TARTU UNIVERSITY
Institute of Computer Sciences
Computer Science curriculum

Markus-Oliver Tamm

Vowel Classification from Imagined Speech Using Machine Learning

Supervisors: Yar Muhammad, Ph.D
Naveed Muhammad, Ph.D

Tartu 2020

Vowel Classification from Imagined Speech Using Machine Learning

Abstract:

Imagined speech is a relatively new EEG neuro-paradigm, which has seen little use in BCI applications. Imagined speech can be used to allow physically impaired patients to communicate and to use smart devices by imagining desired commands and then detecting and executing those commands in a smart device.

The goal of this research is to verify previous classification attempts made and then design a new, more efficient neural network that is noticeably less complex (fewer number of layers) that still achieves a comparable classification accuracy. The classifiers are designed to distinguish between EEG signal patterns corresponding to imagined speech of different vowels and words. This research uses a dataset that consists of 15 subjects imagining saying the 5 main vowels (a, e, i, o, u) and 6 different words 2 previous researches on imagined speech classification done on this same dataset are replicated and the replication results are compared. The pre-processing of data is described and a new CNN classifier with 3 different Transfer Learning methods are described and used to classify EEG signals. Classification accuracy is used as the performance metric.

Keywords:

Electroencephalography, Classification, BCI, Machine learning, Imagined speech, Convolutional neural network, Random forest

CERCS:

P160 programming, P176 Artificial intelligence, B640 neurology

Täishäälikute klassifitseerimine kujutletud kõnest masinõppega

Lühikokkuvõte:

Kujutletud kõne on võrdlemisi uus elektroentsefalograaphia neuro-paradigma, mida on aju-arvuti liidese(ingl. Brain-Computer Interface, lüh. BCI) rakendustes vähe kasutatud. Kujutletud kõne abil saab füüsiliste puudega inimestele anda võimaluse suhelda ning nutiseadmeid kasutada. Selleks tuleb neil lasta ette kujutada vastavaid sõnu või käsklusi ning ettekujutamisest tekkivaid ajulained seejärel BCI seadme abil mõõta ning klassifitseerida.

Selle uurimuse eesmärk on kinnitada eelnevalt tehtud uurimuste tulemusi ning seejärel luua uus, vähemate kihtidega, tõhus närvivõrk mis siiski suudab saavutada võrreldava täpsuse klassifitseerimisel. Klassifitseerijate eesmärk on eristada erinevate sõnade või häälikute kujuteldud hääldamisest tulenevaid signaale. Selles uurimuses kasutatakse andmebaasi, mis sisaldab 15 erineva katsealuse EEG andmeid 5 erineva täishääliku(a,e,i,o,u) hääldamise kohta. Lisaks korratakse kahe eelneva uurimuse tulemusi ning võrreldakse saadud tulemusi uurimuste originaaltulemustega. Töös kirjeldatakse andmete eeltöötlust, närvivõrgu ülesehitust ning kolme erinevat ülekandmisõppe meetodit, mida kasutati. Loodud närvivõrku kasutatakse ajusignaalide klassifitseerimiseks ning tuuakse välja, mis klassifitseerimistäpsused saadi ning mis tegurid neid tulemusi mõjutasid.

Võtmesõnad:

Elektroentsefalograafia, klassifitseerimine, BCI, masinõpe, kujutletud kõne, konvolutsioon-neuraalvõrk, otsustusmets

CERCS:

P160 programmeerimine, P176 Tehisintellekt, B640 neuroloogia

Table of contents

Terms and keywords:

Brain-computer interface (BCI) – communication pathway between an external device and the brain.

Electroencephalography (EEG) – an electrophysiological way of monitoring and recording the electrical activity of the brain.

Random-forest (RF) – learning method for classification that constructs a multitude of decision trees and outputs the class that is the mode of the classes of the individual trees.

Support vector machine (SVM) – a supervised learning model with associated learning algorithms that analyses data used for classification.

1. Introduction

Imagined speech is the act of internally pronouncing words or letters without actually producing any auditory output. Recording and differentiating between these pronounced words could be crucial in allowing physically impaired patients to communicate with their caretakers in a natural way. Some research has already been done on this subject and respectable results have been achieved already and most of the developed classification models offer respectable classification performance.

The objective of this research is to develop a classifier that uses deep learning to classify Electroencephalography(EEG) signals associated with imagining pronouncing vowels which then could be used in BCI applications. Here the main focus is on developing a model, that is less complex to try and achieve classification performances similar to the already developed models to see if higher complexity is required in imagined speech classification. Transfer learning(TL) is also used to try and improve the accuracy optained by the model. This model should be re-trainable on a single subject to achieve even higher individual classification performance.

This research consists of seven sections. The second paragraph gives an overview of what has been done previously on imagined speech classification. The third paragraph describes the methodology used in this research to replicate both of the results of the previous researches and to develop a new classifier. The fourth chapter shows the replication results and the results obtained by the new proposed model and examines them. The fifth chapter discusses the results achieved by the new model as well as the results obtained from the replications. The final chapter gives a summary of the entire paper and suggestion on what to do differently in the future.

2. Literature review

This chapter contains a brief overview of the topics this research deals with. First, an introduction to EEG and BCI is given. Following that is a review of how most EEG data is collected. After that, the most used pre-processing and classification methods are discussed and finally, the recent papers on the dataset used in this research are reviewed.

2.1 EEG and BCI

The American Epilepsy Society [1] states that EEG is the act of measuring the electrophysiological activity of the brain. According to them, EEG is generated by cortical pyramidal neurons in the brain and it becomes detectable when large groups of neurons emit those signals simultaneously. EEG is best used for detecting stroke and epilepsy in patients [1] but it can also detect other neurological conditions like dementia, brain tumours and sleep disorders [2]. Because of its high versatility, EEG has also found use in brain-computer interface systems where it helps partially or fully paralysed patients communicate with others and their environment [3].

