Emotional Sentiment Classification using Convolutional Neural Network

Abstract

Emotion is an unapproachable domain for researchers to understand the casual relationships and even mathematically prove the equation. However, it can be approached by using Electroencephalography (EEG) Test which gives the signals from part of the brain where it is correlated with emotion. This paper proposes a simple Convolutional Neural Network (CNN) to classify emotional sentiment based on Electroencephalography (EEG) brainwave data. Our proposed model architecture has a structure of one dimensional CNN which is mainly used for signal data. MLP model and Random Forest model, which respectively reached 94.89% and 97.89% accuracy with Information Gain method, are defined as baseline models. The experiment and evaluation results demonstrate the robustness of our proposed model which achieved over 97% test accuracy. With reduced dataset which is consisted of the top 2184 attributes applied by Information Gain method, the model achieved 98.13% test accuracy which outperforms both baseline models and our naive 1D-CNN model.

1. Introduction

There is no simple definition of emotional recognition. We need to combine various elements such as facial expression, voice patterns and monitored eye movements in order to assume the current emotional status[14], but it is impossible to consider all of the aspects.

EEG or Electroencephalography measures brain activity with electrical time-series signals, which is used for investigating brain function[1]. As it is incapable of being deceived, EEG is considered as the most reliable component, implying that it is one of the simplest solutions for identifying emotion[10, 15]. Many deep learning technologies are deployed for decoding the EEG signals[13, 7], and significant works related to the emotional sentiment classification have been done[11]. It has shown that the complex neural networks such as Convolutional Neural Networks(CNNs) outperform in classification problems[18]. Moreover, CNNs support various data types [6, 8] besides image datasets with a high performance. In that sense, we apply the neural networks into emotion classification task.

In conclusion, we are going to present the multi-layered Convolutional Neural Network (CNN) architectures to classify emotional sentiment, using the EEG Brainwave dataset.

2. Background

2.1. Electroencephalogram

An Electroencephalogram (EEG) is a test that detects abnormalities in the brainwaves. During the process, EEG measures the electrical activity of neurons in the brain with small, metal electrodes attached to the scalp[9] in order to collect EEG data. It is useful in diagnosing brain disorders such as brain tumor and sleep disorders.

Recent neurophysiological studies have shown that there are some correlations between EEG signals and emotions. The amygdala and the frontal lobe are the two main part correlated with emotional activity in brain, which store the most information about emotional states[17]. This implies that we are able to identify user's emotional state through conducting EEG test which captures brainwaves from those parts.

2.2. Convolutional Neural Network

CNN or Convolutional Neural Network is one of the most popular methods in deep neural networks. CNN has superior performance with image, speech and audio signal inputs. As EEG signals are time-series data, we need to transform it into discrete wavelet and used it as inputs to a 1-D CNN model. 1D Convolutional Neural Networks are mainly used on text and 1D signals.

An example of how 1D-CNN works is described in Figure 1. For 1D-CNN, we have 2-dimensional (spatial dimension, channels) input data, which means that we are training with a 3D tensor (batch, Spatial dimensions, channels).

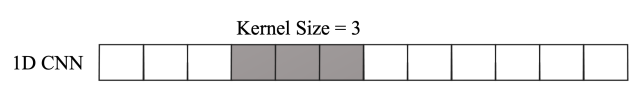

Figure 1. An example of 1D-CNN with kernel size 3. The width of input shape is 12 and there is 1 channel.

3. Related Work

Emotion is a complicated and subtle state to measure, but it can be way more easily expressed as sentiments such as positive, negative and neutral representations. Numerous works have been done for emotion classification using EEG signals[12, 18].

In [18], Suhaimi et al. analyzed emotional classification studies that proposes novel methods using EEG signals, and found out that SVM and KNN are the most popular algorithms used for emotion classification with highest achieved performance which were respectively 97.33% (SVM) and 98.33% (KNN). However, other classifiers such as CNN and RandomForest also made a great result which crossed the 90% margin.

In the research undertaken in [16], Saini et al. proposed a light-weight one-dimensional convolutional neural network architecture for mental task identification and classification. The proposed architecture only consists of two 1D-convolutional layers, one 1D-max pooling layer, one flatten layer with dropout and a final fully connected layer for classification output, but provides good classification accuracy.

