Optimizing Deep Learning Neural Networks: Brain to Computer Interface EEG-Based Imagined Word Prediction for Speech Disability

Babu Chinta[1,*] , Dr. Moorthi M.[2]

[1]*Reseaarch Scholar, Anna University, Chennai, India- 600025.*
[2]*Professor, Department of Biomedical/Medical Electronics, Saveetha Engineering College, Chennai, India-602117.*
[*]*Author email: babuchinta2019@gmail.com*

Abstract.

This idea starts with vowel recognition and then designs a vowel GUI. Five One Versus Rest (OVR) classifiers are created in the next step. QSVM had 91.1% accuracy over 10 trials and 10 patients. Five classifiers (FDT, LDA, QSVM, WkNN, and Subspace Discriminant Classifier) are created using the One Versus Rest (OVR) technique and tested (SDC). Classifiers must identify "A, E, I, O, and U," fake vowels. PCA can improve classifier quality. This upgrade boosts performance by 20%, which is significant. And other side the deep learning CNN model is used. Alexie and training a BCI with EEG-based imagined word prediction, since they can distinguish up, down, right, left, and up to ten words from visual inputs. Alex Net outperformed Google Net in transfer learning. It had a higher accuracy (91.3%), recall (92.4%), precision (91.0%), and F1 score (91.7%) for the seven features. Reducing the number of recoverable attributes from seven to four decreased performance ratings from 85.4% to 84.8%, then 84.9% to 83.6%.

Keywords. Brain Computer Interface, Convolutional Neural Network, Principal Component Analysis, Quadratic Support Vector Machine.

1. INTRODUCTION

When a person uses a BCI to send out messages or orders, those signals bypass the normal output channels of the brain, which consist of the brain's nerves and muscles (Wolpaw et al., 2002). Patterns of electrical brain activity are recorded and stored in an EEG-based BCI, for example. The use of a BCI provides its user with a new means of communicating with the outside world, it can be used to help restore a wide range of senses and abilities, including sight, hearing, movement, speech, and thought. Since the advent of biomedical signal processing techniques, electroencephalography (EEG) data has been widely implemented in the areas of Brain Computer Interface and the detection of neurological diseases. Researchers working in the field of brain-computer interfaces (also known as BCIs) have recently moved their focus to the development of innovative Humans have developed augmentative communication and control technologies for people with severe neuromuscular disorders. Our increased understanding of how the brain works, the widespread availability of powerful, low-cost computer equipment, and the recognition of the needs and potentials of people (Wolpaw et al., 2002).

A wide variety of classification methods are now being used in BCI research and development. Current categorization techniques for EEG data were examined by researchers [5]. Algorithms are broken down into four distinct categories: Matrix/tensor, adaptive learning, and deep learning classifiers. In this study, we explore some of the key challenges associated with correctly categorizing EEG data. Each classifier's design philosophy, plus its pluses and minuses, are laid forth in detail. Also, the study examines the stability, dynamicity, and normalization of each classifier. Consistent with what the researcher has done before [6]. The topic of the classifier's applicability is also explored in this study. Future prospects for the classifiers are examined to wrap up the study. Improved accuracy in EEG feature classifiers is an area where further work is needed. [7].

2. LITERATURE REVIEW

For emotion recognition, a Deep Learning Network (DLN) with Principal Component Based Covariate Shift Adaptation was developed. based on EEG data [1]. (PCBCSA). In this study, employed DLN to investigate the unknown feature correlation between the input signals, which contributed significantly in the learning process. Hierarchical feature-based DLN implementation uses Stacked Auto Encoder (SAE). So, used a three-layer stack of autoencoders with two soft-max layers, one for valence and one for arousal, to estimate the emotion classification.

X. Zhang, L. et al [2] encountered difficulties in correctly identifying the subject's or person's mind from the raw brain signals analyzed due to the low quality and noise inherent in these signals. Pre-processing, which is typically the initial stage, is also a time-consuming process. Therefore, they proposed Unified Deep Learning [UDL] to aid human-machine cognition. The suggested framework initially considered employing electroencephalography (EEG), but has now expanded to include fMRI stands for functional magnetic resonance imaging, while magnetoencephalography measures brain activity (MEG). The raw brain impulses were uploaded to a cloud server, where they were analyzed using a convolutional deep learning model that varies from person to person.

N. Kumar, K. et.al [3] utilized bi-spectral analysis, with phase information obtained via phase signal detection, to define the non-Gaussian components of the EEG signals and reveal the link between the frequency components. The bi spectrum quantitative emotional features were derived using a Valence-Arousal model. The EEG signals were collected using a two-channel EEG signal machine, then preprocessed to reduce artefacts caused by EOG, and then filtered using a Butterworth algorithm.