Brain-computer interface (BCI) is a neural pathway through which signals coming from the brain can control an external device. BCI-s serve 2 main purposes: to facilitate neural recovery and to enable paralysed patients to interact with the environment through the use of external robotic devices [4]. They bring up that BCI applications can largely be divided into 2 categories: invasive and non-invasive. Invasive BCI applications involve implanting electrodes surgically, while non-invasive BCI-s have the electrodes on the surface of the scalp. Work has been going on to develop BCI-s that can help with complete locked-in syndrome [4].

Brain-computer interfaces have found use and are currently being developed to see use in a variety of fields that are brought out in this [5] review. In addition to the previously mentioned medical applications, BCIs have been successfully implemented in games to pilot a helicopter [6]. Even more impressive are advances made in using neural network to control a robot [7] and speech synthesis from decoding spoken sentences [8]. Other potential uses include security authentication, education, neuro-marketing and smart environment.

There are 4 main neuro-paradigms of EEG that have been explored in great detail: slow cortical potentials, motor imagery, P300 component and visual evoked potentials. There is

a fifth paradigm, imagined speech, that has received comparatively less attention from researchers than the 4 main paradigms [3]. Imagined speech is a form of silent communication [9]. Brainwaves related to imagined speech of words or vowels can be recorded with a BCI and then given to a classification model that determines, which word was imagined [10]. Imagined speech offers a promising way for severely disabled people to communicate with their caretakers and as such should be considered an important field of study for researchers.

2.2 Data acquisition methods

Data collection of EEG faces a great number of difficulties as is mentioned by Aina et al. [11]. Firstly, EEG data collection is expensive. It requires specialised equipment, software and personnel to conduct EEG trials. Secondly, EEG data has a very low signal-to-noise ratio, which makes the EEG data very artefact prone [11]. Artefacts are non-wanted signal changes that have no correlation with the task at hand and cause problems in EEG classification but a number of solutions are being used to combat artefacts. Thirdly, EEG trials are a time-consuming process [5]. Classification algorithms require a large number of samples to train effective classifiers and as such require a large number of patients or a large number of trials per patient, both of which can be quite difficult to acquire. Because of these difficulties, there aren't many EEG datasets of imagined speech data publicly available.

Imagined speech data can be acquired through invasive methods such as electrocorticography (ECoG) and through non-invasive methods such as EEG or functional magnetic resonance imaging(fMRI). Non-invasive methods such as EEG are of greater interest to researchers because they are less dangerous to the subjects and represent less of a risk to their health [12]. EEG also has far less complex instrumentation when compared to fMRI [12] and also has greater portability due to its smaller devices. Because of its advantages over other methods, this research also exclusively uses imagined speech data gathered through EEG.

2.3 Data pre-processing methods

Imagined speech decoding process consists of three phases: pre-processing of data, feature extraction and classification. Pre-processing usually involves artefact removal and band-pass filtering. Feature extraction involves typical BCI feature extraction methods like auto-regressive coefficients [13], spectro-temporal features [14] and common-spatial patterns [15]. Several machine learning techniques have been used to classify imagined speech data. Among them are support vector machines(SVM) [16], Linear Discriminant Analysis(LDA) [13] and Random Forests(RF) [17].

Imagined speech data pre-processing is an important step in improving the effectiveness of a classifier. Not all EEG data that is collected is useful in classifying imagined speech. Furthermore, imagined speech signals have a low signal-to-noise ratio [9] and because of this pre-processing is important. The deep learning based EEG review [18] shows that 72% of previous works have applied some form of pre-processing. Most often utilized pre-processing techniques were down-sampling, band-pass filtering and windowing. Down-sampling is used to better bring out the features distinct to each class, band-pass filtering is used to limit data to the most relevant bands and windowing is used to create more samples. This thesis uses down sampling and band-pass filtering.

Artefact removal is also important considering that imagined speech data collected from EEG has a low signal-to-noise ratio. As is mentioned in [19], artefact removal may be important to get good classification performance. Although artefact removal has been shown to make better classifiers [19], in the deep learning based EEG review almost half of the papers did not use any artefact handling [18]. It is possible, that deep neural networks allow you to pass the artefact removal process by giving the task of extracting relevant data from EEG data to the neural network. It is inconclusive whether or not giving the task of artefact handling to the neural network gives better accuracy rather than doing it manually before the neural network. This studydoesn’t use artefact removal in the propose model.

2.4 Classification approaches

Several different machine learning approaches have been used to classify imagined speech data. Among them are support vector machines (SVM) [16], Random Forests (RF) [20] and

Linear Discriminant Analysis(LDA) [21]. SVM has been the most often used method but none of them has proven to be superior to the others. All approaches have received comparable results in imagined speech classification. Deep learning has also successfully been applied to BCI tasks, such as motor imagery [22] and SSVEP [23]. Out of the deep learning approaches convolutional neural networks(CNN) have been the most often used when it comes to BCI and EEG applications. CNNs have already been used to classify imagined speech data, although the number of studies is still quite low. The complete list of deep learning applications related to BCI and EEG can be found in the review by Y. Roy et al. [18].

Deep learning has also been used in decoding EEG data. The review by Roy [18] shows that a large percentage of deep learning based classifiers use some sort of pre-processing with down sampling and band-pass filtering being the most common methods. About 47% of deep learning models didn't use any artefact handling even though research by Yang et al. [19] showed that artefact removal could be crucial in getting high classification accuracy. Out of the feature extraction methods the most popular were raw EEG(automatic feature extraction) and frequency-domain methods [18].

