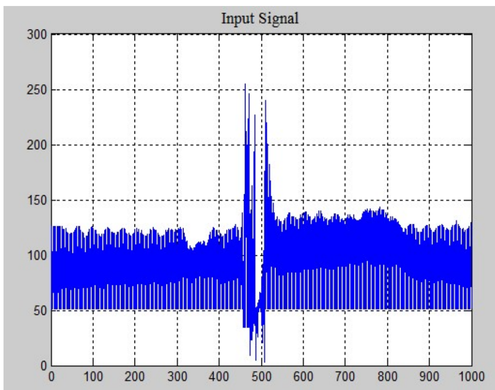

Figure 3. Input noise Signal

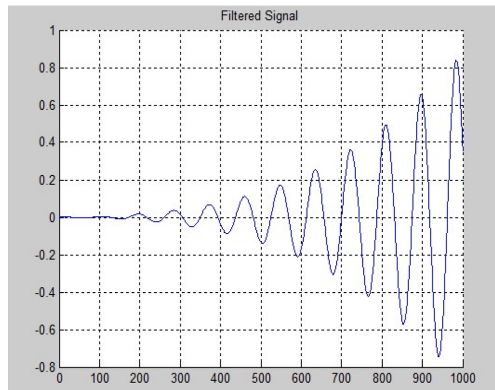

Figure 4. Filtered Signal

2.3. Feature Extraction

EEG is divided into five subband signals, such as Gamma, Beta, Alpha, Theta, and Delta, depending on frequency band, before the features are extracted from it [16]. To extract characteristics from a signal on various scales, DWT is utilized after sequential high pass and low pass filtering [17]. The EEG signals are divided into six frequency sub-bands in this case using the DWT. With the help of EEG investigations, it has been determined that there are five main frequency bands for EEG signals and a relationship exists between a certain brain region's neuronal activity and behavior. Delta (0.1 Hz or 0.5 Hz), Theta (4 Hz), Alpha (8 Hz), Beta (14 Hz), and Gamma (30 Hz–63 Hz are the most commonly used frequency bands [18] - [20].

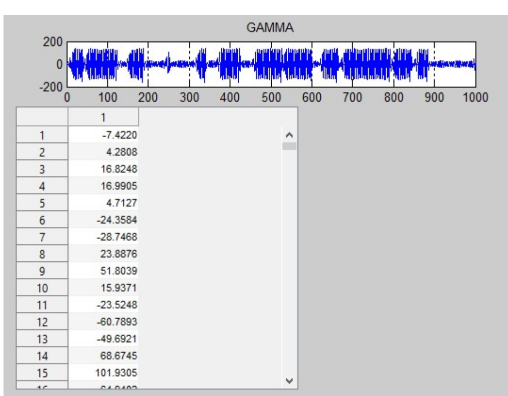

Figure 5. NC GAMA

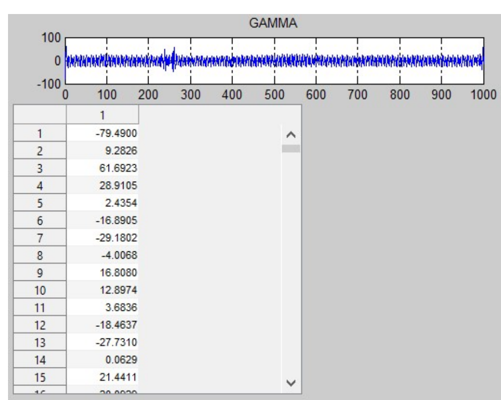

Figure 6. AD GAMA

International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 12-25, June 2024. **DOI: 10.36079/lamintang.ijai-01101.601**

The approximation and detail coefficients continue in the wavelet coefficients in a sequential manner. A transform that generates the fewest coefficients necessary to accurately recover the original signal is favored in applications that call for bilateral transformations. The DWT achieves this parsimony by keeping the range of translation and scale change, often to powers of 2, to a minimum. DWT-based analysis is best explained in terms of filter banks for the majority of signal and image processing applications. Sub-band coding refers to the process of using a set of filters to separate a signal into different spectral components.

Figure 5 to Figure 14 displays the five decomposed bands of Normal and AD signals.

