HIGH SCHOOL EDITION
Journal of Student Research

Using a Convolutional Neural Network to Classify Subvocalized Letters Collected from the Brain by EEG

Shaumprovo Debnath[1] and Dr. Prasanna Kolar[2#]

[1]Keystone School
[2]Southwest Research Institute, San Antonio, TX
[#]Advisor

ABSTRACT

Communication is essential to human behavior. Individuals with amyotrophic lateral sclerosis and severe traumatic brain injury encounter enormous challenges in communicating via speaking, writing or typing with others. Such lack of communication severely limits their independence and quality of life. Recent research employing invasive and expensive surgical implants of neuroprosthesis shows some degree of success in reconstructing language or images. However, these procedures are not user-friendly and not available for all affected individuals. To address these limitations, the researcher created a convolutional neural network model to classify individually thought letters – i.e., subvocalization. Using a commercially available simple electroencephalography (EEG) device data were collected from a study participant. All collected data was formatted for computer interpretation and uploaded to a Python notebook. The data was then augmented and preprocessed to make it easier for the neural network model to make predictions. The convolutional neural network model with the best performance had 63.33% classifying 38/60 samples correctly with a statistically significant p value ($p<0.00001$). Limitations include the reliability of the EEG headset used and time restrictions for collecting data. For future research, better hyperparameters will be investigated using additional data. Finally, the researcher plans to conduct experiments to make real-time predictions with the model once higher accuracies are achieved.

Introduction

Communication is a basic and essential skill to humans and is integral to our lifestyles. However, individuals with amyotrophic lateral sclerosis (ALS)—a progressive neurological disease affecting the brain and spinal cord resulting in restricted voluntary muscle movement—can find difficulty in speaking (dysarthria) and writing. Currently, eye movements or blinking is used to facilitate communication in these cases, however, even this painstaking process is not practical. Similarly, patients with locked-in syndrome present with muscle paralysis and inability to move or speak. One of the causes of locked-in syndrome is traumatic brain injury which is increasingly common in warzones, environmental disasters, and many competitive popular sports. In ALS and locked-in syndrome patients, the quality-of-life is significantly impacted by loss of communication due to inability to write, talk, or type. Currently, there is no treatment for ALS to reverse its progression and or to restore muscle movement including talking or writing.

Brain-computer interfaces (BCI), a technology that decodes electrical information from the brain, has been used in some experiments to restore communication in individuals who have lost the ability to speak or write (Birbaumer 2006; Wolpaw et al., 2000; Wolpaw et al., 2002). In one experiment, microelectrode arrays were surgically implanted in the brain of an ALS patient with locked-in state to form words and phrases using a neurally-based auditory neurofeedback system to choose words from a wordbank (Chaudhary et al., 2022). In another study, Willett et al. (2021) demonstrated that typing speeds of 90 characters per minute could be achieved in a person paralyzed from spinal cord injury using a recurrent neural network decoding approach. Similarly, another study decoded words and sentences from the brain by implanting a multielectrode array in the cortex area of the brain that controls speech in a tetraplegic person with anarthria, a motor speech disorder characterized by an inability to speak (Moses et al., 2021). All these studies are extremely complex, requiring supervision from highly trained personnel, sophisticated computer systems and programming, invasive surgical procedures, and importantly large financial cost. Although these experiments are promising, they are not practical for most of the individuals affected with ALS or locked-in syndrome. Using electromyography (a test that checks the health of the muscles and the nerves that control the muscles), Kapur, et al. (2020) collected electrical data sent to the larynx, requiring users to make movements within the mouth to articulate words. By employing non-invasive functional magnetic resonance imaging (fMRI), one recent study demonstrated the viability of decoding perceived and imagined stimuli into continuous language (Tang et al., 2023). Although this study marks an important step for non-invasive BCI, fMRI is very expensive technology.

In order to address these aforementioned problems and limitations, the purpose of this project is to develop a convolutional neural network (CNN) model capable of recognizing subvocalized (thought to oneself) letters using an electroencephalogram (EEG). EEGs collect electrical data of neurons communicating within the brain using electrodes, often in a range of frequencies. The EEG used has electrodes facing the frontal lobes of the brain, which include motor control (which might involve some activity by subvocalizing, although no muscles are actually used) and speech, specifically in Broca's area.