2.5 Recent approaches

This study uses the dataset provided by Coretto et al. [24]. This dataset contains the EEG data of the imagined speech of 5 vowels (a,e,i,o,u) from 15 subjects. A few researchers have already tried to classify the data from this dataset. The dataset authors themselves [24] provided the initial classification where they down sampled the data and used a RF classifier to get an accuracy of 22.32% for vowels. Garcia-Salinas et al. [25] also down sampled the data and used wavelet transform with Alternated Least Squares approximation with a linear SVM classifier and got an average inter-subject accuracy of 59.70% for words on the first 3 subjects. Cooney et al. [26] used a deep and a shallow CNN to classify word-pairs from the dataset and used independent component analysis with Hessian approximation to achieve an average accuracy of 62.37 and 60.88 for the deep and shallow CNNs respectively. Cooney et al. [27] also classified the five vowels from this dataset. Pre-processing involved down sampling the data to 128Hz and using ICA with Hessian approximation for artefact removal. They used a CNN with 6 convolutional layers and also used 2 different TL methods to improve cross-subject accuracy, both of which were successful.

Another notable approach was done by Tan et al. [28] where on a different dataset they used Extreme Learning Machine(ELM) to classify raw EEG data. ELM is a feed-forward neural network with a single hidden layer and a varying number of hidden units. They trained and tested the ELM and compared it with other machine learning techniques on four different datasets. Their results showed that ELM outperformed the other classifiers in almost all cases. The ELM model was also faster to train than the SVM model, although it was slower than the LDA.

In this study different pre-processing techniques are tested to find the best customization for classifying imagined speech data. Here we will use different down sampling amounts, different data selection techniques and artefact removal and normalization options to find the best combination. Then a convolutional neural network is proposed and used to classify EEG data of imagined speech of the 5 main vowels (a,e,i,o,u).Transfer learning is also used to try and improve upon the accuracy achieved by our proposed neural network.

3. Methodology

This chapter describes the methodology. The first section describes the dataset used by this research and all the studies that are replicated and the second part describes the replication protocol for both replicated studies. Following that is the section for describing our proposed model, including the architecture of the proposed neural network, pre-processing techniques are used and the transfer learning methodology is implemented.

3.1 Dataset

The dataset used in this studyis recorded by Coretto [24] et al. in the Faculty of Engineering at the National University of Entre Ríos (UNER). 15 subjects performed overt and covert speech tasks while their EEG signals were recorded. The data consists of trials in which the subjects had to pronounce the five main vowels “a”, “e”, “i”, “o”, “u” and the six Spanish words corresponding to the English words for “up”, “down”, “left”, “right”, “back”, “forward”, although the word part of the dataset is not used in classification of this study. The experimental protocol for pronouncing the vowels and words had a 2 second pre-trial period where the subjects were shown their target. Following that there was a 4 second period during which the imagined pronunciation of the target took place. The vowel was to be pronounced during the whole 4 seconds while the word was pronounced 3 times during the 4 second period. After that, there was a 2 second rest period. The protocol is illustrated on Figure 2. Only the vowel part of this dataset was used in this study The EEG signals were recorded with an 18-channel Grass analogue amplifier and were sampled at 1024 Hz. The electrodes were positioned according to the 10-20 international system at positions F3, F4, C3, C4, P3 and P4, as shown in Figure 1.

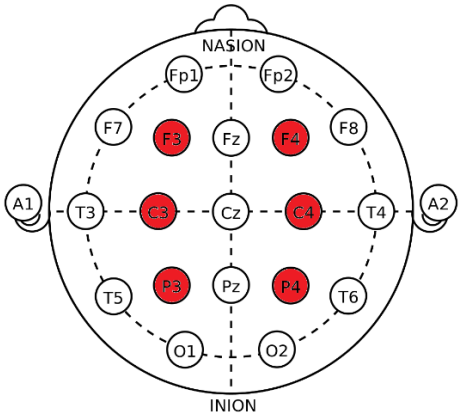

Figure 1. 10-20 international electrode positioning[1]; data is from marked electrodes

[1] https://en.wikipedia.org/wiki/10%E2%80%9320_system_(EEG)

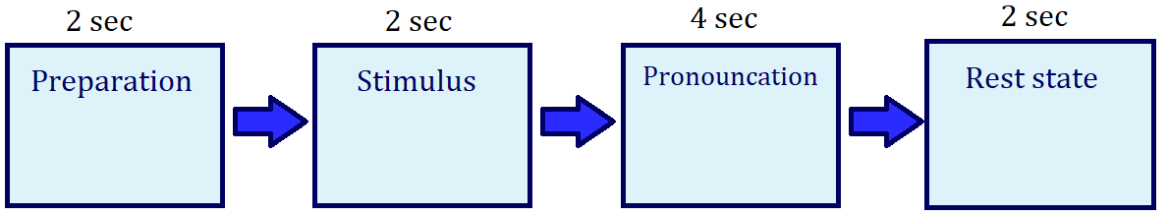

Figure 2. Recording protocol used to collect data from subjects

Following section describes the methodology for replicating the results of the two chosen previous imagined speech classification attempts made on the same dataset used in this study. The first replication is an initial classification attempt by the collectors and authors of the dataset themselves [24] and the second replication is done by Cooney et al. [27] where a CNN with more than 30 layers was used.

3.2 Validating and verifying the existing studies

3.2.1 Replication of the first study

The first replication is done on the results achieved by the same people who collected the data and made the dataset used in this study [24].They used a Random Forest classifier to classify both the words and vowels of the dataset. They downscaled the data eight times and got an effective sampling rate of 128 Hz. After that they computed a Discrete Wavelet Transform (DWT) with five levels of decomposition for each EEG channel by selecting the mother wavelets from the Daubechies family. After that they calculated the Relative Wavelet Energy (RWE) for each channel in each sample and used the RWE for each decomposition except for the first to form the feature vector with which to classify samples. A short guide on how to do this in Python is given here[2].

Like the target paper, the first decomposition (corresponding to frequency bands 32-64) was not used in calculating the RWE. Wavelet transform and decomposition were implemented in Python using the "Pywt" library, which is a publicly available tool for implementing and using wavelet transform in Python. The experiment and Random Forest classifier training were conducted using a publicly available machine learning and feature analysis tool Weka[3].