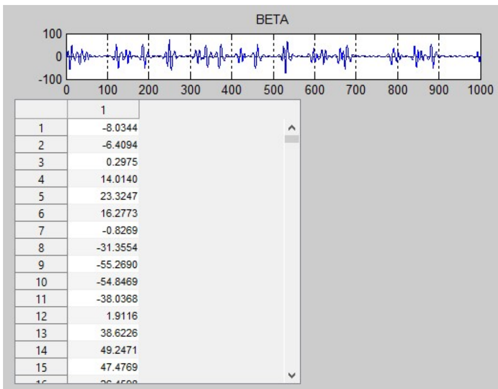

Figure 7. NC BETA

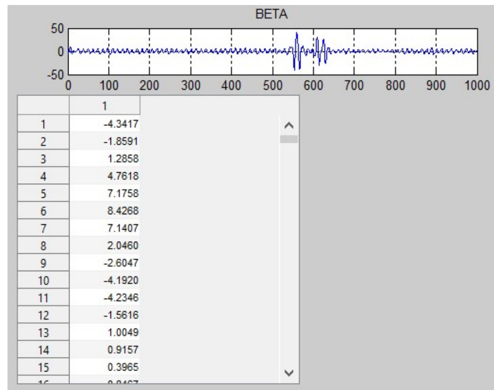

Figure 8. AD BETA

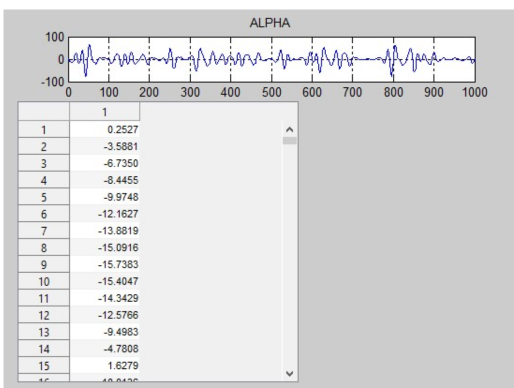

Figure 9. NC ALPHA

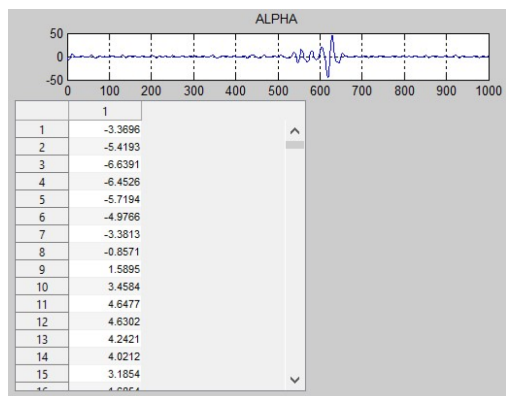

Figure 10. AD ALPHA

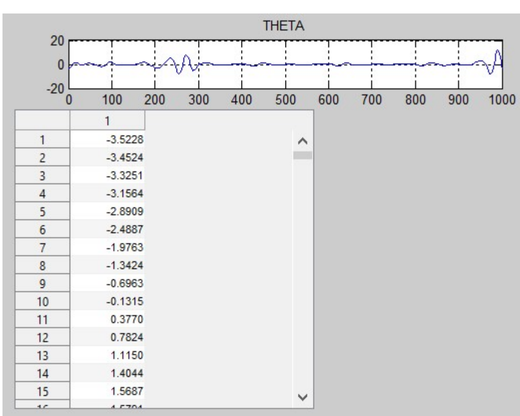

Figure 11. NC THETA

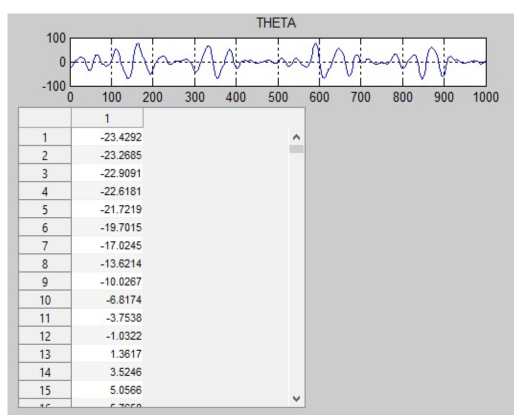

Figure 12. AD THETA

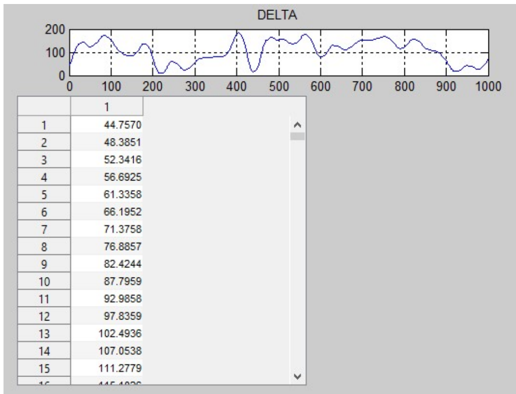

Figure 13. NC DELTA

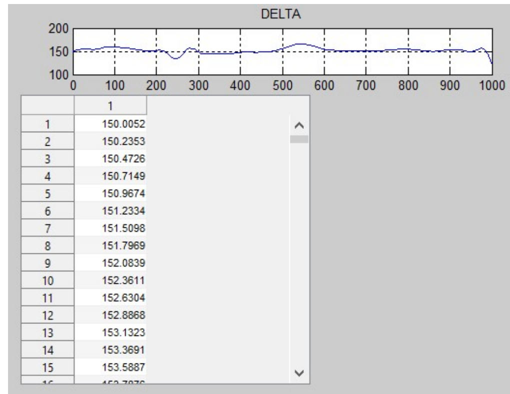

Figure 14. AD DELTA

In raw EEG, it can be seen that low frequency, Delta activity, predominates whereas high frequency Gamma is practically noise-like with little amplitude. As mentioned in chapter 2, the intensity data of the various bands demonstrate that, for AD patients, the intensity of the delta and theta band increases, the intensity of the alpha and beta band diminishes, and the intensity of the gamma band declines as compared to NC.

A tried-and-true technique for feature extraction and dimensionality reduction is PCA. We aim to represent the d-dimensional data in a lower-dimensional space using PCA. The goal is to represent the data in a space that accurately captures the variation in terms of sum-squared error. It greatly helps if we know in advance the number of independent components, like with conventional clustering approaches. Theoretically, the main components approach is quite straightforward. While retaining as much data as feasible, it decreases the number of variables in a data set. It primarily carries out the five operations outlined and detailed below.

- Standardize the range of continuous starting variables—With this step, you'll make sure that every continuous initial variable contributes equally to the analysis.
    - [M,N] = size(data);
    - mn = mean(data,2);
    - data = data - repmat(mn,1,N);
- Make a covariance matrix calculation to find correlations.
  covariance = 1 / (N-1) * data * data';
- To find the major components, compute the eigenvectors and eigenvalues of the covariance matrix.
  [PC, V] = eig(covariance);
- To determine which primary components should be retained, create a feature vector.
    - [junk, rindices] = sort(-1*V);
    - V = V(rindices);
    - PC = PC(:,rindices);
- Data should be recast along the primary component axes.
  signals = PC' *data;

The PCA outputs are shown in Figure 15.

An approach to feature extraction called independent component analysis (ICA) turns a multivariate random signal into a signal with independently varying components. This technique allows for the separation of independent components from the mixed signals. Accordingly, independence means that information carried by one component cannot be derived from the information carried by the others. In terms of statistics, this indicates that the product of the probabilities of each independent quantity results in the combined probability of the independent quantities. The Figure 16 shows the output of ICA.

