Sriwijaya Journal of Informatic and Applications
Vol. 3, No. 1, January 2022, pp. 38-47
e-ISSN 2807-2391

Spelling Detection based on P300 Signal with Convolutional Neural Network (CNN) Algorithm

Kgs. M. Rusdiansyah Muharrom [a,1] , Rifkie Primartha [b,2*]

[a] Undergraduate Student of Informatics Engineering, Faculty of Computer Science, Sriwijaya University
[b] Lecturer of Informatics Engineering Department, Faculty of Computer Science, Sriwijaya University
Jl. Srijaya Negara, Bukit Besar, Palembang 30139, Indonesia
[1] rrrusdiansyah@gmail.com; [2]rifkie@unsri.ac.id*
* corresponding author

ARTICLE INFO

Article history
Received 24 Nov 2021
Revised 30 Nov 2021
Accepted 12 Dec 2021

Keywords
Character Detection
P300 Signal
Convolutional Neural Network
Deep Learning

ABSTRACT

Brain Computer Interface (BCI) is a system that connects the human brain with the outside world for people who have motor skills disability problems. One form of utilization is the P300 speller which is used for character recognition or detection by classifying the P300 signal. The Convolutional Neural Network (CNN) method is a deep learning method that can be used to handle signal problems with ID-CNN. At the initial stage the data signal will be transformed and followed by a duplication process using RandomOverSampling because the amount of data in each class is not balanced. The data will be divided into training, validation, and test data. After that, a training with CNN will be conducted and followed by an evaluation to find the best model. The test results from this study are a good-fitting CNN model with an evaluation value consisting of an accuracy of 94.27%, precision of 90.64%, sensitivity / recall of 98.30%, and f-measure of 94.31%. Based on the test, the CNN method can be used and implemented in authentication detection based on the P300 signal.

1. Introduction

In the health sector, the implementation of bio-signal processing such as electrocardiography (ECG), electromyography (EMG), and electroencephalography (EEG) provides many benefits. EEG signals can be used in the rehabilitation of patients who have problems with neural motor disorders [1]. EEG signals are used in research on brain activity. EEG signals carry important information in research on brain function and neurological disorders [2]. The P300 signal is part of the EEG signal which is included in the Event Related Potential (ERP). The P300 signal appears in the EEG signal when visual stimuli are rare in the subject [3]. One solution that can be done so that people who have problems with motor skills and speech can communicate and relate to the outside world is with the help of technology. Brain Computer Interface (BCI) is a development of system technology that becomes a liaison between the human brain and the outside world for people who have problems with motor skills disabilities or commonly known as Pakrinson's disease [4]. BCI can be used for a variety of different applications, such as video surveillance (CCTV), wheelchair control, and character recognition [3]. In this study, the BCI application used was for character recognition or spelling detection. Spelling detection can be done by classifying the P300 signal based on its class. One way to classify the P300 signal is to use deep learning techniques or algorithms. Convolutional Neural Network (CNN) is one type of deep learning algorithm that can be used to solve signal problems called 1D Convolutional Neural Networks (1D CNNs) [5].

Research related to the classification of EEG signals has been carried out by several researchers. In [6] concluded that classification of EEG signals using a Support Vector Machine (SVM) can be applied for lie detection and produces an accuracy of 75% . In [3] concluded that their proposed method of character recognition based on the P300 signal produced better or the same performance than the previous method. Research entitled the design of a motor imagery (MI) EEG-based finger movement classification system using the Convolutional Neural Network (CNN) was conducted by [1] and resulted in an accuracy of 51.711%. Therefore, from the problems above, in this study, we will use the P300 signal with the Convolutional Neural Network (CNN) algorithm to detect spelling which can help users who have problems with motor and speech abilities, and calculate the performance of the algorithm in completing this research. In addition, this study will see whether the proposed method is better than the previous method.