[2] http://ataspinar.com/2018/12/21/a-guide-for-using-the-wavelet-transform-in-machine-learning/
[3] https://www.cs.waikato.ac.nz/ml/weka/

A random forest classifier made of 100 trees with five randomly chosen attributes was used to classify the data. The classifier used 10-fold cross validation and was trained 10 times.

3.2.2 Replication of the second study

The second replication is of a study made by Cooney et al. [27] on the same dataset used in this study and by the study replicated in the previous section. They used a deep CNN with 34 layers to classify the vowels. Down-sampling, artefact removal and data scaling were used as part of the pre-processing.

The CNN used is divided into seven sections, where the first section is the initial convolution part. This part contains the input layer as well as two initial convolution layers, the first being temporal convolution and the second being spatial convolution. The following five sections are all very similar convolution sections consisting of five layers each. The final section is the classification section which has the softmax activation layer for classification. All of Cooney's research was implemented in Python using Tensorflow[4] and Keras[5] libraries.

The pre-processing started with down-sampling the data to 128Hz. FastICA was used for artefact removal on trials marked by the dataset authors as having artefacts and scikit-learn's robust scaler was used for scaling the data. Data was shuffled prior to feeding it to the network. This data was then given to the 34-layer constructed CNN. All the details except for the pooling layers are provided by Cooney et al. [26]. The parameters of the pooling layers were selected to be the same as Schirrmeister et al. [29], the same CNN that inspired Cooney. 5-fold cross-validation was used and the best model from each fold was used to acquire the testing accuracy. All the folds were trained for 100 epochs, and a callback was used to stop the training when the validation loss had not improved for 50 epochs which helps to reduce overfitting.

3.3 Proposed model

3.3.1 Preprocessing

For the CNN, multiple pre-processing techniques were tried and because this data is recorded at a very high sample rate, down-sampling is safe to use. Data was down-sampled

[4] https://www.tensorflow.org/
[5] https://keras.io/

four times down to 256Hz and data was restructured from a simple array to a 2D array, where each element was stored as an array of the signal values at that time point. For vowels, all classes were balanced to the lowest class count on any subject of any trial type, which was 37 examples for subject 07 on “e”. No artefact removal was used. Data was split into three parts: 70% training data, 15% validation data and 15% testing data.

3.3.2 CNN

The neural network proposed in this study was inspired by the one proposed by Schirrmeister et al. [29]. This neural network consists of an initial convolution block followed by several separately viewable convolution blocks. The goal of this study is to reduce the complexity of the model and keep the number of layers in the CNN below 20 to see if a lower complexity model can still perform imagined speech classification at an optimal level. Figure 3 describes all of the layers along with their parameters. First block consists of an input layer which is immediately followed by 2 2D convolutional layers, a batch normalization and a Relu activation layers. Following that are 2 identical convolution blocks with both of them consisting of: 2Dconvolution, batch normalization, Relu activation, average pooling and dropout. The final classification part consists of a Dense layer followed by a softmax classification layer. The model was implemented in Python with Keras and Tensorflow frameworks.

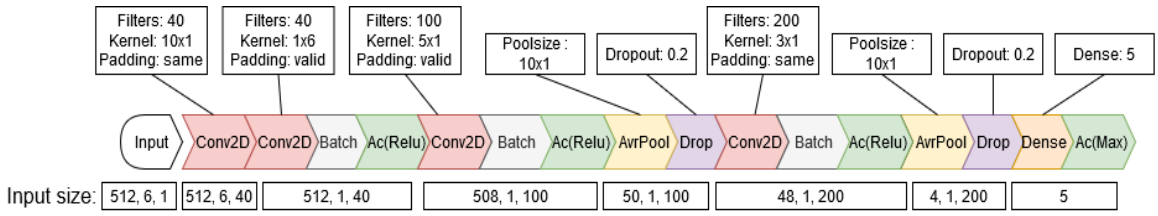

Figure 3: The proposed model’s layers and their parameters

3.3.3 Training

The training of the models took place on the servers of the High Performance Computing Center of Tartu University on the Rocket clusters’ Falcon GPU nodes. The servers are made of 135 CPU nodes and 3 GPU nodes specifically made for machine learning. GPU nodes have 2 24 core CPU-s and 512 GB of RAM and 8 NVIDIA Tesla P100 GPUs. The specific

training server details can be found here[6]. Several callbacks were also used to select only the model with the highest validation accuracy. All three CNN-s were trained with the Adam optimizer and sparse categorical cross entropy loss function. Adam optimizer works well in practice and compares favorably to other adaptive optimizers [30]. A learning rate of 0.001 was also recommended to be used with the Adam optimizer. The initial CNN was trained for 100 epochs. Early stopping was also in place to stop training if the validation accuracy did not improve in 50 epochs in order to stop overfitting.

3.3.4 Transfer Learning

Three TL approaches were used with the proposed model to try and improve the accuracy of the model. Transfer learning is a machine learning technique which aims to improve classification accuracy on a single subject. All methods first train a model with the data from all subjects but one and then use the combined weights to specifically optimise the model for the one subject by fine tuning some or all of the convolutional layers with that one person's data. TL training sessions were 40 epochs long.

Transfer learning method 1

The first TL method freezes the whole base model and then unfreezes the first two convolutional layers to be retrained on the new subject's data. This method locks all weights on the pre-trained network and allows only the first two convolutional layers to change when the network gets fine-tuned with data from S01. The three TL methods are illustrated on Figure 4, Figure 5 and Figure 6 respectively.

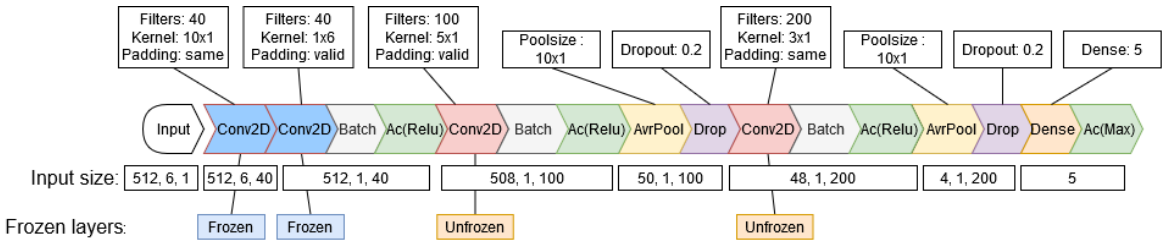

Figure 4. TL method 1: Initial layers are frozen

[6] https://hpc.ut.ee/rocket-cluster/

Transfer learning method 2

The second TL method freezes the whole base model and then unfreezes the last two convolutional layers to be retrained on the new subject's data.