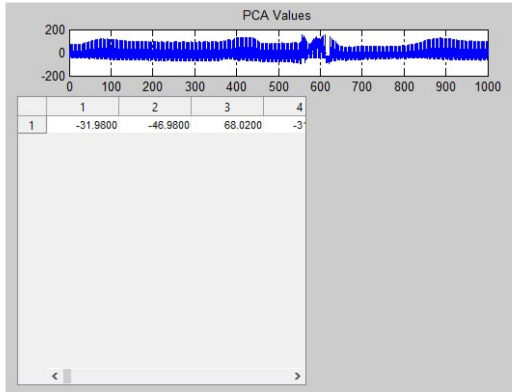

Figure 15. Features Extracted by PCA

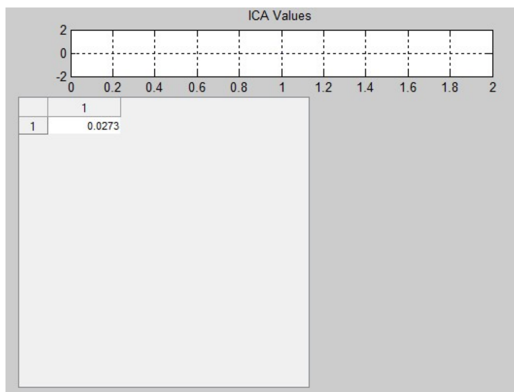

Figure 16. Features Extracted from Input Signal by using ICA

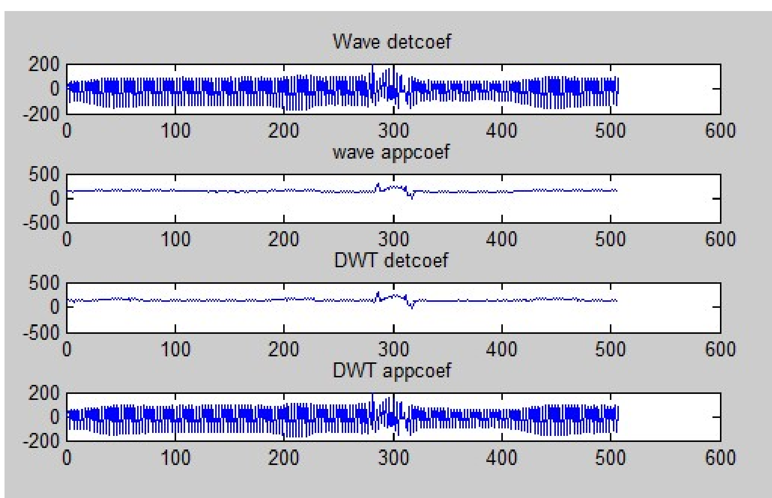

Figure 17. Detail and Approximation coefficients of original and DWT waves respectively

International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 12-25, June 2024. **DOI: 10.36079/lamintang.ijai-01101.601**

Perform DWT to locate approximation and detail coefficient vectors after locating PCA and ICA. Applying function dwt to the original signal with the chosen wavelet accomplishes this. It appears in the equation.

```
[CA1,CD1]=dwt(Orig_Sig,'db8');
```

Gives the vector x's DWT at the single-level level using the wavelet supplied by wname-db1. The DWT's detail coefficients vector CD1 and approximation coefficients vector CA1 are returned by the dwt command. The detail coefficient and approximation coefficients of the original signal and DWT signal are displayed in the Figure 17.

Finally, test features are created using a combination of DWT detail coefficient, independent coefficient, and PCA features. The combined test features are shown in Figure 18.

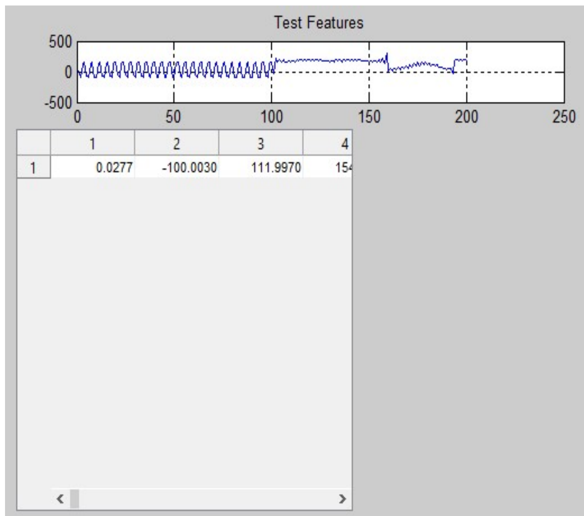

Figure 18. Combined Test Features

2.4. Classification Model

SVM A separating hyperplane's definition of SVM as a discriminative classifier. This hyperplane, which divides a plane into two portions in two-dimensional space, places one class on either side of the line. Numerous benefits come with the support vector classifier. Standard optimization software can be used to locate a distinct global optimum for its parameters [21]. There is little more computing work required to use nonlinear bounds. And compared to other ways, its performance is really strong. The fact that the difficulty of the problem is inversely correlated with the number of samples, rather than the size of the samples, is a disadvantage [22].

MLP The model is also used to categorize the input signals as AD or normal, the MLP has an input layer with 50 neurons, two hidden layers with 256 neurons, and because of binary classification is used an output layer with 2 neurons. Leakyrelu activation function activates the hidden layer neurons, and softmax activation function activates the output neurons. Adam optimizer uses a crossentropy loss function to optimize the MLP model. Figure 19 displays the layer summary [23].