The researcher for this project used a commercially available, relatively inexpensive, and non-invasive EEG (Muse 2, InteraXon, Inc., Toronto, Canada). Data was collected from the researcher thinking a letter, either A, B, or C, to himself without moving any muscles and making any noise (subvocalization). The researcher did this procedure with his eyes closed without stimulation. The EEG includes electrodes placed at AF7, AF8, TP9, and TP10. The AF7 and AF8 electrodes measure data from certain parts of the frontal lobes, while the TP9 and TP10 electrodes measure data from parts of the temporal lobes. Both of these areas of the brain are responsible in some way for language, and so may be activated when subvocalizing letters (Shergill et al., 2002; Tyler and Marslen-Wilson, 2008).

The researcher chose to use a CNN for this study as it is commonly used in artificial intelligence (AI) for image recognition; when this is the case, images are normally converted into arrays containing the RGB (color) data for each pixel. Similarly, the data collected by the EEG (after preprocessing) was formatted into a time series array, such that each thought letter had its own array of data (which could be visualized as a grayscale image).

Materials

In this project, the following materials were used: 1) Alcohol wipes, 2) Saline Solution for eyes, 3) Piece of cloth, 4) Muse 2 EEG, 5) Computer with internet access, Bluetooth, and the BlueMuse app.

Methods

Preparation

The Muse 2 EEG was prepared by wiping electrodes as well as the rubber parts that cradle the ear using an alcohol wipe. Next, the cloth was dampened using the saline solution, then applied to the forehead of the subjects to improve conductivity between the forehead and electrodes, improving signal quality. After turning the EEG on by pressing and holding the "On" button on the right side, the EEG was placed on the forehead such that electrodes are facing it; adjust as necessary. Once the EEG was prepared, the BlueMuse app was downloaded on the computer and connected to the Muse 2 EEG device, ensuring the battery level was above 50%. Next, Python and the EEG-Notebooks Python module were installed. The signal quality of the EEG was checked using the EEG-Notebooks module in the terminal. The standard deviation of the channels TP9, AF7, AF8, and TP10 was ensured to be between 1.5 and 15; the headset was adjusted accordingly if this was not already the case.

Data Collection

Once signal quality was satisfactory, another script was coded that displayed on the screen the letter A, B, or C (cycling through) and played a short beep sound and recorded the data from the headset for 1 second. It wrote this data to a CSV file along with what it prompted. In data collection, this Python script was run and at the first beep, the subject subvocalized "A," at the second, "B," the third, "C," and at the fourth, "A" again, continuing to cycle through. When subvocalizing, no muscles were moved. Occasionally the subject opened their eyes for one sample to verify that they were subvocalizing the correct later. Once complete after the beeping had stopped), the signal quality was verified once more to ensure it has not degraded since recording started. When this was the case, the CSV file containing all the data in the same folder as the Python script opened. The "timestamp" column was deleted, the "words" column (cell B1) was renamed to 0, "terms" (cell C1) to 1, "TP9" (cell D1) to 2, "AF7" (cell E1) to 3, "AF8" (cell F1) to 4, and "TP10" (cell G1) to 5. The remaining columns were deleted. Then, all occurrences of "A" in the file were replaced with 0, all occurrences of "B" within the file to 1, and all occurrences of "C" to 2. The file was renamed to include the date and time of the recording.

Preprocessing the data

This file was uploaded into Kaggle, and from there imported into a Google Colaboratory Ipython notebook with hardware acceleration using a T4 GPU (graphics processing unit). The file was unzipped and converted to an array. Next, a Butterworth filter was applied to each column of data in each file (excluding the first column, the label column). Since there are only about 300 samples for both training and testing, data augmentation was performed based on findings by Wang, et. al. (2018). This involved copying the data twenty-nine times and applying Gaussian noise to the data according to the equation with a standard deviation of 0.2 with a naming convention of aug_1, aug_2, etc. Each of these files was then converted to an array with shape (298, 256, 4), or, an array with 298 elements, each of which is an array with shape 256 × 4. Each 256 × 4 represents one sample (sampling rate of the Muse 2 EEG is 256, the 4 corresponds to the channel number). The label was stored in a separate array with shape (298, 3), or an array with 298 elements, each with shape (3,). For example, the label for letter A would be [1. 0. 0.]. The first 238 of these terms (first 80%) of each array were set off for training, the rest for testing. Each of these training arrays were then concatenated into one with length 238×30 = 7140 samples. The testing array only consisted of the original data and no augmented data, and thus had a length of only 60 samples.