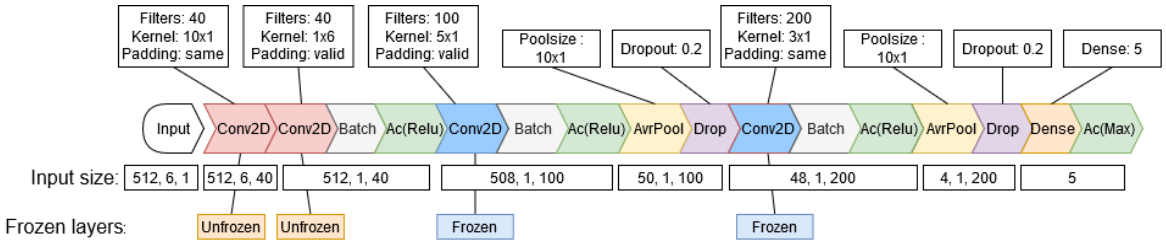

Figure 5. TL method 2: Latter layers are frozen

Transfer learning method 3

The third TL method uses a combination of both previous methods to try and unfreeze all of the convolutional layers to retrain on new subject data.

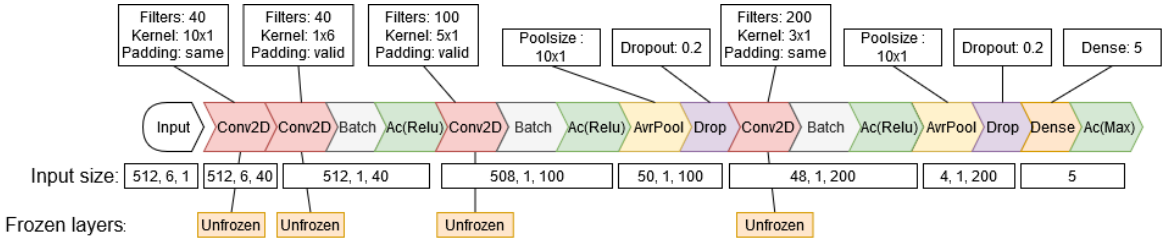

Figure 6. TL Method 3: All layers are unfrozen

4. Results

Here we report both the classification results we got for both the replicated and the proposed model. First there will be results for the replication and then the results for our new convolution model and all of the TL methods.

4.1 Replication results of the first study

The first replication was done on the study by the authors of the dataset used in this study [22]. The mean accuracies for all subjects are presented in Figure 7. The mean accuracy over all subjects over 10 iterations with different seeds was 22.81%, which is just slightly above the first study's mean accuracy. All of the subjects except S13 got a mean accuracy above the chance level (20%). The random seed numbers used by the authors of the first study are not brought out in that paper so the random seeds used here are probably different than the ones used there and are probably the cause of slightly different results.

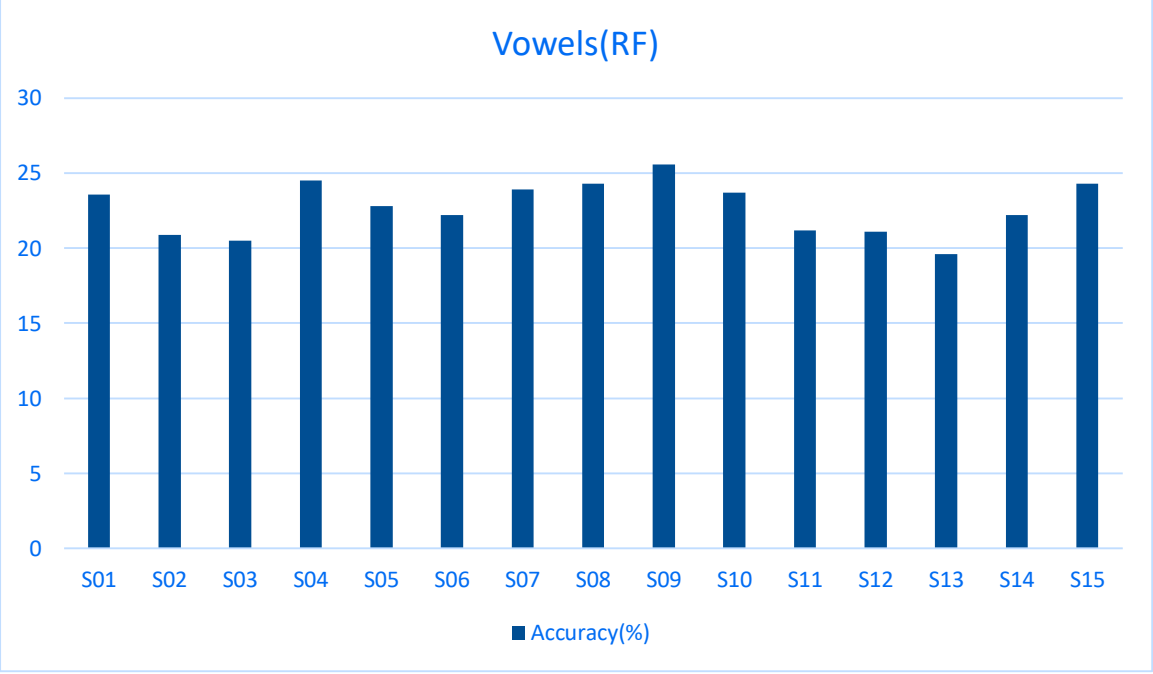

Figure 7: First study replication results per subject

4.2 Replication results of the second study

The second replication was done on the study by Cooney et al. [25]. The mean accuracies for all subjects are presented in Figure 8. The mean accuracy over all subjects was 30.21%. This is considerably lower than the mean accuracy achieved by the authors of the second study (32.75%). This can be explained by the different artefact removal technique used and

also possibly by different parameters used in the pooling layers as those are not clearly defined in the second study. The implementation platform can also slightly influence accuracies.