2. Literature Study

a. Classification

Classification is one of the data mining techniques used to generate a rule that can classify or recognize new data that has not been studied from the results of studying a set of data [7]. Classification is defined as a process of grouping a data based on its class or category. Classification can be used for various applications, such as medical diagnosis, sales prediction, customer classification, and so on [8].

b. P300 Signal

The P300 signal is part of the electroencephalography (EEG) signal which is included in the Event Related Potential (ERP). The P300 signal appears in the EEG signal when less visual stimuli occur in the subject [3]. An EEG signal is a signal that records the spontaneous electrical activity of brain waves by measuring voltage fluctuations in brain neurons over a certain period of time [9]. Measurement of brain signals using EEG can be done in two ways, namely invasive and non-invasive. In invasive measurements, the electrode will be placed on the surface of the brain, but this measurement will be expensive. In non-invasive measurements, the electrodes will be placed on the surface of the scalp. Measurement in this way is preferred because it is cheaper and safer. P300 event-related potential (ERP) is one of the EEG signals commonly used in speller systems. The P300 speller is a non-invasive type of BCI system [3]. The character spelling paradigm or commonly called the P300 speller was first introduced by Farwell and Donchin. There is a 6 x 6 matrix of characters and numbers whose rows and columns each glow randomly. The user is asked to look and focus on the target character in the matrix. P300 will be obtained if the target character is in a glowing row or column. P300 found ERP was detected as a positive peak EEG voltage 300 ms after the stimulus. The classifier will detect one row and one column with P300 as the desired target character [10]. An example of the P300 speller matrix can be seen in Figure 1 below.

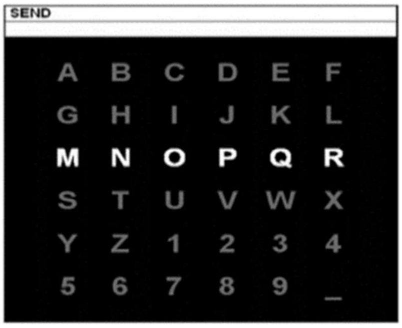

Fig 1. P300 *Speller Matrix*

c. RandomOverSampling

RandomOversampling is one of the oversampling techniques used to increase data in the minority class. This technique is used to increase the amount of data in the minority class by duplicating the minority class sample at random until the number of data for the majority and minority classes is balanced [11]. This technique is used because the amount of data in each class is not balanced.

d. Convolutional Neural Network

Convolutional Neural Network (CNN) is one type of deep learning algorithm which is the development of the Multilayer Perceptron (MPL) which is usually used for image data processing [12]. Deep Learning is part of machine learning that has the ability to define features or feature extraction automatically so that it is suitable for large data and has many features [8].

1. 1D Convolutional Neural Network (1D-CNN)

   CNN can also be used on 1-dimensional data (signal) called 1D-CNN which is a modified version of 2D-CNN. The attribute used in the signal data is the amplitude of the signal. In the 1D-CNN architecture, there are two layers used for the feature extraction and classification process. The first layer is the CNN layer (CNN-Layer) which is used for the feature extraction process where there is a convolution and sub-sampling (pooling) process. The second layer is a fully-connected layer used for the classification process [5].

2. Convolutional Layer

   Convolutional layer is a layer that performs the convolution process on CNN. The convolution process is the core process of the CNN algorithm in performing feature extraction from input data, which mostly performs deep learning computing [8]. In convolution, the process of multiplying the input with a filter or kernel produces an output (feature map).

3. Pooling Layer

   Pooling layer is a layer used to perform data dimension reduction (downsampling). The purpose of pooling is to make the data representation smaller, easier to manage so as to speed up computing because the parameters that need to be updated are getting smaller and overfitting conditions are easy to control [8]. The pooling methods commonly used are max pooling and average pooling. Max pooling will take the maximum value of the output value in the filter, while average pooling will take the average value [1].