4. Approach

In this section, we presents our approach to the problem by describing the dataset and the performance of baseline models. Our objective is to build the CNN model which could outperform the baseline models. By describing the architecture and loss function of the proposed CNN model with additional data preprocessing procedure, we details our method. Lastly, we elaborate the evaluation metrics which would be used for measuring the model performance.

4.1. Dataset

Our data comes from the EEG Brainwave Dataset: Feeling Emotions[2, 3], which consists of 2132 EEG signal data, and each data point has 2549 features. The dataset is consisted of 2132 x 2549 numerical values without any null values as we can see in Table 1. The sample of data is shown in Figure 2. The dataset was collected from two people (1 male, 1 female, aged 20-22) who were watching 6 different film clips with a Muse EEG headband. EEG signals were categorized into 3 different emotional states - positive, neutral, negative. We can notice that the dataset is balanced in Table 2.

4.2. Baseline Models

The baseline models are the MLP model and RandomForest model which achieved a high accuracy of 94.89% and 97.89% respectively on the given dataset[2] with feature selection algorithms. Those models were applied with Information Gain method which generates the reduced dataset from the 2,549 source attributes.

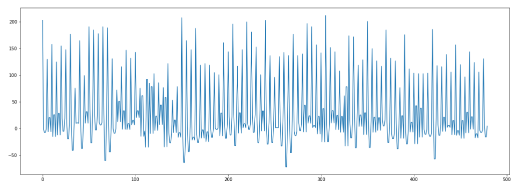

Figure 2. A sample of EEG Brainwave Dataset. This is a EEG signal with feature ranges from fft-0-a to fft-479-a for the fifth data point.

These are the best models that Bird et al [2] found among the various models such as BayesNet, Logistic Regression and Random Tree, each of which were trained on every dataset generated by the four methods - OneR, BayesNet, InfoGain and Symmetrical Uncertainty. 10-fold cross validation process was used for the model selection, and the parameters for deep neural network were 30 and 20 neurons on each layer and 500 epoch training time.

However, we need to identify that whether those high performances are due to the model themselves or the feature selection method or both. Moreover, the performance of MLP model implies that more complex neural network models such as Convolutional Neural Networks (CNNs) can be effective. Considering them, we will build the CNN models with applying various levels of feature selection method, and check whether CNN models are effective and robust even if no attribute selection was applied into it.

4.3. Method

4.3.1 Data Preprocessing

The first data preprocessing we have done is to scale the dataset into a standard range. We standardized features by centering and dividing by the standard deviation so that the mean of observed values is 0 and have a unit variance.

Then, we have split the scaled dataset into training set and test set with ratio of 80:20. We can notice that the labels in each of the datasets are all balanced in Table 3. 25% of the training set will be used as a validation set for the model selection at the end of each epoch. We decided not to define the validation set in advance in order to validate the model performance on a different random set. The ratio of training set, validation set and test set would be 60:20:20.

4.3.2 Architecture

As we are classifying signal time series data into 3 categories, the architecture should feed the signal through a 1D convolutional deep neural network. We referred two following CNN architectures in order to construct our own model:

	mean_{0a}	mean_{1a}	mean_{2a}	mean_{3a}	...	fft_{747_b}	fft_{748_b}	fft_{749_b}	label
0	4.620	30.3	-356.0	15.60	...	-162.00	-162.00	280.00	NEGATIVE
1	28.800	33.1	32.0	25.80	...	-31.60	-31.60	2.57	NEUTRAL
2	8.900	29.4	-416.0	16.70	...	-148.00	-148.00	281.00	POSITIVE
...	...	...	...	...	...	...	...	...	...
2130	16.800	19.9	-288.0	8.34	...	-271.00	-271.00	552.00	NEGATIVE
2131	27.000	32.0	31.8	25.00	...	22.80	22.80	-6.71	NEUTRAL

Table 1. The EEG brainwave dataset consists of 2132 rows and 2549 columns with numerical values. The data is resampled by statistical extraction since waves must be mathematically described in a temporal fashion[2, 4]

Label	Number of data
NEUTRAL (1)	716
POSITIVE (2)	708
NEGATIVE (0)	708

Table 2. The data distribution per label. Each emotional state has fair number of data, therefore the dataset is balanced.

Label	Training set	Test set
NEUTRAL (1)	573	143
POSITIVE (2)	566	142
NEGATIVE (0)	566	142
Total number	1705	427

Table 3. The dataset has been split into training set and test set with ratio of 80:20. The datasets are all balanced.

