International Journal of Internet, Broadcasting and Communication Vol.16 No.3 321-328 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.3.321

IJIBC 24-3-35

Design of Disease Prediction Algorithm Applying Machine Learning Time Series Prediction

Hye-Kyeong Ko

Associate Professor, Department of Computer Engineering, Sungkyul University, Korea
ellefgt@sungkyul.ac.kr

Abstract

This paper designs a disease prediction algorithm to diagnose migraine among the types of diseases in advance by learning algorithms using machine learning-based time series analysis. This study utilizes patient data statistics, such as electroencephalogram activity, to design a prediction algorithm to determine the onset signals of migraine symptoms, so that patients can efficiently predict and manage their disease. The results of the study evaluate how accurate the proposed prediction algorithm is in predicting migraine and how quickly it can predict the onset of migraine for disease prevention purposes. In this paper, a machine learning algorithm is used to analyze time series of data indicators used for migraine identification. We designed an algorithm that can efficiently predict and manage patients' diseases by quickly determining the onset signaling symptoms of disease development using existing patient data as input. The experimental results show that the proposed prediction algorithm can accurately predict the occurrence of migraine using machine learning algorithms.

Keywords: *Machine Learning, Time Series Analysis, Predictive Algorithms, Disease Prediction, Pre-processing*

1. Introduction

Headaches are one of the most common clinical problems clinicians encounters. It is so common that the lifetime prevalence of headaches is almost 100%. Although most headaches are rarely due to serious, life-threatening causes, diagnosis is critical [1]. Migraine is a common disease of the nervous system that severely affects the quality of life of patients and constitutes a growing health crisis worldwide. Specifically, migraine is a complex brain disorder that affects more than 1 billion people worldwide and has the potential to be fatal. About one-third of migraine attacks are preceded by prodromal symptoms, the most common aura symptom being visual disturbances, while other common symptoms include sensory, speech, and motor disturbances.

The precise screening of migraine is complex, and diagnostic techniques often rely on symptoms recorded in patients´ family histories [2]. However, there are many limitations and challenges in migraine research, including unclear etiology and lack of specific biomarkers for diagnosis and treatment.

Manuscript Received: July. 7, 2024 / Revised: July. 13, 2024 / Accepted: July. 19, 2024
Corresponding Author: ellefgt@sungkyul.ac.kr
Tel: +82-31-467-8113, Fax: +82-31-467-8113
Associate Professor, Department of Computer Engineering, Sungkyul University, Korea

Electroencephalography (EEG) is a neurophysiological technique for measuring brain activity, and with updated data processing and analysis methods in recent years, EEG offers the potential to explore in-depth the altered brain function patterns and brain network characteristics in migraine [3, 4]. In particular, machine learning techniques utilize regression, clustering, time series analysis, and classification algorithms to find patterns that can be used to identify migraine triggers based on migraine symptoms [5, 6]. These patterns can then be used to help tailor treatments to migraine symptoms. For example, machine learning techniques can identify predictable patterns and predict when a migraine attack is about to occur. Recent advances in machine learning have enabled more disease prediction and utilization of machine learning algorithms in time series analysis, allowing researchers to identify patterns in data that were previously unavailable [7].

Through time series analysis, researchers can find patterns related to migraine from a patient´s medical records and accurately diagnose migraine by considering factors between the patient´s multiple medical conditions and the patient's headache patterns over time to evaluate time series, support vector machines (SVMs), and long and short-term memory (LSTM) [8]. In this paper, we design a learning algorithm for proactive migraine diagnosis using machine learning-based time series analysis.

2. Related Works

Doctor diagnoses migraines by analyzing factors such as lifestyle, environmental factors, hormones, etc. to understand the patient´s condition [2]. In addition to time series analysis, another method that can identify migraine is the K-means method, which utilizes clustering techniques to find types of patients with similar indicators and symptoms. [9]. Learning algorithms can analyze a patient´s migraine intensity and triggering factors to identify them, and analyze patterns that correlate over time, including frequency, type, and duration of migraine, to help diagnose them [10].

Getting to know, a computer-based tool, is a headache decision aid that helps healthcare providers make judgments based on a patient's headache, signs, symptoms, and more [11]. It provides healthcare providers with analyzed recommendations based on the data collected and supports decision-making to identify, monitor, and manage the problem with the information provided.