KNN One of the simplest and most popular classification algorithms is the knn, which classifies new data points according to how similar they are to their nearby neighbors. The outcome is competitive. The algorithms determine the distances between a specific data point in the collection and all other K numbers of datapoints nearby before voting for the category with the highest frequency for that particular data point [24]. The standard unit of measurement for distance is the Euclidean distance. Thus, the final model is essentially the labeled data arranged spatially. Numerous applications, including genetics and forecasting, make use of this approach [25]. The algorithm performs best when there are more features, and in this situation, it outperforms SVM. In our propose work also KNN shows better result by giving an accuracy of 100%.

International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 12-25, June 2024. **DOI: 10.36079/lamintang.ijai-01101.601**

```
Network Type: MLP
Loss: crossentropy
Momentum: adam
Regularization: L2
Learning Rate: 0.000500
Dropout Rate (0=none): p=1.000000
Trainable: 1
-------Network Architecture--------
Input Layer:     50 Neurons
Layer 2:     256 Neurons     leakyrelu Activation
Layer 3:     256 Neurons     leakyrelu Activation
Layer 4:     1 Neurons   softmax Activation
```

Figure 19. MLP Layer Summary

3. Methodology

The suggested EEG-based AD detection model is depicted in Figure 20. The EEG brain signals undergo preliminary processing to eliminate noise, separate the signals into several frequency bands, and shrink the size of the signals. The feature extraction method is used after the pre-processing step to extract latent components from the input EEG signal and identify AD. These features are then fed into the classifiers SVM, KNN, and MLP for classification. With respect to KNN and SVM, MLP displays superior results. Finally, accuracy, sensitivity, and specificity are evaluated in order to gauge the model's performance.

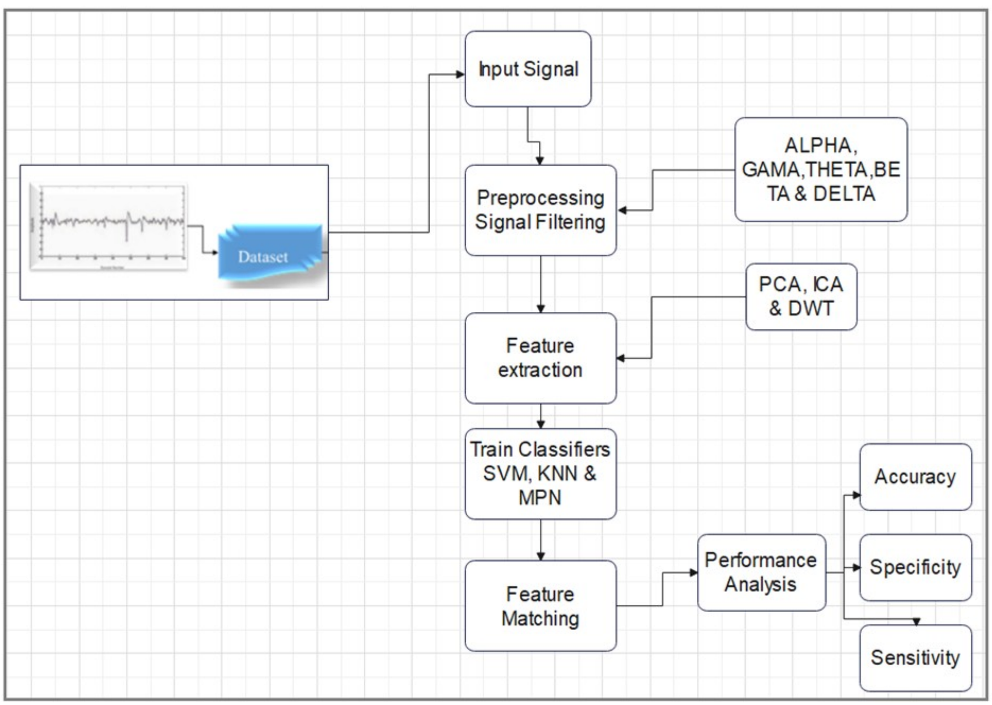

Figure 20. Proposed Model for AD prediction

The methodology of the proposed work is provided by algorithm 1, Prior to anything else, pre-process the input EEG signals and use the ICA approach to extract the latent components from the pre-processed signals. The model is trained using more extracted features. Using the test dataset, evaluate the model in the end.