1. DeepECG: In [5], the researchers proposed 1D-CNN ECG classification program, which is consisted of [Conv1D, MaxPooling1D, Dropout]x3 + [Conv1D, Global-AveragePooling1D] + [FullyConnected, Dropout]x3 + [FullyConnected], and gives about 86% accuracy as the best validation accuracy.

2. ShallowConvNet : In [19], well-validated shallow CNN model for EEG signal processing and classification has been proposed, which has a structure of two Conv2D layer, Batch Normalization, AveragePooling2D layer, dropout and final fully connected layer with softmax activation function.

With merging characteristics of above architectures, we built our initial CNN architecture which has shown in Figure 3. However, the initial architecture has not given high performance, so we had to try various experiments such as deepening the layers and add some dropout layers. Finally, the overview of our proposed model architecture is shown in Figure 4, which comprises two set of 1D convolutional layer with ReLU activation, batch normalization layer and 1D max pooling layer with dropout, which is followed by the final fully connected layer with Softmax activation for classification output. By including batch normalization, max pooling and dropout, we can expect the model can scale the dataset even further and also be generalized.

4.3.3 Loss function

We are using sparse categorical cross-entropy loss which is generally used for multi-class classification problem as the loss function. Sparse categorical cross-entropy and categorical cross entropy have the same loss function, but the only difference is that the true labels y_i are integers or one-hot encoded. The loss function, which follows the Softmax activation from the output, is defined as follows:

$$L_i = -\log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$$

where L_i is the loss of the ith example in the mini batch of training data, f is the score for the particular class, j is one of the three possible classes, and y_i is the correct class for the ith example. Thus, our objective would be minimize the cross-entropy loss.

4.4. Evaluation Metric

The proposed model performance would be assessed by following evaluation metrics:

$$\text{(Classification) Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP refers true positives which is the number of cases where model correctly predicted the actual positive classes, TN refers true negatives which is the number of cases where model correctly predicted the actual negative classes, FP refers false positives which is the number of cases where model predicted as positive for the actual negative classes and FN refers false negatives which is the number of cases where model predicted as negative for the actual positive classes. Also, the confusion matrix would be provided in order to visualize the performance for multi-class classification.

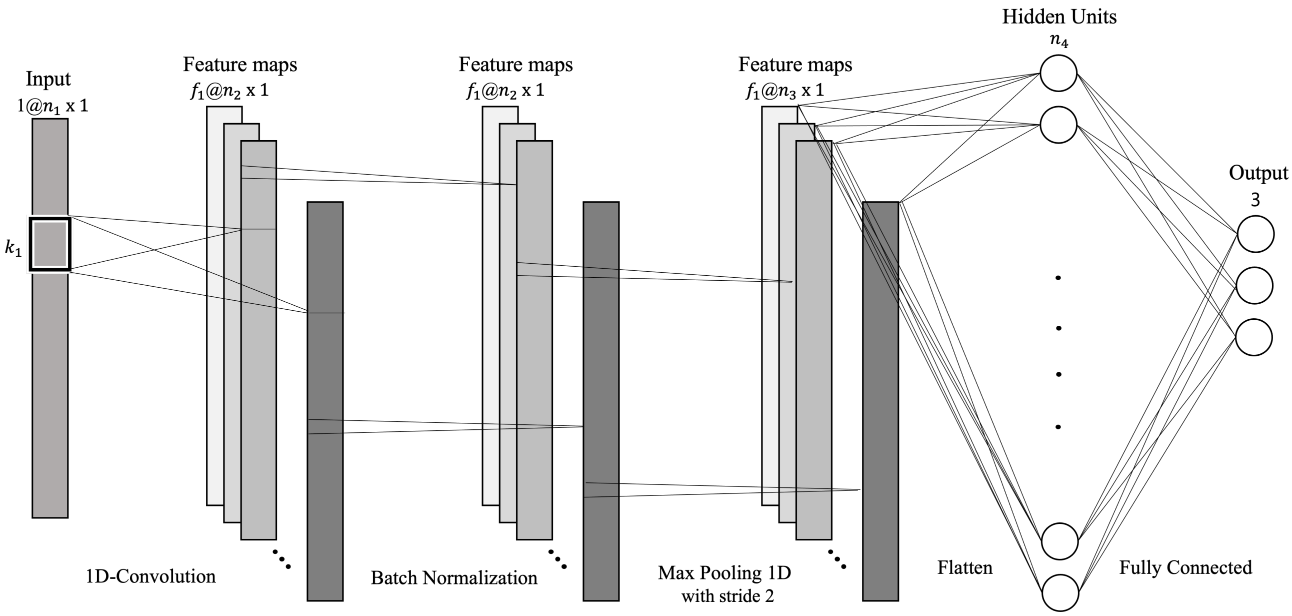

Figure 3. An initial architecture of proposed 1D-CNN model for emotional sentiment classification where n_1 is the number of feature, n_2 equals to $n_1 - k_1 + 1$, n_3 equals to $n_2/2$, n_4 equals to $f_1 \cdot n_3$ and output is three nodes as we are classifying emotional sentiment into positive, negative and neutral states.