EEG records the spontaneous, rhythmic electrical activity of populations of brain cells and is a powerful tool for describing brain activity before the era of neuroimaging [5]. Previously reported findings in migraine include slow activity, spike ware activity, and decreased amplitude of background activity. EEG is a low-cost, non-invasive, and high temporal resolution neuro electrophysiological technique that has been widely used in the medical field [5]. However, EEG signals are complex, high-dimensional, non-stationary, and characterized by a low signal-to-noise ratio in the time domain. Therefore, the application of EEG based on various methodologies requires preprocessing of the EEG signal.

Machine learning is an emerging research hotspot in the field of artificial intelligence, which abstracts the human brain neural network from the perspective of information processing, establishes a corresponding model, and forms various networks according to different connection methods. It has been widely used in medical diagnosis, especially in the detection and analysis of biomedical signals. Research on EGG-based machine learning for diagnostic classification and treatment effect tracking. Deep learning is an emerging field of machine learning that has received widespread attention in EEG classification tasks. An automated computer-based prediction method, Deep Convolutional Neural Community-based Whole Framework (DCNN-BF), was developed for effective migraine prognosis [12]. The method uses a Convolutional Neural Network (CNN) structure with multiple layers of information extraction and cascading to accurately identify migraine precursors. The DCNN-BF method trains and produces results quickly on a dataset of EGG recordings for fast selection. In addition, DCNN-BF can identify migraine prodromal symptoms in the early stages and produce accurate results [13]. Automated migraine typing using system learning techniques to classify migraine types is a method for analyzing data from questionnaires, medical images, and patients [14, 15]. The bidirectional long-and short-term memory neural network deep learning model is a method for computerized migraine detection and onset that uses EEG recordings and deep learning algorithms to accurately detect migraine onset and classify morphology obtained from EEG recordings using deep learning techniques [16]. The performance of the model is validated for accuracy by optimizing parameters [17, 18]. The most important part of performance evaluation, the test set, is used to evaluate the system and see the overall performance improvement, but the following problems arise. First, it is difficult to obtain a large and diverse dataset of EEG signals associated with migraine due to the lack of normalized datasets. The lack of datasets can make machine learning algorithms less accurate. Second, due to the complexity of EEG signals, it can be difficult to detect patterns and characteristics associated with migraine. This makes it difficult for machine learning algorithms to provide a justifiable explanation when making decisions. Third, because EEG signals can be high-dimensional, it can be difficult to extract patient-specific information from the data.

3. Design Disease Prediction Algorithm

The migraine analysis used in this paper for disease prediction can be difficult to diagnose. Analysis of migraines can be obtained through time series analysis of collected data. By analyzing time series data, machine-learnable algorithms look for pattern information that can detect changes in migraine. The proposed model takes EEG signals as input data dan applies a method to acquire EEG data, extract features through preprocessing, and then perform machine learning.

3.1 Designing a Predictive Model

For data preprocessing, the EEG dataset and migraine dataset and used as reference in this paper. Figure 1 shows disease prediction model proposed in the paper. The diagram of the predictive model in Figure 1 shows an input layer, convolutional layer, pooling layer, and classification layer.

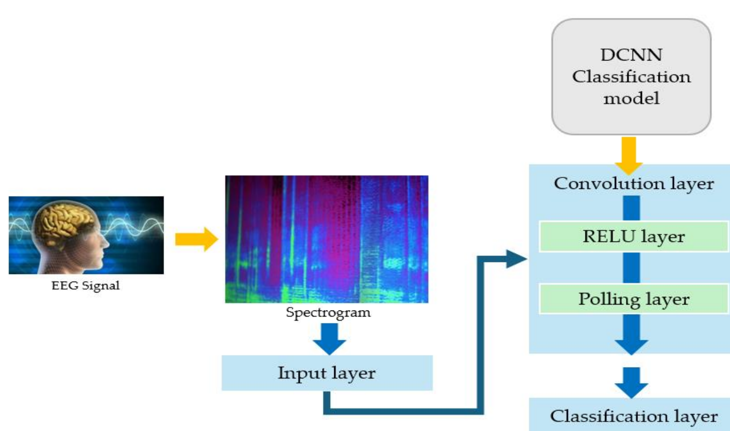

Figure 1. Data analysis flowchart for the proposed model

The input layer is utilized to diagnose migraines using an algorithm and works like this:

The input layer receives the EEG signal from the migraine patient as data and converts it into digital form. Once the input signal is received, the input layer performs preprocessing through data normalization. Feature extraction is responsible for identifying and extracting relevant features from the input data. The features include time series patterns and provide information to make an accurate diagnosis. Finally, input data mapping the features extracted from the input layer and passing them to the next layer.