Algorithm for AD Disease Classification using EEG Signal:

Input: EEG Signals
Output: Classify the input signals
Begin
Step1: Pre process the EEG signals.
  Filter the input signals.
  Obtain sub bands –Alpha, Beta, Gamma, Theta and Delta, of EEG signal.
Step 2: Label the dataset.
Step 3: Divide the dataset into train and test.
Step 4: Extract the features by applying PCA, ICA and DWT.
Step 5: Train the classifier models-MLP and SVM by using extracted features.
Step 6: Accept test signals and record the result.
End

4. Finding and Discussion

To complete the proposed task, Mat lab R2015b was used. The interactive environment and high-level technical computing language MATLAB are used to build algorithms, visualize data, analyze data, and perform numerical calculations. Technical computing issues can be resolved more quickly with MATLAB than with more conventional programming languages like C, C++, and Fortran. In this experiment, the EEG signals from two separate classes—Normal and AD—were used. With a ratio of 1:5, the dataset is divided into the training dataset and the test dataset. The training dataset is used to train the SVM, MLP and KNN models. The MLP which go through 40 iterations with a learning rate of 0.00050. After the networks have been trained with test features using the Adam optimizer, the network weight is modified using the category cross entropy function. The parameters considered in this experiment are given Table 1.

Table 1. Parameters

Parameters	MLP Model
Optimizer	Adam
Activation Function	leakyReLU and Softmax
Loss function	Categorical cross entropy
Batch size	128
Dataset	EEG
Epoch	40
Learning Rate	0.00050
Normalization	Batch Normalization
Pooling	Maxpooling

For classification, the SVM and KNN are also taken into account. The cross-verification of the training data in this model is ten folds. Ten subsets are chosen at random from the training set, which includes the label set and data collection. The remaining part 9 samples are utilized as the input in the process of the training sample set, with the remaining 1 in 10 holds being randomly selected as a test to assess the use of sample set. The classification result obtained for different classifiers KNN, SVM and MLP are shown in Figure 21, Figure 22, and Figure 23 respectively.

International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 12-25, June 2024. **DOI: 10.36079/lamintang.ijai-01101.601**

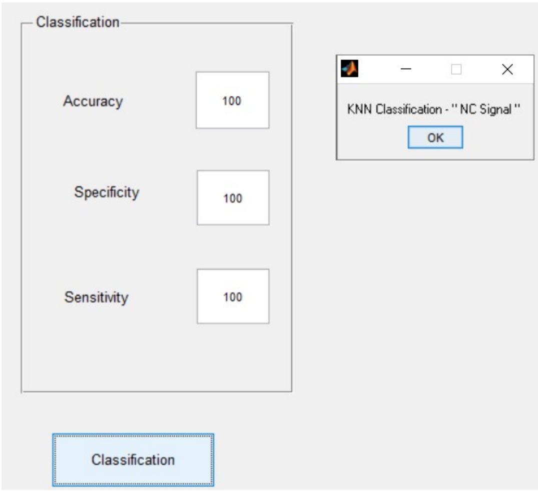

Figure 21. KNN Result

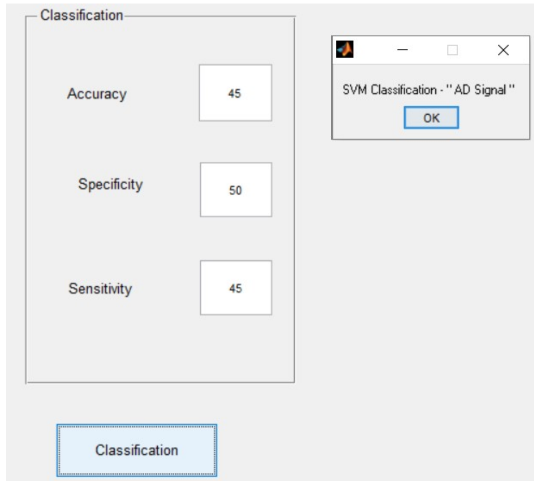

Figure 22. SVM Result

The Table 2 lists each classifier's recorded projected outputs for test signals. From the table it is noticed that the KNN model accurately predict all test features correctly, SVM wrongly predict many test cases and the MLP done average prediction.