J. Atkinsona and D. Campos et al [4] recognized the difficulties in EEG signal categorization and recognition, specifically that automatic recognition is typically limited to a small number of emotion classifications. Scientists have proposed a new method of emotion identification combining brain-computer interfaces and electroencephalograms in an effort to solve these problems.

3. PROPOSED METHOD

Implementation loops

a. EEG signal properties

The EEG machine serves as the foundation of the entire setup. The EEG gadget has 16 sensors, and their positions are predetermined according to the 10-20 system. Out of the total of sixteen sensors, two are set as benchmarks. These two sensors will rest on the mastoid bones, which are located behind the ears. Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2 are the locations of the sixteen sensors. Each of these sensors provides a unique "channel" of information. The sensors' readings will be based on the brain's electrical impulses, generated by neurons.

The data from the EEG gadget is transferred to the computer through Bluetooth. There are a total of 25 channels in the incoming data, an increase of 9 channels from before. These nine channels carry timestamps, counters, marker signals, synchronization signals, gyro measurements, and more. Assuming a sample rate of 128Hz, the EEG equipment collects data once every 0.0078125 seconds. A total of 256 samples are taken over the course of the two-second recording and then analyzed using a matrix. In this case, the matrix would have 256 columns and 25 rows. The signal is then passed through a high pass filter, which attenuates the DC offset and eliminates any low-frequency disturbances. There is a cutoff frequency of 5 Hz in the high pass filter. In order to use MATLAB for classifier training, data is typically stored in a computer in a matrix format.

b. Combined PCA/Quadratic SVM for vowel recognition

Figure 3.1 shows the enhanced EEG-based vowel detection system block architecture using PCA and quadratic SVM. Data is sampled at 128 hertz using the EEG machine. Reduced sample sizes per recording as a result of the improved sampling rate will drastically cut down on the time and energy required to train and test classifiers. Each activity recording will have 256 sets of data. Because Principal Component Analysis (PCA) will rotate data and identify the direction with the highest degree of variation, it is used by the system instead of Common Spatial Features (CSP). This technique works well with information that does not cover every available experiment.

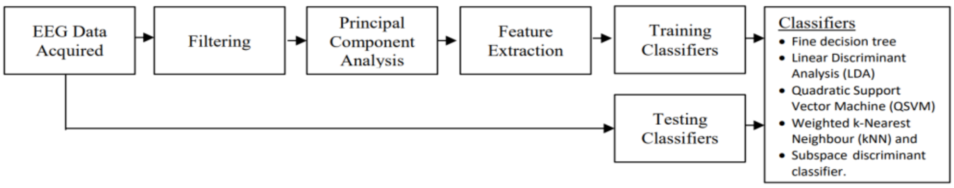

Figure 3.1. The improved EEG vowel identification system showing the PCA and quadratic SVM integrating

To train classifiers, PCA is used. The data in the predictor columns are transformed into the answer field. FDT, LDA, SVM, WkNN, and SD classifiers are the first five classifiers trained. In order to aid these classifiers in their classification efforts, principal component analysis can be used to determine which direction has the most data variance. Five-fold cross-validation is used to train the classifiers' data. As a result, the initial dataset was split into five parts, two of which would be used for training and the other two for validation. A new batch of data is then used to train the classifier and validate it. The training process will continue to repeat as long as the classifiers validate with each successive set and acquire new values for their weights. One classifier model is retained from the data and utilized for prediction. Resulting from repeated recording and classifier-training sessions, a final recording is made to test the classifiers. The data is scrubbed and then stored. Afterward, the saved classifier model is applied to the archived data for predicting purposes. The classifier model will provide a number between one and five to indicate its recommended course of action.

c. Word Prediction Using CNN

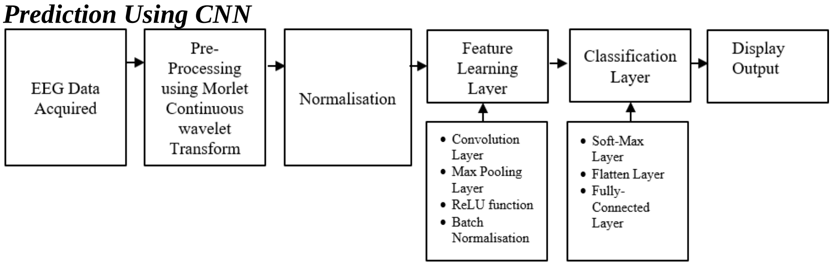

Figure 3.2. Block diagram for word recognition

Figure 3.2 shows the block diagram for word recognition. The EEG signals are converted into a 3D scalogram image using the Continuous Wavelet Transform (CWT), and the word imagined is predicted using a convolutional neural network (CNN), that can capture image time- and spatial-dependencies with the right filter.