Training and Testing the CNN Model

The architecture of the Keras 1-dimensional CNN model was determined by hyperparameter tuning using a Bayesian optimization algorithm to find the best performing models. The model with the highest accuracy had an architecture of: C36@4 × 4- P2 - GAP - F3 according to notation used by Wang, et al (2018); during training, learning rate was set to 0.01 and the Adam optimizer was used for stochastic gradient descent. This model was trained with a batch size of 100 over 500 epochs.

Results

In total, 298 samples were collected. Figure 1 shows visual representations for one sample according to the label. Each image is four pixels wide and 256 pixels long. Four represents the number of channels: TP9, AF7, AF8, and TP10. The 256-pixel length represents data based on the sampling rate of the EEG, 256 Hz. The original image array was not modified, and these are merely for visual purposes. The values were normalized where the minimum was set to zero and the maximum 255. Lower values are represented with cooler colors, higher values as warmer colors.

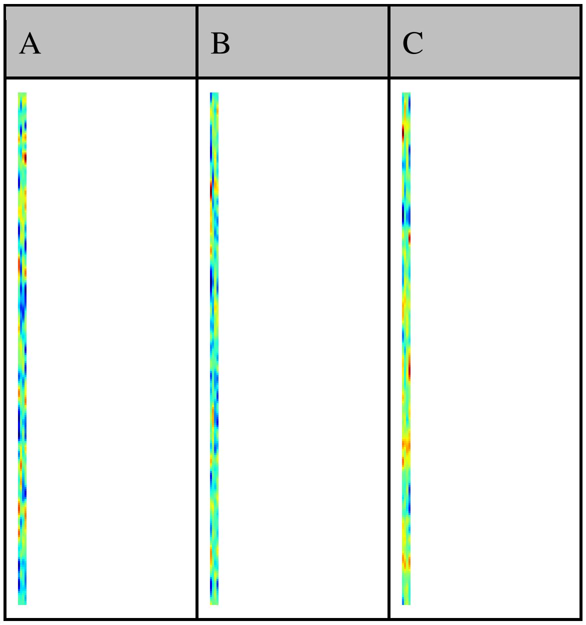

Figure 1. Pixels in each column correspond to the values for every EEG channel in a single sample corresponding to the label. Cool colors (blues and greens) represent low electrical activity and warm colors (reds, oranges, and yellows) represent high electrical activity measured by the EEG.

The neural network model with the highest testing accuracy achieved 63.33% by correctly identifying 38/60 of the terms in the testing dataset. This is visualized in Figure 2 as the model is trained and tested over 500 epochs. Figure 3 shows the value of the loss function over epochs.

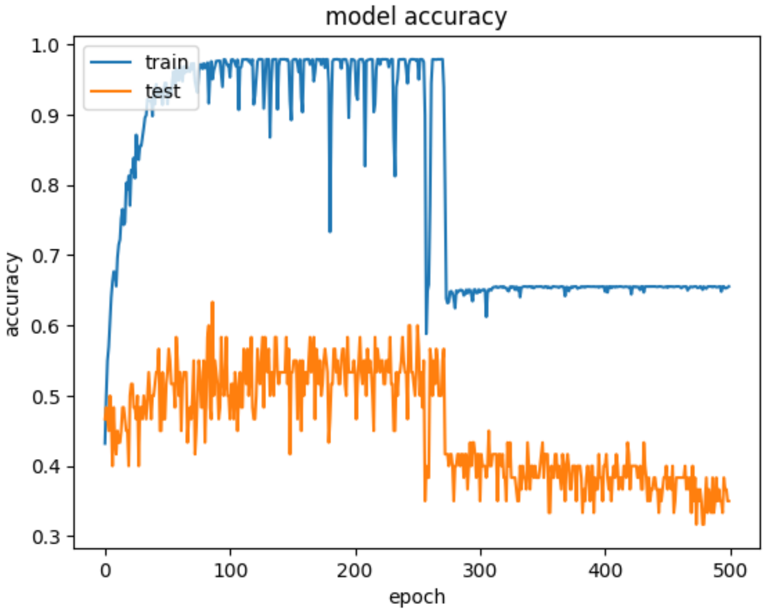

Figure 2. The model accuracy over time (epochs).