4. Fully Connected

   Fully Connected Layer is a layer where every neuron in the previous layer is connected to the neuron in the next layer as in an ordinary neural network. This layer is used in the application of MLP (Multi Layer Perceptron) which has the aim of transforming the data dimensions so that they can be classified linearly [12]. This layer is used at the end of the network to carry out the classification process. MLP training can be carried out using the Back Propogation algorithm. This algorithm has two stages carried out, namely forward and backward calculations. Forward calculation functions to calculate the level of error between the predicted and actual results. The backward calculation serves to update the weights on the neurons in order to minimize errors [8].

5. Activation Function

   The activation function is a function that is used to decide whether a neuron should be activated or not in an artificial neural network [13]. This function allows the perceiver to adjust the pattern for non-linear data. ReLU activation or Rectified Linear Units layers is an activation layer on CNN which is used to apply the function in equation 1 [8]. This function makes the input value of the neurons that are less than zero or negative will be made 0 [12].The Sigmoid activation function is a non-linear activation function that has values in the range of 0 to 1 [13]. The sigmoid function can be formulated by equation 2 below, where $f(x)$ is activation function, e is euler's number, and x is output before activation.

$$f(x) = \max(0, x) \tag{1}$$

$$y = f(x) = \frac{1}{(1 + e^{-x})}$$

6. Loss Function

The Loss function or loss function is a function that determines the penalty for deviations between prediction results and labels in the training process [8]. The lower the error generated by the loss function, the better. Cross Entropy is one type of loss function that is used to measure the difference between the probability of the hypothesis result and the probability of the original truth.

e. Model Evaluation

The confusion matrix is a tool used to measure the performance of classification problems and get the accuracy of class classification predictions from a model against the actual class [7]. The confusion matrix can be seen in Table 1 below.

Table 1. Confussion Matrix

		Predicted	
		Negatif	**Positif**
Actual	**Negatif**	TN	FP
	Positif	FN	TP

This confusion matrix used is a confusion matrix for binary classes consisting of four values, namely TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). Measures obtained from the confusion matrix to evaluate the classification model are accuracy, sensitivity / recall, precision, and f-measure with equations 3 to 6 below.

1. Accuracy is the value of how accurate the classification model is in determining the correct prediction results on a test data. The calculation of accuracy can be seen in equation 3 below.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ of\ data} \quad (3)$$

2. Sensitivity or Recall is the value of the number that is predicted to be correct in a class divided by the amount of data that should be predicted for that class. The calculation of sensitivity can be seen in equation 4 below.

$$Sensitivity_c = \frac{TP_c}{TP_c + FN_c} \quad (4)$$

3. Precision is the value of the number that is correctly predicted in a class divided by the number of predicted results for that class. The calculation of precision can be seen in equation 5 below.

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \quad (5)$$

4. F-Measure is the average of the harmonics of precision and sensitivity. The f-measure calculation can be seen in equation 6 below.

$$F - Measure = \frac{2\ x\ precision\ x\ sensitivity}{precision +\ sensitivity} \quad (6)$$

f. Waterfall Model

The Waterfall model is a sequential software development method in which the progress process is considered to flow downwards like a waterfall through the stages that must be carried out for success in building the software [14]. The Waterfall model provides a sequential or sequential software lifeflow approach starting from analysis, design, coding or implementation, testing, and support or maintenance [15].

3. Methodology

a. Data Collection

The data used for this research is secondary data in the form of a public dataset, namely the BCI Competition III dataset II regarding the P300 speller paradigm. The dataset consists of two data subjects, namely subject A and B [3]. In this study, the dataset used is the subject A which is stored in the "Subject_A_Train.mat" file which consists of 85 character epochs. The dataset is obtained by downloading the data through the http://www.bbci.de/competition/iii/ website in matlab format.

b. Data Understanding

The dataset is stored in the form of a dictionary type, each key consisting of certain data. An explanation of understanding the data can be seen in Table 2 below.