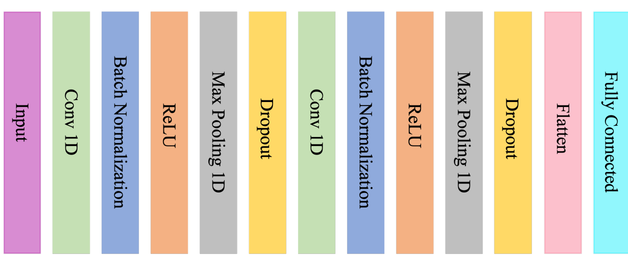

Figure 4. The final architecture of our proposed 1D-CNN model for emotional sentiment classification.

5. Experiment

5.1. Hyperparameter Tuning

With our final proposed architecture of CNN model for emotional sentiment classification, we expected to enhance the model performance by tuning hyperparameters. We have experimented with learning rate, kernel size, filters, dropout parameters and the batch size.

5.1.1 Learning Rate

The experiment of tuning learning rate is shown in Table 4. We have set other hyperparmeters all the same and only differed learning rate with the value of 0.0001, 0.0005, 0.001. As we are getting both the highest training accuracy and the highest validation accuracy, we selected learning rate value as 0.001.

Learning Rate	Training Accuracy	Validation Accuracy
0.0001	0.9531	0.9415
0.0005	0.9812	0.9742
0.001	**0.9961**	**0.9789**

Table 4. The conditions were all same which was set as batch size=28, dropout1=0.5, dropout2=0.25, epochs=20, filters1=64, filters2=16, kernel size1=5, kernel size2=5.

5.1.2 Kernel Size

We have to define two kernel sizes which are respectively applied in two 1D-Convolutional layers. The experiment result of tuning kernel size is reported in Table 5. Both resulted in high accuracies, but as shown in Figure 5, the loss much stably decreases if we set the both of the kernel sizes as 5, compared to individually setting as 10 and 5. Therefore, we select the kernel size value for both 5.

Kernel Size	Training Accuracy	Validation Accuracy
5, 5	**0.9961**	**0.9789**
10, 5	0.9977	0.9696

Table 5. Conducted on batch size=28, dropout1=0.5, dropout2=0.25, epochs=20, filters1=64, filters2=16, learning rate=0.001.

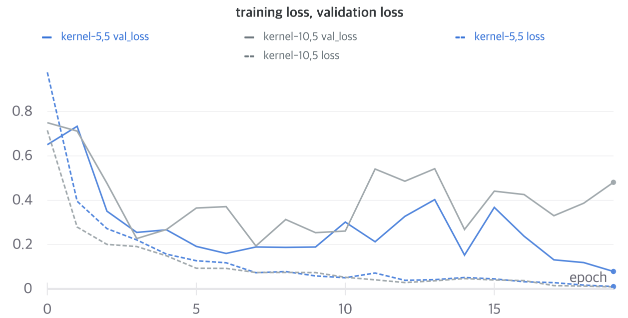

Figure 5. The validation loss decreases much more stably while setting kernel size1=5 and kernel size2=5, rather than setting 10 and 5 respectively.

5.1.3 Filters

The number of filters are yet another parameter which are applied in 1D Convolutional layers. We tuned them with pairs of (16,16), (64,16), (64,64) as described in Table 6. All candidates had fairly high training accuracies, but we selected the pair of (64,64) based on the highest validation accuracy among them.

Filters	Training Accuracy	Validation Accuracy
(16,16)	0.9992	0.9719
(64,16)	0.9961	0.9508
(64,64)	**0.9992**	**0.9789**

Table 6. Conducted on batch size=28, dropout1=0.25, dropout2=0.25, epochs=20, kernel size1=10, kernel size2=5, learning rate=0.001.