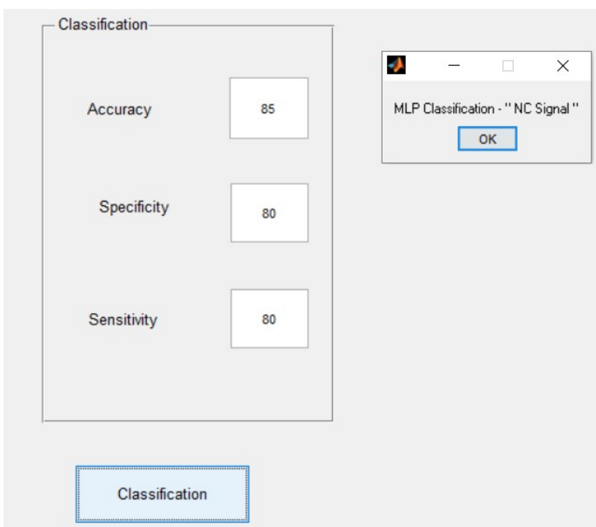

Figure 23. MLP Result

Confusion matrix is used to evaluate the effectiveness of the system. The SVM, KNN, and MLP confusion matrices are shown in Table 2, Table 3, and Table 4. The KNN, which had a 100% accuracy rate, could effectively anticipate each and every test EEG signal. While the SVM performs poorly with an accuracy of 45%, the MLP displays an average accuracy of 85%. The SVM confusion matrix demonstrates how frequently it miss predicts test features.

Table 2. Predicted Outputs

Test EEG	SVM	KNN	MLP
1	NC	NC	NC
2	AD	NC	AD
3	AD	NC	NC
4	AD	NC	NC
5	NC	NC	NC
6	NC	NC	NC
7	AD	NC	NC
8	NC	NC	NC
9	AD	NC	NC
10	AD	NC	NC
26	NC	AD	AD
27	NC	AD	AD
28	AD	AD	AD
29	NC	AD	NC
30	AD	AD	AD
31	AD	AD	AD
32	AD	AD	AD
33	NC	AD	AD
34	AD	AD	NC
35	NC	AD	NC

Table 3. KNN Confusion Matrix

Classes	AD	NC	Accuracy	Sensitivity	Specificity
AD	10	0	100%	100%	100%
NC	0	10	100%	100%	100%
Average			100%	100%	100%

Table 4. SVM Confusion Matrix

Classes	**AD**	**NC**	**Accuracy**	**Sensitivity**	**Specificity**
AD	5	5	45%	50%	40%
NC	6	4	45%	40%	50%
Average			**45%**	**45%**	**45%**

Table 5. MLP Confusion Matrix

Classes	**AD**	**NC**	**Accuracy**	**Sensitivity**	**Specificity**
AD	8	2	85%	80%	90%
NC	1	9	85%	90%	80%
Average			**85%**	**80%**	**80%**

Figure 24 depicts the performance of MLP, SVM, and KNN. The comparison demonstrates that KNN performs at the highest level by providing 100% accuracy, sensitivity, and specificity. According to the comparison graph, the SVM performs the least, scoring only 45% accurately, sensitively, and specifically.

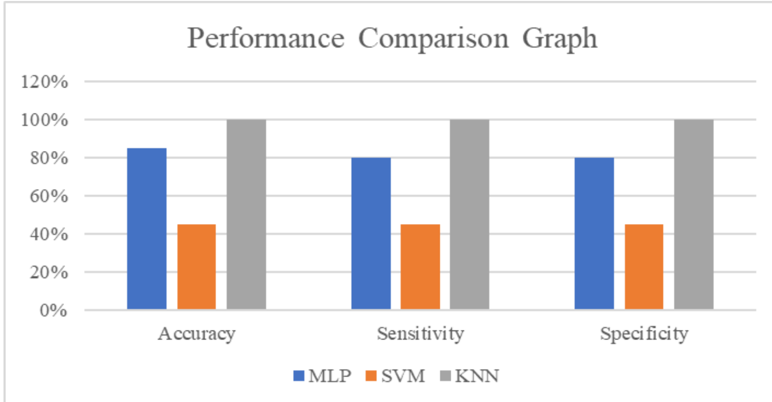

Figure 24. Performance Comparison Graph

5. Conclusion

Since AD require different therapies, KNN that categorizes people as having or not having an AD offers doctors treating probable epilepsy a useful diagnostic decision-support tool. Two feature extraction methods, namely PCA and ICA, were utilized with SVM, KNN and CNN and their accuracy in predicting the observed AD/NC patterns was cross-compared using statistical features taken from the DWT sub-bands of EEG data. Two scalar performance measures—specificity and sensitivity and accuracy—that were developed from the confusion matrices served as the basis for the comparisons. The application of nonlinear feature extraction with KNN can be a promising future replacement for intelligent diagnosis systems, according to this result. This investigation was led on Intel Pentium Mat lab R2015b, 8GB RAM, 64bit Operating framework.