Table 2. Explanation of Dataset Contents

No.	Key	Description
1	Flashing	Stimulus of intensification. 1 when the row or column is intensified and 0 otherwise. It has dimensions of 85 x 7794.
2	StimulusCode	Intensification of raw signal class rows or columns. 0 when the row or column is not intensifying, 1-6 when the column is intensifying, 7-12 when the row is intensifying. It has dimensions of 85 x 7794.
3	Signal	Raw signal. The length of this signal data is the same for each data, which is 64 electrode data. It has dimensions of 85 x 7794 x 64.
4	StimulusType	The class of the raw signal. 0 when the intensified row or column does not contain the target character and 1 when the row or column that is intensified contains the target character. It has dimensions of 85 x 7794.
5	TargetChar	Correct character label for each character epoch. It has as dimensions of 85.

c. Data Preparation

In the data preparation process, the first thing to do is transform the signal into signal data that has dimensions of 85 x 12 x 15 x 64. Also transform the StimulusCode and StimulusType data into data with dimensions of 85 x 12 x 15, so the dataset consists of four keys, namely signal, code, label, and targetchar. After that, the process of changing the shape of the dataset is carried out with the key signal being N x C and the key label being N data. The amount of data that will be used is 15,300 data. After that, the RandomOverSampling dataset process is carried out. Before that process, the number of target class data was 2,550 data and 12,750 non-target data. After that process, the number of data for each class becomes 12,750 data. The last stage is to divide the data into training data (20,655), validation data (2,295), and test data (2,550).

d. Convolutional Neural Network Architecture

This research will examine several proposed architectures and hyperparameters which can be seen in Table 3. The input is signal data from the form (64.1), where 64 is the number of features and 1 is the dimension. In addition, it consists of a convolution layer and max pooling with Conv-ID (Filter, Kernel, Activation) and Max Pooling (Size) formats, as well as an FC layer with 512 neurons. The output layer uses a Sigmoid activation function with 1 neuron.

Table 3. Proposed 1D-CNN Architecture and Hyperparameter

No	Architecture and Hyperparameter					
1	Conv-1D (32,3,ReLU)	-	Flatten	FC (512, ReLU)	-	-
2	Conv-1D (32,3,ReLU) + Max Pooling (3)	-	Flatten	FC (512, ReLU)	-	-

3	Conv-1D (32,3,ReLU) + Max Pooling (3)	Conv-1D (64,3,ReLU)	Flatten	FC (512, ReLU)	-	-
4	Conv-1D (32,3,ReLU) + Max Pooling (3)	Conv-1D (64,3,ReLU) + Max Pooling (3)	Flatten	FC (512, ReLU)	-	-
5	Conv1D (32,3,ReLU)	-	Flatten	FC (512, ReLU)	FC (512, ReLU)	-
6	Conv-1D (32,3,ReLU) + Max Pooling (3)	-	Flatten	FC (512, ReLU)	FC (512, ReLU)	-
7	Conv-1D (32,3,ReLU) + Max Pooling (3)	Conv-1D (64,3,ReLU)	Flatten	FC (512, ReLU)	FC (512, ReLU)	-
8	Conv-1D (32,3,ReLU) + Max Pooling (3)	Conv-1D (64,3,ReLU) + Max Pooling (3)	Flatten	FC (512, ReLU)	FC (512, ReLU)	-
9	Conv1D (32,3,ReLU)	-	Flatten	FC (512, ReLU)	FC (512, ReLU)	FC (512, ReLU)
10	Conv-1D (32,3,ReLU) + Max Pooling (3)	-	Flatten	FC (512, ReLU)	FC (512, ReLU)	FC (512, ReLU)
11	Conv-1D (32,3,ReLU) + Max Pooling (3)	Conv-1D (64,3,ReLU)	Flatten	FC (512, ReLU)	FC (512, ReLU)	FC (512, ReLU)
12	Conv-1D (32,3,ReLU) + Max Pooling (3)	Conv-1D (64,3,ReLU) + Max Pooling (3)	Flatten	FC (512, ReLU)	FC (512, ReLU)	FC (512, ReLU)

e. Framework

The framework in this research can be seen in Figure 2 below.