5.1.4 Dropout

In this section, we report the experiment that we have done in terms of tuning dropout parameters p_1 and p_2. Table 7 shows that the model tends to be prevented from overfitting as we increase the value of dropout parameters. Both (0.5,0.25) and (0.5,0.5) reach the same validation accuracy, but we decided to select the latter one which stably decreases the validation loss.

Dropout	Training Accuracy	Val Accuracy	Val Loss
0.5,0.25	0.9977	0.9696	0.4809
0.5, 0.5	**0.9883**	**0.9696**	**0.2808**
0.25,0.25	0.9961	0.9508	0.2442

Table 7. Conducted on batch size=28, epochs=20, kernel size1=10, kernel size2=5, filters1=64, filters2=16, learning rate=0.001.

5.1.5 Batch Size

We have experimented the batch size with 13, 28, 52 values as described in Table 8. It shows that setting batch size as 52 is too big, while 13 and 28 are moderate values. We selected batch size 28 which gives the highest validation accuracy among them.

Batch Size	Training Accuracy	Validation Accuracy
13	0.9789	0.9625
28	**0.9687**	**0.9696**
52	0.9671	0.9204

Table 8. Conditions were epochs=10, dropout1=0.5, dropout2=0.5, filters1=64, filters2=64, kernel size1=5, kernel size2=5, learning rate=0.001.

5.2. Evaluation

We have built our final proposed CNN model which has been tuned as Table 9. The Figure 6 describes our model architecture with input, output volume size from each layer.

Batch	Dropout	Filters	Kernels	LR	Epochs
28	0.5, 0.5	64, 64	5, 5	0.001	30

Table 9. The best combination of hyper parameters.

The plot of our proposed model performance measured by accuracy is shown in Figure 7. We achieved 99.53% training accuracy and 97.42% validation accuracy which are both fairly high.

The model performed 97.19% accuracy on test set, and the performance measured with different evaluation metrics on each class is shown in Table 10. The confusion matrix is shown in Figure 8.

Class	Precision	Recall	F1-score	support
0	0.97	0.96	0.96	142
1	1.00	0.99	0.99	143
2	0.95	0.97	0.96	142

Table 10. The classification report on our final proposed CNN model.

We compared our proposed 1D-CNN model with the baseline models, and the Table 11 demonstrates that the CNN model performs well even if no feature selection was applied. Moreover, considering that MLP model trained with 500 epochs, our model is competitive in that it only took 30 epochs.

5.3. Experiment with Feature Selection

Another experiment that we have done is to find out whether feature selection is effective. As our dataset has