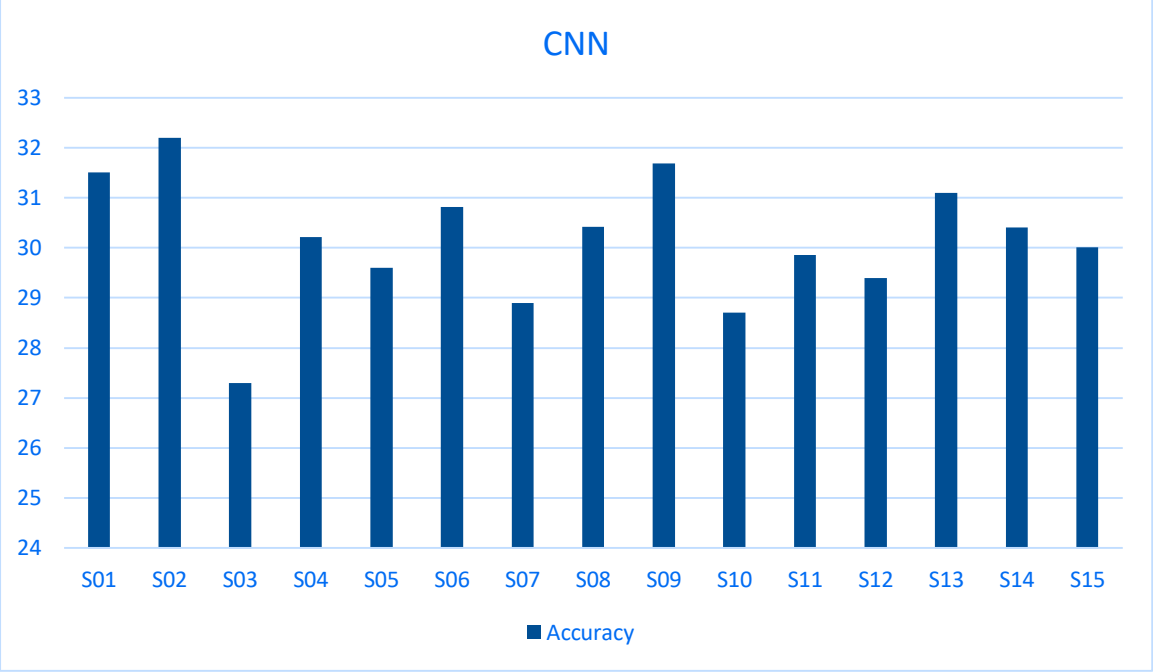

Figure 8: Second study replication results per subject

4.3 Results of the proposed model

The mean accuracies for the CNN and all of the TL methods are brought out in the Table 1 below. The proposed CNN with 3 convolutional layers managed to beat the Random Forest classifier (23.98% vs 22.72%) which is presented in Table 1, that was used by the initial makers of the dataset, but it didn't beat the significantly deeper and more complex neural network proposed by Cooney et al. on this same dataset (23.98% vs 32.75%)(Table 1) but proposed model in this study is less complex. Out of all the TL methods the best results were achieved by the first TL method.

	First replicated study	Second replicated study	Proposed model
Accuracy	22.72%	32.75%	23.98%
Replication	22.81%	30.21%	-