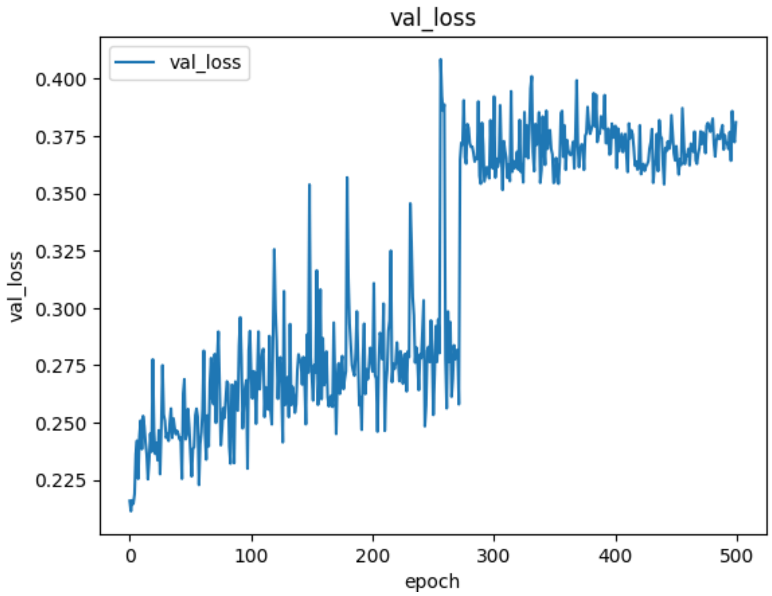

Figure 3. The model loss over time (epochs).

In more detail, the 38/60 statistic comes from, 13/20 being correctly identified as "A," 12/20 as "B," and 13/20 as "C." Figure 4 shows the confusion matrix of the model, comparing the true label to the predicted.

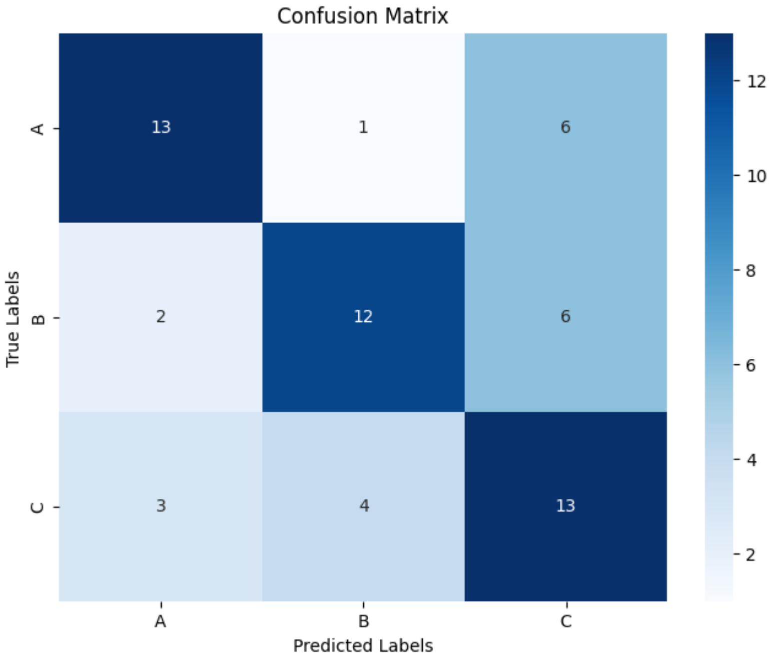

Figure 4. Confusion matrix of the model.

To confirm if these accuracies were achieved because the model was learning—and not just getting lucky with guesses—the researcher calculated the p values. The p value was calculated using the binomial distribution formula to find the probability of classifying 38/60 samples correctly with the assumption that it was achieved through luck with a 33.33% chance of classifying each of these samples correctly. The resulting p value was approximately 1.9316×10^{-6}. Values less than 0.05 are generally considered statistically significant.

Discussion

The goal of this study was to create a neural network model capable of recognizing data of subvocalized letters. In this study, statistically significant results indicate that the model has in fact learned to differentiate between the subvocalized letters “A,” “B,” and “C.” This model was trained on data collected using a relatively low-cost and commercially available EEG device.

Interestingly, the optimal model found by the hyperparameter tuning with Bayesian optimization did not include a fully connected dense layer except for the output layer. This model was relatively simple to other CNN models used for processing EEG data. The model also was clearly quickly overfitting to the training dataset by achieving accuracies higher than 95% while having much lower accuracies in the testing dataset, as seen in Figure 2. Depicted in Figure 3, the loss of the model over time also seemed to steadily increase during the first 250 epochs, even though the testing accuracy was quite high and still increasing.