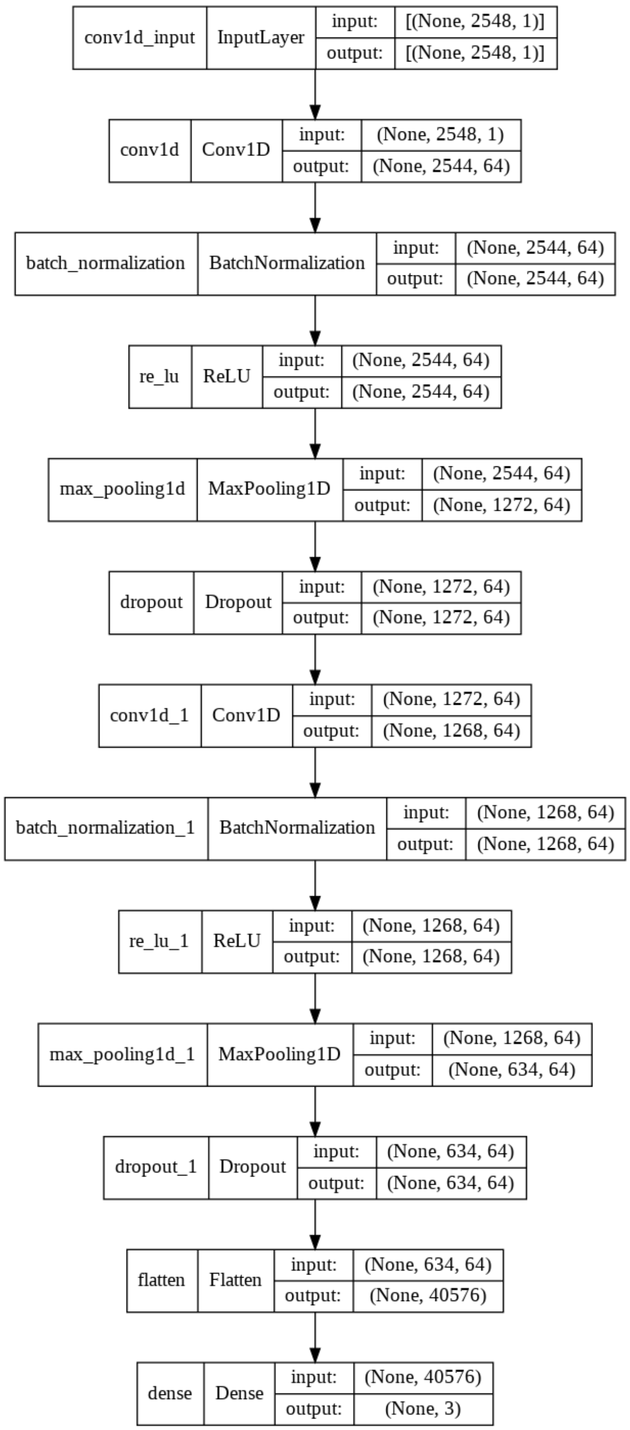

Figure 6. The architecture of proposed 1D-CNN model for emotional sentiment classification. We have built the model architecture via the Tensorflow framework and hyperparameters are tuned as described in Table 9.

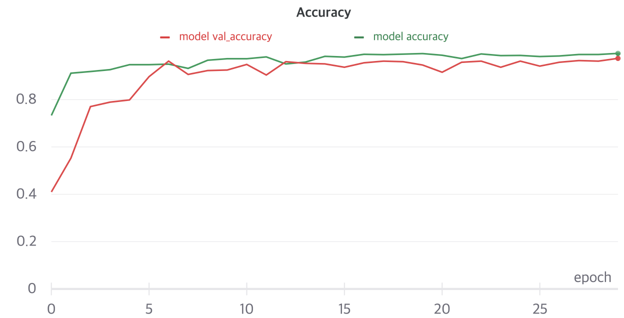

Figure 7. The performance of our proposed 1D-CNN model with hyperparameter tuning, which achieves 99.53% training accuracy and 97.42% validation accuracy after 30 epochs.

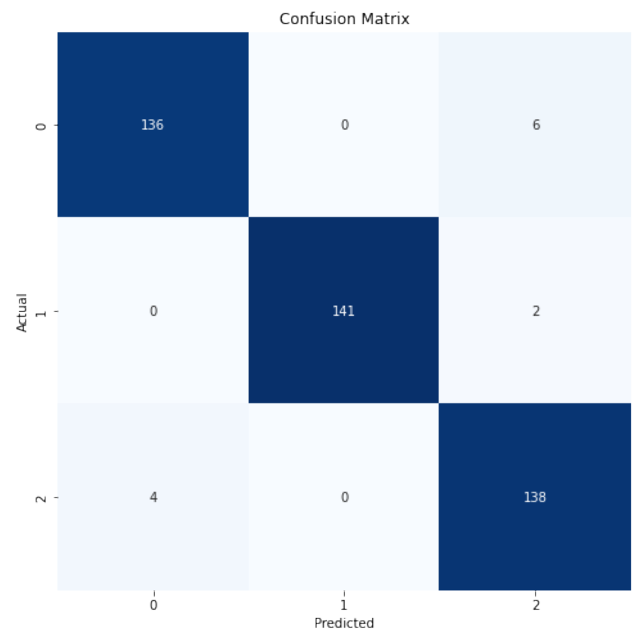

Figure 8. The confusion matrix of our 1D-CNN model.

Study	Method	Accuracy
This study	**1D-CNN**	**97.19**
Bird, et al.[2]	InfoGain, RandomForest	97.89
Bird, et al.[2]	InfoGain, MLP	94.89

Table 11. The comparison of baseline models' performance and our proposed 1D-CNN model.

high number of features which is larger than the data points, we wondered if reducing the dataset with feature selection method would result in a higher performance.

We used Information Gain attribute evaluation method in order to generate reduced dataset. During data preprocessing procedure, we identified k top features which are mostly related to our dataset using Mutual Information. Mutual Information, which is known as Information Gain, quantifies the dependencies between the variables. Higher MI score refers higher dependency, and we selected the top k attributes which gives the k largest MI scores, before we split the dataset. In the experiment, we selected 2184 features, all of which has more than 10% MI scores.