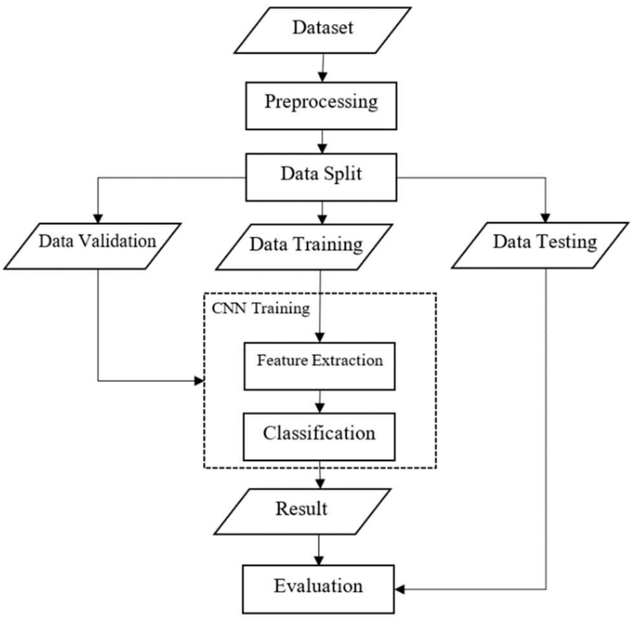

Fig 2. Research Framework

The stages of research work will begin by preprocessing the datasets that have been collected. An explanation of this preprocessing process can be seen in the data preparation section. After that, the dataset is divided into training data, validation data, and testing data. Furthermore, the training process is carried out using the CNN algorithm. There are two processes carried out by CNN training, namely feature extraction and classification. The input of this CNN training is one-dimensional signal data, while the output is the result of the classification of the target or non-target class of the signal. This training process will use training and validation data to find the best model from several proposed architectures and hyperparameters. After finding the best model, the model will be evaluated using data testing. This is done to see the performance of the resulting model. Evaluation uses a confusion matrix with evaluation values in the form of accuracy, precision, sensitivity / recall, and f-measure.

4. Result and Discussion

The experiment was carried out using the P300 speller dataset which had been carried out in the data preparation process. This test was carried out with the CNN algorithm using the proposed architecture and hyperparameter model. This model will be trained using training data and validation with 15 epochs. The model that has been trained will be evaluated using test data (test) with a confusion matrix to obtain evaluation values in the form of accuracy, precision, sensitivity / recall, and f-measure.

In this study, the test values for each proposed model have been generated. The test values are displayed in a table which can be seen in Table 4. In addition, each evaluation value will be displayed through the graphs in Figures 3 to 6.

Table 4. Model Test Results

Architecture	Accuracy	Precision	Sensitivity / Recall	F-Measure
1	88,63%	82,75%	96,59%	89,14%
2	82,55%	77,01%	91,07%	83,45%
3	94,27%	90,64%	98,30%	94,31%
4	91,73%	87,62%	96,51%	91,85%
5	86,43%	80,51%	94,89%	87,11%
6	71,29%	64,64%	89,61%	75,10%
7	93,37%	89,46%	97,81%	93,45%
8	91,33%	86,92%	96,59%	91,50%
9	89,73%	86,63%	93,10%	89,75%
10	77,10%	72,63%	84,42%	78,08%
11	92,94%	89,61%	96,59%	92,97%
12	91,45%	88,41%	94,72%	91,46%