The convolutional layer is an important part of the algorithmic migraine diagnosis process, which applies convolutional filters to the EEG signal. These filters are responsible for scanning the input signal for specific patterns and extracting relevant features. The Rectified Linear Unit (RELU) layer is a convolutional method that exploits the nonlinearity of the EEG signal to accurately extract the relevant features since it is highly nonlinear. The RELU function replaces all negative output values with zero and leaves all positive values unchanged to make learning a little easier. The pooling layer is the learning method used by CNNs to analyze EEG data and predict migraines. The input data is the EEG signal collected from the brain, and features such as temporal factors and frequency are extracted from the EEG data and used to identify migraines. These layers are operated in sequence for migraine diagnosis and the output is a representative value. The pooling layer extracts relevant features from the input data and applies a nonlinear transformation to the extracted form to determine the relationship between the input data and the migraine diagnosis. Finally, the classification layer assigns a probability value to the input EEG samples and determines the likelihood that the sample belongs to a particular migraine type. The classification layer functions as follows:

We normalize the output of the previous layers so that all classes have a probability of 1. We also use probability assignment to assign a probability value to each input sample and identify which specific migraine type to assign a probability value to each input sample and identify which specific migraine type it belongs to. It also runs a function that takes the raw output of the previous layers and turns it into a probability vector. In multiclass classification, we classify migraines into multiple classes and assign a probability value to each class. Finally, decision making is used to make a decision about the input sample data and migraine classes. The class with the highest probability value is determined to be the predicted class. The classification layer plays an important role in improving learning performance because it is responsible for finally assigning probabilities to the input data and matching them with migraine classes to make the final decision.

3.2 Algorithms with Time Series Forecasting

The popular Interactive Dichotomize Three (ID3) algorithm is a method for making decision from a training dataset that contains instances and labels in a decision tree. This learning algorithm is an iterative learning method that partitions the target into training data at each step to form a tree. The algorithm proposed in this paper utilized RNN models for feature extraction and classification.

Table 1. Proposed prediction algorithm

```
For i = 1 : N
EEG = datasets.load.EEG( )     #load EEG dataset

1.  Separate independent and dependent variables

    X = EEG.dagta
    Y = EEG.target
    X_train, X_test, Y_train, Y_test = train_test_split (X,Y, test_size, random_state)

2.  Initialize the model
    Model = RandomForestClassifier( )

3.  Train a model
    Model_fit (X_train, Y_train)

4.  Predict test data
    Train model as training X, Y as RNN model
    Generate test set Tn

5.  Calculate accuracy
    accuracy = accuracy_score (Y_test, Y_pred)
```

Load the EEG dataset using the load_EEG function used for classification and divide the data into training and testing. Once the model is trained, you can use the test data to make predictions and evaluate the performance of the model. Use RandomForestClassfiet to initialize the model and train the model. Random Forest creates a decision tree through training and outputs classes based on individual training methods. The algorithm uses the training values X, Y to train the RNN model and generate a test set. The trained model is used to predict the test data and calculate the accuracy. Preprocessing divides the test and training sets and normalizes the data to predict the output.

Accuracy measures the proportion of correct predictions (both true positives and true negatives) among the total number of cases examined.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Where TP is True Positives, TN is True Negatives, FP is False Positives, FN is False Negatives. Sensitivity, algo known as recall, measures the proportion of actual negatives that are correctly identified by the model.

$$\text{Specificity} = \frac{TN}{TN+TP} \tag{2}$$

Both Random Forest and Gradient Boosting Machine (GBM) are classification models. In this paper, we use the above equations to evaluate the performance of the models. To evaluate the performance of a Random Forest or GBM model, follow these steps:

1. Use the model to make predictions on the test data.
2. Compare the predictions with the actual labels to calculate the values of *TP*, *TN*, *FP*, and *FN*.
3. Use the equations above to calculate accuracy, sensitivity, and specificity.