Table 1: Comparsion between CNN accuracies

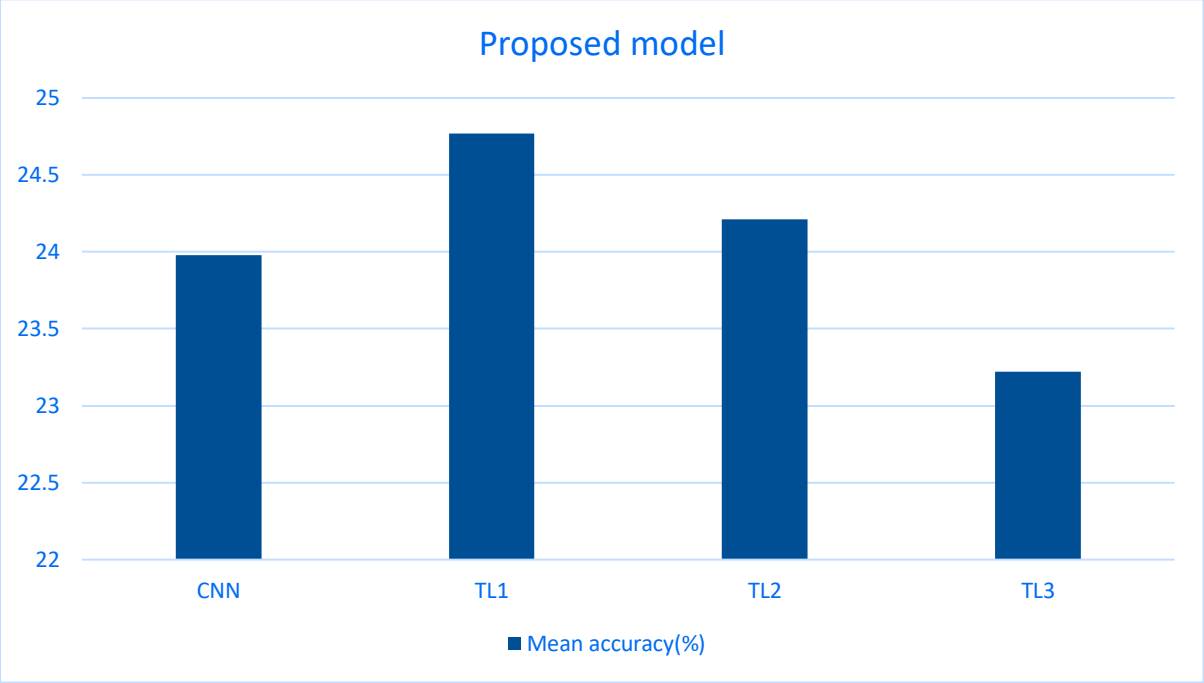

Figure 9: Comparsion of different TL methods used on the proposed model

Figure 9 gives an overview of how well TL methods performed when compared to not using any TL methods. TL improved the accuracy in two out of three cases. The biggest improvement was seen when the fine-tuning took place on the initial convolutional layers and a considerable improvement was also seen when fine-tuning the latter layer. Using TL on all layers lowered the overall accuracy.

5. Discussion

Replication provided very similar results in the first replicated study and considerably lower results in the second replicated study but that can be attributed to slightly differing tools and methods used in the replication. The proposed model was moderately successful but definitely underperformed when compared to a more complex model but the proposed model is less complex

5.1 Replicated results

The overall mean accuracy from the first replication of study is very similar to the result achieved in this study. The overall accuracies differ by only 0.09%. Accuracies between other subjects differ more. For example, the replicated papers' best individual accuracy was achieved on subject S14 but in this case was achieved on subject S09. The lowest accuracy was achieved on S12 on the replicated paper's case and S13 in this case. These differences can be attributed to the fact that the replicated paper didn't specifically mention the seed numbers used in training their RF classifier. These numbers were guessed when replication took place and most probably differ from the ones chosen in the original work. Another place where a minor divergence could have taken place is in the Wavelet Transform implementation. Both original work and this work implemented their classifier in Weka but the original work didn't mention in which language or platform they did the Wavelet Transform. In this work, it is done in Python with the 'pywt' library but that might not be the case in the replicated paper and different platforms can implement the Wavelet Transform in a slightly different way.

The overall mean accuracies between the second study of replication differ a bit more. 30.21% vs 32.75% represent significantly worse results. The main difference between this and replicated study is that a different artefact removal technique was used to remove artefacts from data. In the replication artefact removal was used only on the trials marked by the authors of the dataset. The second study doesn't mention if they used artefact removal on all of the trials or only on the ones marked by the dataset authors themselves. Also again the original work doesn't mention which platforms were used to implement the artefact removal and classifier training. As such the chosen platforms might be different and implement the CNN in a slightly different way which can cause differences in accuracies.

5.2 Results of the proposed model

The overall mean accuracy achieved over all subjects with the new model is 23.98%, which is above the chance line(randomly guessing) of 20.00% and is better than the accuracy from the first replication study (22.72%) but is a lot lower than the results from the second replication study (32.75%).

The TL methods helped slightly to increase the accuracy but still didn't manage to achieve desirable results. This all indicates that classifying EEG data takes quite a bit of effort, since EEG data is highly personalized and the same imagined action gives off different signals when it is measured from different people. This could mean that using neural networks with a small amount of layers to classify EEG data could be an impossible task and deeper networks are to be preferred.

One reason for generally low accuracies across the board could be that this is a relatively high class-count classification task with a very limited amount of data acquired from quite a high number of subjects. Machine learning models need a lot of data per class do learn the specific features relevant to each class and this dataset has a relatively small sample size to provide. The low accuracies achieved in this study and also previous studies into this dataset could also be indicative of the poor quality of this dataset, where the relevant features are not easily accessible for the classifiers. Perhaps the recording protocol could be optimised to have larger pauses between experiments as they have been shown to give better results[31].

6. Summary

First, the results of two different previousstudies on the same dataset were replicated and confirmed. This dataset, which contains the imagined speech data from 15 subjects was used to train our own classifier, which was then used to classify the imagined speech data. Three different TL methods were tried to improve the accuracy of the model. The first method fine-tuned the first twoconvolutional layers, the second fine-tuned the latter layers and the third tried to improve all convolutional layers.

The replication results were very similar to the ones achieved by their original authors. The dataset paper used a Random Forest classifier to classify both words and vowels. The replication achieved very similar results in vowels (22.81% vs 22.72%). For the Cooney´s results the accuracies slightly differed (30.21% vs 32.75%) but this can be explained with using a different artefact removal technique that was native to the platform that was used to implement the CNN. After the replication a new CNN was constructed to try to achieve similar results to the previously done works but with limiting the size of the CNN to a maximum of 18 layers. Different TL techniques were also tried. Achieved accuracies for the CNN (23.98%) and all the TL methods (24.77%, 24.12%, 23.22%) were better than the one achieved by the authors of this dataset (22.72%) but fell short of the one achieved by a considerably deeper network.

In the discussion, it was argued that when it comes to classifying the highly personalised EEG imagined speech data, deeper neural networks should be used and additional data generation could prove very useful in improving the performance of the classifiers. TL in neural networks also showed promise and should be used in the future to help get better accuracies. It was shown that TL improved the classification accuracy the most when it was applied to the latter layers of the neural network.

In the future when it comes to using EEG data in imagined speech classification it is recommended that if the amount of data is low then models with a high number of layers should be used or some additional data generation should be used to generate better classification accuracies. TL showed potential and it’s effect on both shallow and deep neural networks as well as in the cases of low and high amount of data could be further explored.

References

[1] J. W. Britton, L. C. Frey, J. L. Hopp, P. Korb, M. Z. Koubeissi, W. E. Lievens, E. M. Pestana-Knight and E. K. S. Louis, Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants [Internet], Chicago: American Epilepsy Society, 2016.

[2] K. Blocka, "EEG (Electroencephalogram)," Healthline, [Online]. Available: https://www.healthline.com/health/eeg. [Accessed 10 01 2020].