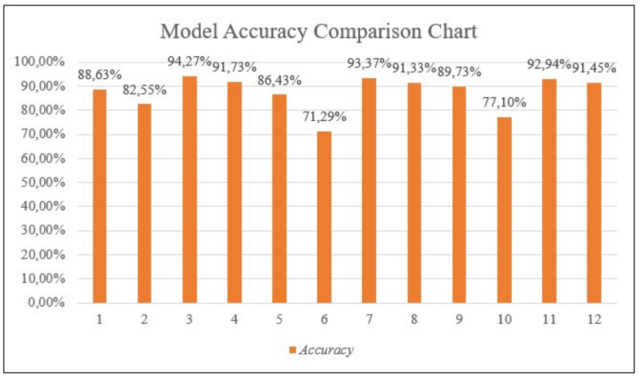

Fig 3. Model Accuracy Comparison Chart

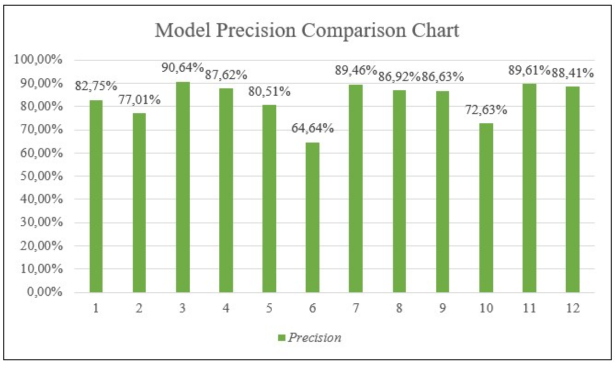

Fig 4. Model Precision Comparison Chart

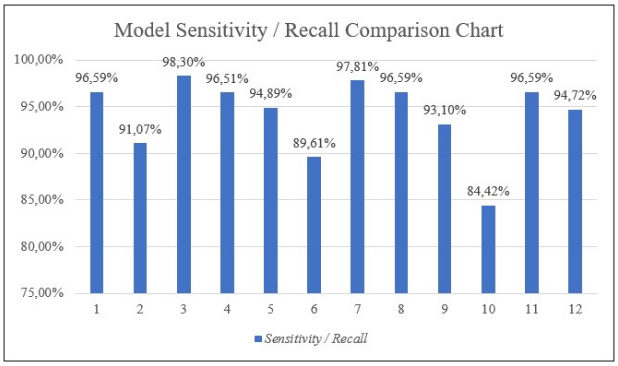

Fig 5. Model Sensitivity / Recall Comparison Chart

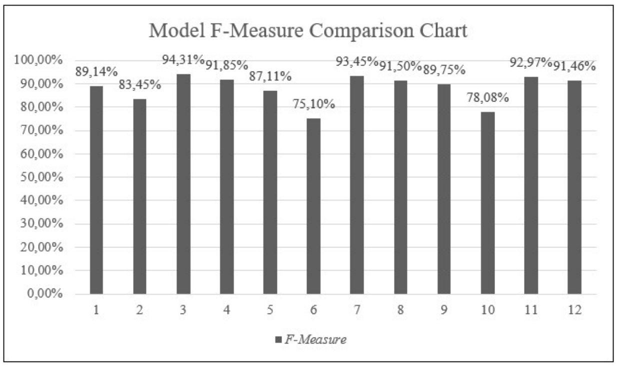

Fig 6. Model F-Measure Comparison Chart

Based on Figure 3 to 6, it can be seen that the cause of the decline in the performance of the model is caused by the addition of the max pooling layer. This is indicated by the decrease in accuracy in model 1 to model 2, where in model 2 there is an addition of a max-pooling layer. The decrease also occurred between model 3 to model 4, 5 to 6, 7 to 8, 9 to 10, and 11 to 12 due to the addition of the max pooling layer. Performance degradation also occurs in the value of precision, sensitivity / recall, and f-measure. Based on the above, it can be said that the addition of a max pooling layer causes a decrease in performance in this problem. In addition, it can be seen that the highest accuracy, precision, sensitivity / recall, and f-measure are found in the model with architecture 3. Based on this, the best model that has been produced is **model 3**.