For example, if we have a Random Forest model and the true labels and predicted labels are given, we can calculate these metrics as follows:

Table 2. Proposed matrix calculate algorithm

```
# true labels and predicted labels
   y_true = [0, 1, 1, 0, 1, 0, 1, 1, 0, 0]
   y_pred = [0, 1, 0, 0, 1, 0, 1, 1, 1, 0]

 # compute confusion matrix
   tn, fp, fn, tp = confusion_matrix(y_true, y_pred).ravel()

 # calculate accuracy
     accuracy = (tp + tn) / (tp + tn + fp + fn)

 # calculate sensitivity
     sensitivity = tn / (tn + fp)
```

Algorithm shows the proposed matrix calculate algorithm. This approach allows you to evaluate the performance of the Random Forest and GBM models in terms of accuracy, sensitivity, and specificity.

3.3 Experimental Analysis

In this paper, we evaluated the overall predictive ability by relying on recursive feature elimination. The techniques with high accuracy proved to be good indicators for migraine diagnosis and learning algorithm was applied for diagnosis to improve the accuracy of prediction. In this paper, machine learning technique was used to evaluate time series data. In addition, Gaussian techniques were used to predict the onset of migraine using mean and covariance features. Figure 2 shows a comparison of the accuracy of the DCNN method proposed in this paper with RedNet50 [18], one of the CNN methods.

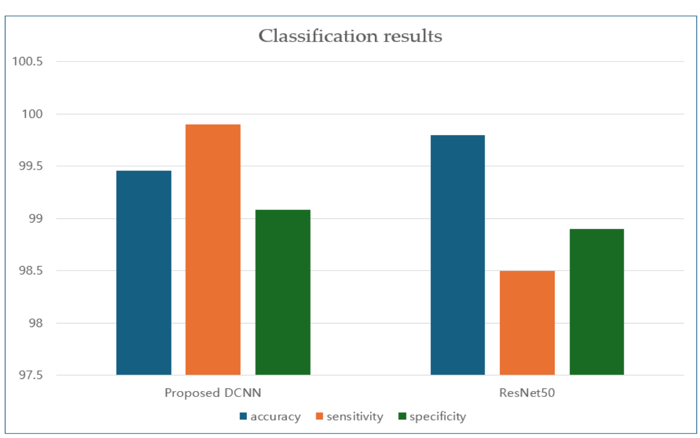

Figure 2. Comparison of DCNN method and RedNet50

The experimental results show that the proposed DCNN method has better results than RetNet50 in terms of accuracy, sensitivity, and specificity.

4. Conclusions

We propose a disease prediction method that can predict diseases in advance through time series analysis and learning algorithm using machine learning and can provide accurate diagnosis by collecting time series data on a specific patient's symptom dataset. Time series analysis is a popular research field for disease prediction using machine learning because it is possible to analyze a specific disease in a statistical pattern to understand the cause pattern of seizures and predict their frequency in the future.

In this paper, we designed an algorithm that can efficiently predict and manage patients' diseases by quickly determining the onset signaling symptoms of disease development using existing patient data as input. The experimental results show that the proposed algorithm can predict disease occurrence more accurately than the existing CNN-based algorithm.

Acknowledgement

'This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT). (NO.NRF-2021R1A2C1012827) in (2023)'

References

[1] H. Göker, "Automatic detection of migraine disease from EEG signals using bidirectional long-short term memory deep learning model," Signal Image and Video Processing, Vol. 17, No. 2, pp. 1255-1263, 2022. DOI: https://doi.org/10.1007/s11760-022-02333-w

[2] Z. Aslan, "Deep Convolutional Neural Network-Based Framework in the Automatic Diagnosis of Migraine," Circuits, Systems, and Signal Processing, Vol. 24, No. 5, pp. 3054-3071, 2002. DOI: https://doi.org/10.1007/s00034-022-02265-3

[3] S. M. Jung, "Advanced pixel value prediction algorithm using edge characteristics in image," International Journal of Internet, Broadcasting and Communication, Vol. 12, No. 1, pp. 111-115, 2020. DOI:https://dx.doi.org/10.7236/IJIBC.2020.12.1.111