[3] R. Ramadan and A. Vasilakos, "Brain Computer Interface: Control Signals Review," *Neuroscience,* vol. 223, pp. 26-44, 2017.

[4] U. Chaudhary, N. Birbaumer and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Reviews Neurology,* no. 12, pp. 513-525, 2016.

[5] S. N. Abdulkader, A. Atia and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egyptian Informatics Journal,* vol. 16, no. 2, pp. 213-230, 2015.

[6] A. S. Royer, A. J. Doud, M. L. Rose and B. He, "EEG Control of a Virtual Helicopter in 3-Dimensional Space Using Intelligent Control Strategies," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 18, no. 6, pp. 581 - 589, 2010.

[7] B. J. Edelman, J. Meng, D. Suma, C. Zurn, E. Nagarajan, B. S. Baxter, C. C. Cline and B. He, "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Science Robotics,* vol. 4, no. 31, 2019.

[8] G. K. Anumanchipalli, J. Chartier and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences.," *Nature,* vol. 568, no. 7753, pp. 493-498, 2019.

[9] K. Brigham and B. V. K. V. Kumar, "Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy," in *2010 4th International Conference on Bioinformatics and Biomedical Engineering,* Chengdu, 2010.

[10] K. Brigham and B. V. K. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS),* Washington, DC, 2010.

[11] A. Puce and M. S. Hämäläinen, "A Review of Issues Related to Data Acquisition and Analysis in EEG/MEG Studies," *Brain sciences,* vol. 7, no. 6, p. 58, 2017.

[12] R. Bogue, "Brain-computer interfaces: Control by thought," *Industrial Robot,* vol. 37, no. 2, pp. 126-132, 03 2010.

[13] Y. Song and F. Sepulveda, "Classifying speech related vs. idle state towards onset detection in brain-computer interfaces overt, inhibited overt, and covert speech sound production vs. idle state," in *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings,* Lausanne, 2014.

[14] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Brisbane, 2015.

[15] C. S. DaSalla, H. Kambara, M. Sato and Y. Koike, “Single-trial classification of vowel speech imagery using common spatial patterns,” *Neural Networks,* vol. 22, no. 9, pp. 1334-1339, 2009.

[16] C. Cooney, R. Folli and D. Coyle, “Mel Frequency Cepstral Coefficients Enhance Imagined Speech Decoding Accuracy from EEG,” in *2018 29th Irish Signals and Systems Conference (ISSC)*, Belfast, 2018.

[17] W. Chen, Y. Wang, G. Cao, G. Chen and Q. Gu, “A random forest model based classification scheme for neonatal amplitude-integrated EEG,” *BioMedical Engineering OnLine,* vol. 13, no. Suppl 2, 2014.

[18] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk and J. Faubert, “Deep learning-based electroencephalography analysis: a systematic review,” *Journal of Neural Engineering,* vol. 16, no. 5, 2019.

[19] B. Yang, K. Duan, C. Fan, C. Hu and J. Wang, “Automatic ocular artifacts removal in EEG using deep learning,” *Biomedical Signal Processing and Control,* vol. 43, pp. 148-158, 2018.

[20] L. A. Moctezuma, M. Molinas, A. A. Torres-García and L. Villaseñor-Pineda, “Towards an API for EEG-Based Imagined Speech classification,” in *International conference on Time Series and Forecasting*, Granada, 2018.

[21] X. Chi, J. B. Hagedorn, D. Schoonover and M. D. 'zmura, “EEG-Based Discrimination of Imagined Speech Phonemes,” *International Journal of Bioelectromagnetism,* vol. 13, no. 4, pp. 201-206, 2011.

[22] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche and M. S. Hossain, “Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion,” *Future Generation Computer Systems,* vol. 101, pp. 542-554, 2019.

[23] N. Waytowich, V. J. Lawhern, J. O. Garcia, J. Cummings, J. Faller, P. Sajda and J. M. Vettel, “Compact Convolutional Neural Networks for Classification of Asynchronous Steady-state Visual Evoked Potentials,” *Journal of Neural Engineering,* vol. 15, no. 6, 2018.

[24] G. A. P. Coretto, I. Gareis and H. L. Rufiner, “Open access database of EEG signals recorded during imagined speech,” in *12th International Symposium on Medical Information Processing and Analysis*, Tandil, 2017.

[25] J. S. García-Salinas, L. Villaseñor-Pineda, C. A. Reyes-García and A. Torres-García, “Tensor Decomposition for Imagined Speech Discrimination in EEG,” in *Advances in Computational Intelligence. MICAI 2018*, Guadalajara, 2018.

[26] C. Cooney, A. Korik, F. Raffaella and D. Coyle, “Classification of imagined spoken word-pairs using convolutional neural networks,” in *Proceedings of the 8th Graz Brain Computer Interface Conference 2019*, Graz, 2019.

[27] C. Cooney, F. Raffaella and D. Coyle, “Optimizing Input Layers Improves CNN Generalization and Transfer Learning for Imagined Speech Decoding from EEG,” in *IEEE International Conference on Systems, Man, and Cybernetics, 2019*, Bari, 2019.

[28] P. Tan, W. Sa and L. Yu, “Applying Extreme Learning Machine to classification of EEG BCI,” in *2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, Chengdu, 2016.

[29] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard and T. Ball, “Deep learning with convolutional neural networks for EEG decoding and visualization,” *Human Brain Mapping,* vol. 38, no. 11, pp. 5391-5420, 2017.

[30] S. Ruder, “An overview of gradient descent optimization algorithms,” 15 9 2016.

[31] Y. Muhammad and D. Vaino, “Controlling Electronic Devices with Brain Rhythms/Electrical Activity Using Artificial Neural Network (ANN),” *Bioengineering,* vol. 6, no. 46, 2019.

Appendices

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Markus-Oliver Tamm,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

Vowel Classification from Imagined Speech Using Machine Learning,

(lõputöö pealkiri)

mille juhendaja on Yar Muhammad,

(juhendaja nimi)

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Markus-Oliver Tamm

03.05.2020