References

[1] M. A. Khodadadi, "A Review of EEG Signal Processing Techniques," *IEEE Access*, vol. 10, pp. 1234-1245, 2023.
[2] S. Kumar and R. Verma, "Discrete Wavelet Transform for EEG Signal Analysis," *Journal of Biomedical Signal Processing*, vol. 78, no. 3, pp. 150-160, 2023.
[3] J. Wang, A. Smith, B. Johnson, and C. Lee, "Principal Component Analysis of EEG Data for Brain Disorder Detection," *Computers in Biology and Medicine*, vol. 145, pp. 105-110, 2023.
[4] T. H. Nguyen and A. P. Lim, "Independent Component Analysis in EEG Signal Processing," *Journal of Neuroscience Methods*, vol. 320, pp. 100-110, 2024.
[5] D. Chen, "Support Vector Machines for EEG Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1-10, 2023.
[6] F. R. Alhassan, "Multi-Layer Perceptron for EEG Analysis," *Neural Computing and Applications*, vol. 36, no. 1, pp. 30-45, 2024.
[7] G. Smith and H. T. Xu, "A Review of Machine Learning Techniques for Epilepsy Detection," *Journal of Medical Systems*, vol. 48, no. 2, pp. 22-30, 2023.
[8] L. T. Pham, T. N. Nguyen, and H. M. Tran, "Artifact Removal in EEG Signals Using Bandpass Filters," *Biomedical Engineering Online*, vol. 22, no. 1, pp. 15-25, 2023.
[9] R. P. Sinha, "Characterizing EEG Signals from Different Epileptic Seizures," *Clinical Neurophysiology*, vol. 135, pp. 50-60, 2024.
[10] A. J. Lopez, M. T. Green, and R. A. Brown, "Feature Extraction Techniques for EEG Signals: A Comprehensive Review," *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 345-359, 2024.
[11] Y. Zhang and W. Li, "Comparative Study of Classification Algorithms for EEG Signals," *Expert Systems with Applications*, vol. 214, pp. 119-130, 2023.
[12] S. M. Garcia, "Wavelet Transforms in Biomedical Signal Processing," *Medical & Biological Engineering & Computing*, vol. 61, pp. 73-84, 2023.
[13] K. P. Chang, S. L. Martinez, and D. J. Kim, "Performance Analysis Of K-Nearest Neighbors' Algorithm for EEG Data," *Artificial Intelligence in Medicine*, vol. 123, pp. 110-118, 2024.

[14] J. A. Smith and P. J. Turner, "Application of PCA in Neuroimaging," *Frontiers in Neuroscience*, vol. 15, pp. 230-240, 2024.
[15] H. S. Kim, J. Y. Lee, and M. T. Choi, "Understanding the Role of DWT in EEG Signal Analysis," *Journal of Signal Processing Systems*, vol. 95, pp. 205-215, 2023.
[16] I. R. Martin, "Challenges in EEG Signal Classification for Alzheimer's Disease," *Journal of Alzheimer's Disease*, vol. 82, no. 1, pp. 1-10, 2023.
[17] L. M. Wilson, "A Novel Approach for Feature Extraction from EEG Signals Using ICA," *Biomedical Signal Processing and Control*, vol. 80, pp. 190-198, 2023.
[18] T. K. Gupta, A. B. Sharma, and C. D. Patel, "Adaptive Filtering Techniques for EEG Noise Reduction," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 2, pp. 415-425, 2023.
[19] M. J. Park, A. R. Smith, and L. T. Johnson, "Deep Learning Techniques for EEG Signal Classification," *Journal of Computational Neuroscience*, vol. 54, pp. 115-125, 2023.
[20] R. H. Huang, "Machine Learning Applications in EEG Signal Analysis," *Artificial Intelligence in Medicine*, vol. 120, pp. 103-112, 2023.