In Figure 7, the results of the evaluation of the best model on the test data are shown in the form of a confusion matrix. The figure shows that the value of the target signal classification that is correctly predicted as a target signal (TP) is 1,211, a non-target signal that is correctly predicted as a non-target signal (TN) is 1,193, a non-target signal is incorrectly predicted as a target signal (FP). amounted to 125, and the target signal that was incorrectly predicted as a non-tagert (FN) signal was 21. Based on the confusion matrix plot, the results of the research test were the values of accuracy, precision, sensitivity / recall, and f-measure as shown in Table 4.

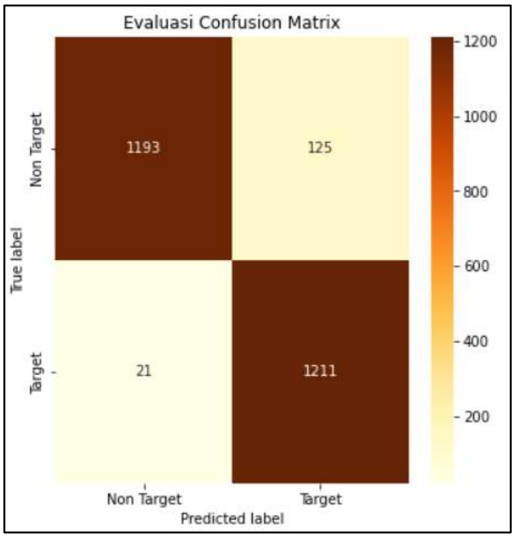

Fig 7. The Confusion Matrix of Best Model

In Figure 8, the results of the best model training process have been obtained which are displayed in the form of accuracy and loss plots. In the resulting loss plot, it can be seen that the loss training and validation values go to zero so that the model can be said to be good. In addition, the evaluation value model has a high value, which is above 90% for each metric value. Based on these things, the resulting model is a good model or goodfitting.

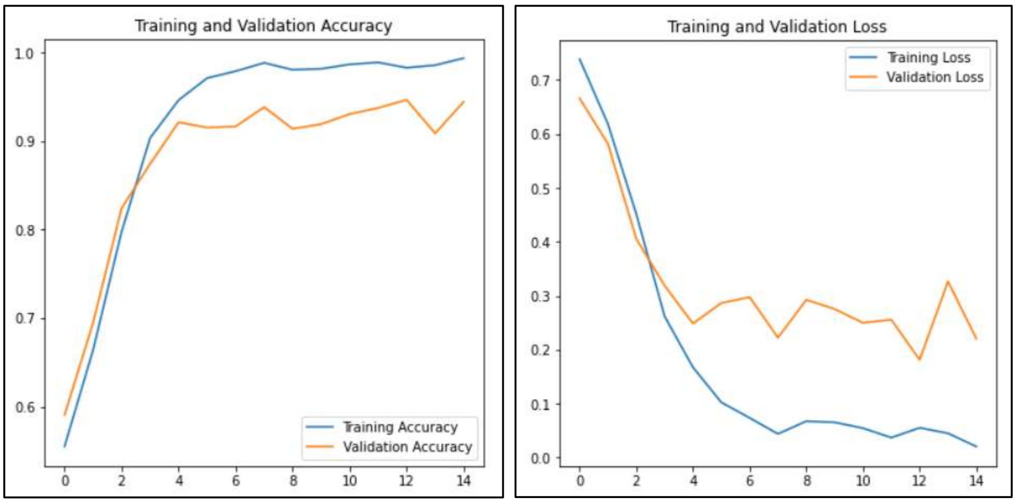

Fig 8. Training and Validation Accuracy and Loss Plot of Best Model

The resulting model has a good accuracy value of 94.27% in handling spelling detection based on the P300 signal with the BCI Competition III dataset II subject A. However, in previous research conducted by [3], in handling cases and objects which results in an accuracy of 100% for the SAE-ESVM method and 99% for the SSAE-ESVM method. Based on this, the proposed CNN method can be implemented for this problem, but this method is not better than the previous method.