i. Pre-processing using Morlet Continuous Wavelet Transform

The size and structure of the dataset were verified by loading before any preprocessing or normalization of label value could be carried out. The window function in CWT controls the primary wavelet function and can be resized and moved to suit the features of the desired wavelet. Thus, windowing can be performed over a longer time for low frequency and a shorter period for high frequency. Additionally, the variance analysis window can be split in two, simultaneously allowing for a detailed investigation of low- and high-frequency information in a non-stationary EEG signal. Because Morlet wavelet CWT is better suited for non-stationary EEG signals, it was used for the spectral analysis.

ii. Alex net

Alex net is an 8-layer deep convolutional neural network that has been trained on more than a million signals, making it capable of signal prediction across 1000 object categories. Both the fully linked layer and the output layer have 1000 classification outputs, and this number is utilized to categories the signals. However, this task requires the production of 10 distinct classifications. In order to get to an epoch where the fully connected layer can do the classification, weights and biases were fine-tuned at each iteration using backpropagation. Finally, the trained CNN model will be tested by feeding its output from the fully connected layer into a soft-max classification layer for labelled final classification. The SoftMax layer produces 10 outputs, which is the desired number of outputs (left, right, up, down, front, back, stop, pick, red, blue).

iii. Training Model Setup

In this part, we compare two trained models and pick the one that performs best in terms of accuracy and training time. The Alex net and the Googlenet are two trained models that are compared in this experiment. Googlenet is trained with an epoch of 165 because it needs more time to learn, while Alex net is educated with an epoch of 85. The Alex net is more accurate than the Googlenet despite increased epoch and training time. Performance & result analysis.

iv. PCA and Quadratic SVM & 9-trail classifier accuracy

This test an appearance at the accuracy of classifiers trained with only one trial the classifiers' efficiency when only one trial, or recording session, is used during training. This presents a problem for the concept of developing a reliable and precise classifier with a minimum of training time. The classifier must be flexible enough to accommodate subtle shifts in the data. First session of Subject 1's recordings is used to train all five classifiers, as indicated in Table 3.1. The purpose of this experiment is to see how much of an effect principal component analysis (PCA) has on classifier performance, as well as how much of an effect more data has on classifier performance. As demonstrated in Table 3.2, this test will also reveal whether or not direct correlations between training data availability and classifier performance are statistically significant. To replicate Subject 1's result from the preceding experiment.

Table 3.1. Accuracy of one-trial classifier performance

Classifier	Accuracy (%)
Fine Decision Tree (FDT)	81.06
Linear Discriminant Analysis (LDA)	82.03
Quadratic (SVM)	89.01
Weighted (WKNN)	80.06
Subspace Discriminant (SD)	87.08

Table 3.2. Accuracy of 9-trials classifier performance

Classifier	Accuracy (Without PCA %)	Accuracy (With PCA %)
Fine Decision Tree (FDT)	68.02	87.06
Linear Discriminant Analysis (LDA)	72.02	89.07
Quadratic (SVM)	68.05	91.01
Weighted (WkNN)	73.08	88.00
Subspace Discriminant (SD)	60.08	89.08

v. Vowel classifier accuracy

The assumption is that different classifiers have different capabilities when it comes to identifying vowels. To describe the vowels "a," "e," and I for instance, a quadratic support vector machine (SVM) may be preferable, while a linear discriminant analysis (LDA) may be preferable, as indicated in Table 3.3. In order to identify the most appropriate class of responses, classifiers are trained utilizing the OVR principle.

Table 3.3. Accuracy of classifiers in classifying each task.