We applied the feature selection method on our final proposed 1D-CNN model. Figure 9 demonstrates the effectiveness of feature selection method, which we mostly achieved higher validation accuracy for each epoch.

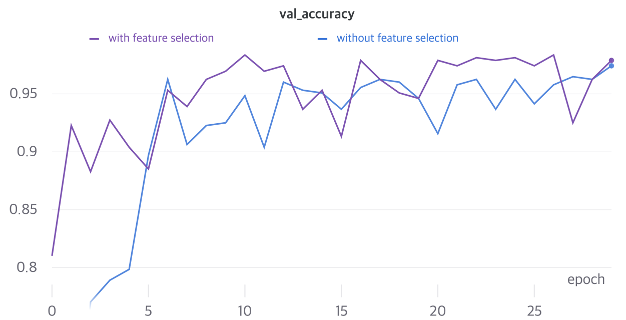

Figure 9. The validation accuracy of the model without feature selection and the one with feature selection which has the reduced dataset of 2184 attributes, each of which has more than 10% MI scores.

We evaluated its performance on test set with the same evaluation metrics: Classification accuracy, Precision, Recall, F1-Score and Confusion matrix, and each are described in Table 12 and Figure 10. The comparison with other models are shown in Table 13.

Class	Precision	Recall	F1-score	support
0	0.99	0.97	0.98	142
1	0.99	1.00	0.99	143
2	0.97	0.97	0.97	142

Table 12. The classification report on our final proposed CNN model with feature selection.

Study	Method	Accuracy
This study	**InfoGain, 1D-CNN**	**98.13**
This study	**1D-CNN**	**97.19**
Bird, et al.[2]	InfoGain, RandomForest	97.89
Bird, et al.[2]	InfoGain, MLP	94.89

Table 13. The comparison of baseline models' performance and our proposed 1D-CNN models.

The evaluation results shows that the feature selection method helps the model even more robust and enhanced.

6. Conclusion

In this paper, we have discussed about existing methods for emotional sentiment classification based on EEG brainwave dataset and other related works. Considering that Convolutional Neural Networks have been utilized in various domains, we decided to propose our new model using

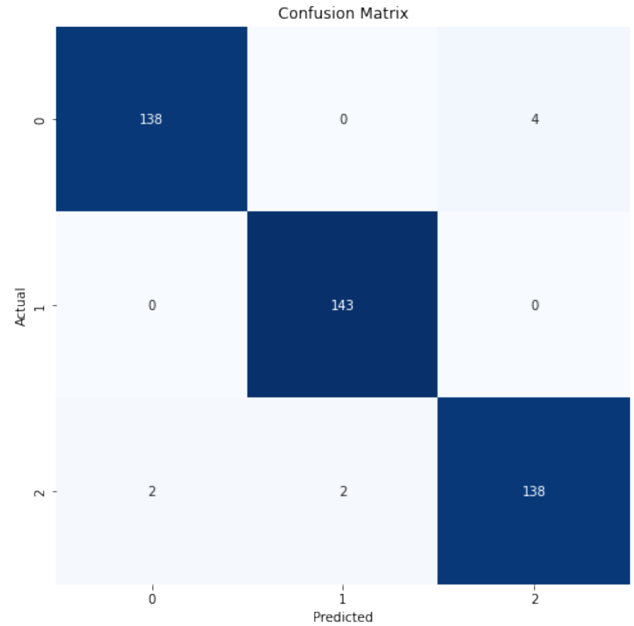

Figure 10. The confusion matrix of our 1D-CNN model with feature selection.

1D-CNN architecture to classify emotional sentiment. We have demonstrated that the model can reach over 97% accuracy, and even higher accuracy which is over 98% accuracy with Information Gain method. The performance results show that the proposed architecture outperforms the existing methods in terms of both classification accuracy and efficient training.

Our future work contains three parts as follows:

1. Utilize data augmentation which collects more labeled data, and expect to have a better performance.

2. Apply other feature selection methods with diverse levels, and see if it can lead to another performance gain.

3. Conduct an experiment with a recurrent neural net (RNN) and long short-term memory (LSTM) network. As both are suitable for working with time series data, we can expect high performance in EEG signal classification task. It would be also interesting to compare their performance with the proposed CNN model.