Although a 63.33% accuracy is evident and indicates that using a model to do such a task is possible, higher accuracies are desirable. It should also be tested whether the same model can work on multiple people, or if a unique model is required for every person or even for every session. In this study, the researcher only applied a bandpass filter on the data, whereas in the future the researcher hopes to apply more advanced methods of preprocessing, such as feature extraction using differential entropy or principal component analysis. Another possible method of improving accuracy would be allowing the model to make use of multiple convolution layers during hyperparameter tuning, although this would require more powerful GPUs. Using a more advanced EEG device, the researcher hopes to collect more data of more letters, as well as with more samples per letter. Once accuracies of about 90% or higher are achieved with a similar setup, the researcher plans to run the model in real time, stringing together classified letters to form strings of commands or sentences and/or to map individual letters to predetermined tasks. This could be done through an application on a phone or computer running the model on the back end.

Conclusion

Based on the extremely low likelihood that the accuracy achieved by the model were the result of guesswork, it can be reasonably concluded the model was able to learn from this data. Whereas previous research done to use BCIs/EEGs to extract thoughts from the brain involves expensive and/or invasive techniques, the researcher was able to use a non-surgical method, relatively low-cost and commercially available EEG to recognize, with statistically significant accuracy, the subvocalized letters “A,” “B”, and “C.” Further research could provide a practical solution to thousands of people who find difficulty in the traditional methods of communication (writing, typing, speaking, etc.). This includes those with ALS or locked-in syndrome who are unable to communicate or find difficulty in doing so. The methods used in this experiment are also intuitive since it only involves subvocalizing letters.

Limitations

Possible limitations of the study included only being able to get 5 minutes of data due to a certain degree of unreliability of the EEG used for long periods of time when the connection may falter, as well as the ability to focus. The EEG used also has noticeable differences in data quality across multiple recording sessions, so the researcher opted to use the data from a single session. Breathing patterns may also have influenced the data, such as always taking a breath coinciding with the subvocalizing of the letter “A,” although during recording the researcher tried to be conscious of this to avoid it as much as possible. A way to mitigate this would be adding breaths as noise to the data during augmentation.

References

Birbaumer, N. (2006). Breaking the silence: brain–computer interfaces (BCI) for communication and motor control. Psychophysiology, 43(6), 517-532. https://doi.org/10.1111/j.1469-8986.2006.00456.x

Chaudhary, U., Vlachos, I., Zimmermann, J. B., Espinosa, A., Tonin, A., Jaramillo-Gonzalez, A., ... & Birbaumer, N. (2022). Spelling interface using intracortical signals in a completely locked-in patient enabled via auditory neurofeedback training. Nature communications, 13(1), 1-9. https://doi.org/10.1038/s41467-022-28859-8

Kapur, A., Sarawgi, U., Wadkins, E., Wu, M., Hollenstein, N., & Maes, P. (2020). Non-invasive silent speech recognition in multiple sclerosis with dysphonia. In Machine Learning for Health Workshop (pp. 25-38). PMLR. https://proceedings.mlr.press/v116/kapur20a.html

Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., ... & Chang, E. F. (2021). Neuroprosthesis for decoding speech in a paralyzed person with anarthria. New England Journal of Medicine, 385(3), 217-227. https://doi.org/10.1056/nejmoa2027540

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Jr, Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. Human brain mapping, 16(4), 219–227. https://doi.org/10.1002/hbm.10046

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non invasive brain recordings. Nature Neuroscience, 1-9. https://doi.org/10.1038/s41593-023-01304-9

Tyler, L. K., & Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363(1493), 1037–1054. https://doi.org/10.1098/rstb.2007.2158

Wang, F., Zhong, S. H., Peng, J., Jiang, J., & Liu, Y. (2018). Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. In MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part II 24 (pp. 82-93). Springer International Publishing. https://doi.org/10.1007/978-3-319-73600-6_8

Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. Nature, 593(7858), 249-254. https://doi.org/10.1038/s41586-021-03506-2

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., ... & Vaughan, T. M. (2000). Brain-computer interface technology: a review of the first international meeting. IEEE transactions on rehabilitation engineering, 8(2), 164-173. https://doi.org/10.1109/tre.2000.847807

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology, 113(6), 767–791. https://doi.org/10.1016/s1388-2457(02)00057-3