Vowels	Fine Decision Tree (FDT) (%)	Linear Discriminant (LDA)(%)	Quadratic SVM (%)	WeightedKNN (%)	Subspace Discriminant (%)
A	82.02	89.02	96.00	86.03	89.02
E	86.03	89.07	91.00	89.06	88.01
I	87.09	86.09	92.00	89.03	89.02
O	88.00	91.00	88.05	85.09	90.03
U	87.04	90.02	88.03	88.01	89.02

Among these five classifiers, quadratic SVM performs the best. As a result, PCA has been demonstrated to enhance classifier quality. This enhancement boosts performance by around 20% in accuracy, which is a notable amount. The test outcomes in section four detail the overall performance of each classifier. A quadratic support vector machine (SVM) classifier is the most often used algorithm for classification.

vi. Word Prediction using CNN

A majority of the data (80%) was used for training the model, with the remaining (20%) used for testing and validation. According to Model 1, the results of the 10 tests given to each participant are presented in Table 4.4, whereas those according to Model 2 are presented in same Table 3.4. These tables show the generated and recorded results for each trained model used in the learning of EEG signal classification using 4 features extracted, including the Recall, Precision, Accuracy, and F1 scores for each trained model.

Table 3.4. Model-1 and Model-2 performance in tests using 4-feature data

Participants	Recall (%)		Precision (%)		Accuracy		F1 score	
	Model1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
1	91.01	81.08	89.06	89.08	90.04	89.08	90.03	85.06
2	86.04	77.08	83.02	76.01	85.04	77.05	84.08	76.09
3	87.07	79.01	85.03	79.04	86.03	79.03	86.05	79.02

4	86.02	77.05	84.07	75.05	85.07	75.05	85.04	76.05
5	85.08	75.04	83.02	74.07	83.02	74.02	84.05	75.00
6	86.06	87.02	84.01	86.05	85.05	86.01	84.08	86.03
7	90.08	88.09	88.05	89.01	89.08	88.01	89.06	88.05
8	91.02	93.03	93.02	92.05	94.05	92.02	92.02	93.09
9	87.06	86.01	86.01	85.08	86.01	85.01	86.08	87.04
10	93.07	90.06	91.04	88.07	92.08	88.07	92.05	89.06
Average	88.07	83.07	86.08	83.06	88.00	84.06	88.08	83.06

4. DISCUSSION & PERFORMANCE ANALYSIS

This study shows how to use a brain-computer interface (BCI) and electroencephalogram (EEG) to anticipate words. In this experiment, we first compared three distinct ILR values (0.0001, 0.0003, and 0.0005). We used an ILR of 0.0001 because it permits a room to attain an accuracy of over 70.68%, taking into account both the maximum feasible accuracy and the largest difference in validation and training loss. The proceeding two models are performance in tests using four-feature data. In terms of training time, Alexnet's 85 epochs were chosen since they required less time in the lab yet produced better results. Furthermore, Table 8 describes the results of a comparison of the two CNN models' performance when tested on data containing either 7 or 4 extracted features.

Table 4.1. Two model performance comparison

Used several features extracted	**Performance of Models**	**Recall (%)**	**Precision (%)**	**Accuracy**	**F1-score**
7	Model -1	92.04	91.00	91.03	90.07
	Model- 2	87.07	85.08	87.00	79.03
4	Model- 1	84.04	82.09	83.08	83.06
	Model- 2	82.07	82.06	82.06	82.06

Table 4.1 shows that when using all of the performance evaluation criteria for 7 and 4 extracted features, model 1 (Alex net) performs better than the rest. When 7 extracted features were applied, Alexnet's accuracy increased to 91.3%. In addition, When the number of extracted features was lowered from 7 to 4, all performance measures in both trained models decreased, though only by an average of 3.3%.

Table 4.2. Comparing developed models to literature

Performance of parameters analysis	**Model -1 Outperforms**		**Model -2 Outperforms**		**Kocturova& Juhar, 2021 [31]**	**Netzer, Frid & Feldman, 2020 [32]**
Used several features extracted	4	7	4	7	9	-
Recall (%)	85.04	92.04	83.07	88.07	89.085	-
Precision (%)	83.09	91.00	83.06	86.08	85.096	-
Accuracy	84.08	91.03	83.06	88.00	85.062	71.06
F1- score Measure	84.06	91.07	83.06	80.03	87.080	-

Table 4.2 contrasts the new models with those already in the literature [31, 32]. In comparison to model 2 (Googlenet), [31], and [32], it is clear that the accuracy of the generated model 1 has been enhanced by a significant margin (3.79%, 6.71%, and 27.90%). They have extracted a more manageable subset of features and evaluated the performance of the systems and models we have developed. Alexie was the better model than Googlenet due to its 91.3% accuracy after being trained with 80 epochs, 64 batches, the scalogram pre-processing strategy, an 80:20 split between the training and validation sets, and an initial learning rate of 0.0001. With this level of accuracy, researchers can expand their efforts to cross-participant analysis to gather a larger sample size for testing and refining the system's deep-learning neural-network foundations to make them appropriate for use in mobile apps based on EEG.