5. Conclusion

The Convolutional Neural Network method can be implemented on spell detection based on the P300 signal, but this method is not better than the previous method in dealing with the problem. The resulting model has a good performance with evaluation values in the form of 94.27% accuracy, 90.64% precision, 98.30% sensitivity / recall, and 94.31% f-measure.

References

[1] R. Widadi, B. A. Widodo, and D. Zulherman, "Klasifikasi Sinyal EEG pada Sistem BCI Pergerakan Jari Manusia Menggunakan Convolutional Neural Network," *Techno.Com*, vol. 19, no. 4, pp. 459–467, 2020, doi: 10.33633/tc.v19i4.4119.

[2] Y. O. Halchenko, S. J. Hanson, and B. A. Pearlmutter, *Multimodal Integration: fMRI, MRI, EEG, MEG*. 2005.

[3] S. Kundu and S. Ari, "P300 based character recognition using sparse autoencoder with ensemble of SVMs," *Biocybern. Biomed. Eng.*, vol. 39, no. 4, pp. 956–966, 2019, doi: 10.1016/j.bbe.2019.08.001.

[4] N. Birbaumer and L. G. Cohen, "Brain-computer interfaces: Communication and restoration of movement in paralysis," *J. Physiol.*, vol. 579, no. 3, pp. 621–636, 2007, doi: 10.1113/jphysiol.2006.125633.

[5] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, pp. 1–20, 2021, doi: 10.1016/j.ymssp.2020.107398.

[6] B. Kumbara and A. Turnip, "Klasifikasi Sinyal Eeg Menggunakan Support Vector Machine (Svm) Untuk Deteksi Kebohongan," vol. 3, no. 3, 2015.

[7] W. Yunus, "Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik," *J. Tek. Elektro CosPhi*, vol. 2, no. 2, pp. 51–55, 2018.

[8] D. Suyanto, *Data Mining Untuk Klasifikiasi dan Klasterisasi Data*, Revisi. Bandung, 2019.

[9] I. Wijayanto, "Identifikasi Kondisi Sinyal Electroencephalogram (Eeg) Yang Terpapar Radiasi Sinyal Gawai 4g Lte 1800 Mhz Menggunakan Learning Vector Quantization (Lvq)," *TEKTRIKA - J. Penelit. dan Pengemb. Telekomun. Kendali, Komputer, Elektr. dan Elektron.*, vol. 1, no. 2, pp. 138–143, 2019, doi: 10.25124/tektrika.v1i2.1746.

[10] F. Akram, S. M. Han, and T. S. Kim, "An efficient word typing P300-BCI system using a modified T9 interface and random forest classifier," *Comput. Biol. Med.*, vol. 56, pp. 30–36, 2015, doi: 10.1016/j.compbiomed.2014.10.021.

[11] A. Syukron and A. Subekti, "Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit," *J. Inform.*, vol. 5, no. 2, pp. 175–185, 2018, doi: 10.31311/ji.v5i2.4158.

[12] S. Ilahiyah and A. Nilogiri, "Implementasi Deep Learning Pada Identifikasi Jenis Tumbuhan Berdasarkan Citra Daun Menggunakan Convolutional Neural Network," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.*, vol. 3, no. 2, pp. 49–56, 2018.

[13] Julpan, E. B. Nababan, and M. Zarlis, "Analisis Fungsi Aktivasi Sigmoid Biner Dan Sigmoid Bipolar Dalam Algoritma Backpropagation Pada Prediksi Kemampuan Siswa," *J. Teknovasi*, vol. 02, no. 1, pp. 103–116, 2015.

[14] Y. Bassil, "A Simulation Model for the Spiral Software Development Life Cycle," *Int. J. Eng. Technol.*, vol. 2, 2012.

[15] R. A. S and M. Shalahuddin, *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*, Revisi. Bandung: INFORMATIKA, 2018.