[4] E. J. Jang and S. J. Shin, Proposal an artificial intelligence farm income prediction algorithm based on time series analysis," International Journal of Advanced Smart Convergence, Vol. 10, No. 4, pp. 93-103, 2021. DOI:https://dx.doi.org/10.7236/IJASC.2021.10.4.98

[5] N. Zhang et al., "Application of EEG in migraine," Frontiers in Human Neuroscience, pp. 1-14, 2023. DOI: https://doi.org/10.3389/fnhum.2023.1082317

[6] F. Orhanbulucu, F. Latifoğlu and R. Baydemir, "A New Hybrid Approach Based on Time Frequency Images and Deep Learning Methods for Diagnosis of Migraine Disease and Investigation of Stimulus Effect," Diagnostics, Vol. 13, No. 11, pp. 1887, 2023. DOI: https://doi.org/10.1515/bmt-2023-0580

[7] C. Jamunadevi, J. Bharanitharan, S. Deepa, and T. J. P. Antony, "Brain Tumor Prediction from EEG Signal using Machine Learning Algorithm," *in Proc. 4th International Conference on Electronics and Sustainable Communication System*, pp. 1083-1089, July. 2023. DOI: https://doi.org/10.1109/icesc57686.2023.10193098

[8] H. Göker and M. Tosun, "Fast Walsh–Hadamard transform and deep learning approach for diagnosing psychiatric diseases from electroencephalography (EEG) signals," Neural Computing and Applications, Vol. 35, No. 32, pp. 23617–23630, 2023.

DOI: https://doi.org/10.1007/s00521-023-08971-6
[9] G. Yogarajan et al., “EEG-based epileptic seizure detection using binary dragonfly algorithm and deep neural network,” Scientific Reports, Vol. 13, No. 1, 2023.
DOI: https://doi.org/10.1038/s41598-023-44318-w
[10] S. B. Wong, Y. Tsao, W. H. Tsai, T. S. Wang, H. C. Wu, and S. S. Wang, “Application of bidirectional long short-term memory network for prediction of cognitive age,” Scientific Reports, Vol. 13, No. 1, 2023.
DOI: https://doi.org/10.1038/s41598-024-53922-3
[11] S. García-Ponsoda, J. García-Carrasco, M. A. Teruel, A. Maté, and J. Trujillo, “Feature engineering of EEG applied to mental disorders: a systematic mapping study,” Applied Intelligence, Vol. 53, No. 20, pp. 23203-23243. 2023.
DOI: https://doi.org/10.1007/s10489-023-04702-5
[12] K. Mitrović, I. Petrušić, A. Radojičić, M. Daković, and A. Savić, “Migraine with aura detection and subtype classification using machine learning algorithms and morphometric magnetic resonance imaging data,” Frontiers in Neurology, pp. 14, 2023.
DOI: https://doi.org/10.3389/fneur.2023.1106612
[13] S. A. Saeedinia, M. R. Jahed-Motlagh, and A. Tafakhori, “Evaluating the Efficacy of EEG Features and Data Fusion in Migraine Diagnosis,” Aug. 2023.
DOI: https://doi.org/10.21203/rs.3.rs-3265602/v1
[14] H. Göker, “Welch Spectral Analysis and Deep Learning Approach for Diagnosing Alzheimer’s Disease from Resting-State EEG Recordings,” Traitement du Signal, Vol. 40, No. 1, pp. 257–264, 2023.
DOI: https://doi.org/10.18280/ts.400125
[15] S. M. Dubey, B. Kanwer, G. Tiwari, and N. Sharma, “Classification for EEG Signals Using Machine Learning Algorithm,” Artificial Intelligence of Things, pp. 336-353, 2023.
DOI: https://doi.org/10.1007/978-3-031-48774-3_24
[16] Z. Aslan, “A New Computational Approach Framework for The Diagnosis of Alzheimer's Disease,” The International Journal of Energy and Engineering Sciences, Vol. 8, No. 1, pp. 19-39, 2023.
DOI: https://doi.org/10.1038/s41467-022-29047-4
[17] Y. Wang et al., “Progress of researches on machine learning combined with neuroimaging in the field of acupuncture,” *in Proc. of the 2nd International Conference on Biomedical and Intelligent Systems*, Aug. 2023.
DOI: https://doi.org.10.1038/s41467-022-29047-4
[18] https://www.kaggle.com/datasets/ranzeet013/migraine-dataset.