5. CONCLUSION

This concept begins with vowel recognition and subsequently develops a vowel user interface. Using the One Versus Rest (OVR) method, five classifiers (FDT, LDA, QSVM, WkNN, and Subspace Discriminant Classifier) are developed and evaluated (SDC). Classifiers must identify the fictitious vowels "a, e, I, o, and u." With an accuracy of 90.1% across 10 trials and 10 individuals, QSVM is the most precise classifier of the five. PCA helps increase classifier quality. This improvement increases performance by 20%. And the CNN model with deep learning is utilised. The Alex net transfer learning model was deemed to be superior to Googlenet because it achieved an accuracy of 91.3% whereas Googlenet had only 89.7% accuracy. Cross-participant analysis, which will expand participant numbers, will benefit from this high accuracy.

6. References

[1] S. Jirayucharoensak, S. Pan-Ngum and P. Israsena, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaption", The Scientific World Journal, Vol. 1, pp. 1-10, 2014.

[2] X. Zhang, L. Yao, S. Zhang, S. Kanhere, Salil, Sheng, Michael and Y. Liu, "Internet of Things Meets Brain-Computer Interface: A Unified Deep Learning Framework for Enabling Human-Thing Cognitive Interactivity", Journal of Latex Class Files, Vol. 14, No. 8, pp. 1-8, 2015.

[3] N. Kumar, K. Khaund and M. S. Hazarika, "Bi-spectral Analysis of EEG for Emotion Recognition", Procedia Computer Science, Vol. 84, No. 1, pp. 31-35, 2016.

[4] J. Atkinsona and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers", Expert Systems with applications, Vol. 47, No. 3, pp. 35-41, 2016.

[5] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10-year update", Journal of Neural Engineering, Vol. 15, Vol. 3, 2018, pp.1-29.

[6] L. Ravi, S. Sathish Kumar, Chow Him Mun, "Improved Speech Recognition using Neural Network", The International Journal of Applied Engineering Research, Vol.9, No. 18, 2014, pp. 4297-4325.

[7] L. Wookey, J. S. Jessica, O. Busra, S. S. Bong, M. Azizbek, L. Suan, "Bio signal Sensors and Deep Learning-based speech recognition: A Review", Sensors, 2021, pp. 1-22.

[8] K. Gautam, T, Co, C. Mason, H. Yan, T. H. Ahmed, "Improving EEG based Continuous Speech Recognition", 2019, arXiv:1911.11610.

[9] M. Beomjun, K. Jongin, P. Hyeong-Jun, L. Boreom, "Vowel Imagery Decoding toward Silent Speech BCI using Extreme Learning Machin with Electroencephalogram", BioMed Research International, 2016, p. 1-11.

[10] A. D. Khalid, H. Laila, Z. S. Saadeh, H. Mohammed, S. Ramzi, H. A. Sharhabeel, "An efficient speech recognition system for arm-disabled students based on isolated words" Computer Applications in Engineering Education, Vol. 26, No. 2, 2018, pp. 285-301.

[11] J. T. Panachakel, A. G. Ramakrishnan, "Decoding Covert Speech from EEG - A Comprehensive Review", Neuro Science, 2021

[12] C. Sandhya, S. R. Anandha, A. Kavitha, "Identification of vowels in consonant-vowel-consonant words from speech imagery-based EEG signals", Cognitive Neurodynamic, Vol. 14, 2020, pp. 1- 19.

Biographies

BabuChinta Completed his B.Tech in Electronics and Communication Engineering from Jawaharlal Nehru Technological University Kakinada, Andhra Pradesh, India. Post-Graduation (M.Tech) in Communication Systems obtained from Andhra University, Vishakapatnam, Andhra Pradesh, India. Present he pursuing Ph.D. at Anna University, Chennai. He had 5 years of teaching experience and he published papers on reputed journals. His research interests on Signal & Image processing, Artificial Intelligence, Brain to Computer Interface, and Speech signal Processing.

Dr.M.Moorthi is a Professor & Head of the Department in the Biomedical Engineering at Saveetha Engineering College, Chennai. He has professional teaching experience of over 20 years. His areas of research interest are Digital image processing and Multimedia Compression, Microprocessor and Microcontroller. He is Anna University Recognized Supervisor for Ph.D/M.S Scholars. He has obtained several funding from AICTE,CSIR,BRNS, IEEECurrently, he is the IEEE Student branch and Society advisor at Saveetha Engineering College. He is currently acting as an Executive member in IEEE Photonics Society, Madras section.Duzce University. His research areas include mobile security, deep learning, and social network analysis.