References

[1] Martin-Lopez D Valentin A Alarcon G Sanei S Cheong Took C. Antoniades A, Spyrou L. Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial eeg. In *IEEE Trans Neural Syst Rehabil Eng. 2017 Dec;25(12):2285-2294. doi: 10.1109/TNSRE.2017.2755770. Epub 2017 Sep 22. PMID: 28952945*, 2017. 1

[2] Jordan J Bird, Anik'o Ekart, Christopher D Buckingham, and Diego R Faria. Mental emotional sentiment classification with an eeg-based brain-machine interface. In *Proceed-*

ings of the International Conference on Digital Image and Signal Processing (DISP'19), 2019. 2, 3, 6, 7

[3] Jordan J Bird, Diego R Faria, Luis J Manso, Anik'o Ek'art, and Christopher D Buckingham. A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction. *Complexity*, 2019, 2019. 2

[4] Jordan J Bird, Luis J Manso, Eduardo P Ribeiro, Anik'o Ek'art, and Diego R Faria. A study on mental state classification using eeg-based brain-machine interface. In *2018 International Conference on Intelligent Systems (IS)*, pages 795–800. IEEE, 2018. 3

[5] Nick Mikhailovsky Boris Pyakillya, Natasha Kazachenko. Deep learning for ecg classification. *Journal of Physics: Conference Series*, 913:1–5, 2017. 3

[6] S. Dieleman, P. Brakel, and B. Schrauwen. Audio-based music classification with a pretrained convolutional network. *In ISMIR, pages 669–674*, 2011. 1

[7] Zhongke Gao, Dang Weidong, Xinmin Wang, Xiaolin Hong, Linhua Hou, Kai Ma, and Matjaž Perc. Complex networks and deep learning for eeg signal analysis. *Cognitive Neurodynamics*, 15, 06 2021. 1

[8] A. Graves, A. r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. *In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, pages 6645–6649. IEEE*, 2013. 1

[9] Minow F et al Light GA, Williams LE. Electroencephalography (eeg) and event-related potentials (erps) with human participants. In *Curr Protoc Neurosci. 2010;Chapter 6:Unit-6.25.24. doi:10.1002/0471142301.ns0625s52*, 2010. 1

[10] Jung Tzyy-Ping Lin Yuan-Pin. Improving eeg-based emotion classification using conditional transfer learning. *Frontiers in Human Neuroscience*, 11:334, 2017. 1

[11] Junxiu Liu, Guopei Wu, Yuling Luo, Senhui Qiu, Su Yang, Wei Li, and Yifei Bi. Eeg-based emotion classification using a deep neural network and sparse autoencoder. *Frontiers in Systems Neuroscience*, 14:43, 2020. 1

[12] Gil Rosa Guilera Teresa Batalla Iolanda Soler-González Jorge García Roberto López-Gil Juan-Miguel, Virgili-Gomá Jordi. Method for improving eeg based emotion recognition by combining it with synchronized biometric and eye tracking technologies in a non-invasive and low cost way. *Frontiers in Computational Neuroscience*, 10:85, 2016. 2

[13] Ewan Nurse, Benjamin S. Mashford, Antonio Jimeno Yepes, Isabell Kiral-Kornek, Stefan Harrer, and Dean R. Freestone. Decoding eeg and lfp signals using deep learning: Heading truenorth. In *Proceedings of the ACM International Conference on Computing Frontiers*, CF '16, page 259–266, New York, NY, USA, 2016. Association for Computing Machinery. 1

[14] Mark Purdy, John Zealley, and Omaro Maseli. The risks of using ai to interpret human emotions. 1

[15] Soumava Kumar Roy, Chetan Ralekar, and Tapan K. Gandhi. Emotion classification from eeg signals. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 2543–2546, 2016. 1

[16] Manali Saini, Udit Satija, and Madhur Deo Upadhayay. Light-weight 1-d convolutional neural network architecture for mental task identification and classification based on single-channel eeg, 2020. 2

[17] Riyanarto Sarno, Muhammad Nadzeri Munawar, and Brilian Tafjira Nugraha. Real-time electroencephalography-based emotion recognition system. *International Review on Computers and Software*, 11:456–465, 2016. 1

[18] Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo. Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Computational Intelligence and Neuroscience*, vol. 2020(Article ID 8875426):19 pages, 2020. 1, 2

[19] Schirrmeister Robin Tibor, Springenberg Jost Tobias, Fiederer Lukas Dominique Josef, Glasstetter Martin, Eggensperger Katharina, Tangermann Michael, Hutter Frank, Burgard Wolfram, and Ball Tonio. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420. 3