Determining the Appropriate Machine Learning Classifier in Electroencephalography Data for Brain Controlled Wheelchair

Shrey Jain

Supervisor: Dr. Ghassemi at Vector Institute of Technology and the University of Toronto

1.0 Introduction

Utilizing brain control interface platforms for individuals that are physically disabled has attracted a lot of attention due to its increased convenience, non-invasiveness nature, increased mobility, and comparatively low cost. Converting brain signals into control signals depends on the development of a brain computer interface (BCI). A variety of different inputs can be collected from the brain to work with this brain computer interface, often times taking the electroencephalography (EEG) data as the main source of input. This brain computer interface bridges the gap between the computer and the brain of humans. In order to employ human brain activity, converting brain signals into control signals is required and depends on the development of an interface between the brain and the computer. A brain-computer interface (BCI) is a system that bridges the gap between the computer and the brain of humans. This interface depends on the changes of rhythms in brain signal. These brain activities can be detected using electroencephalographic (EEG) signals. The BCI will transform the EEG signals produced by brain activity into control signals which can be used for controlling the wheelchair. The detected EEG signals are based on the change of frequencies and change of amplitudes. For example, during voluntary thoughts, the frequencies of signals change, and during movement, synchronisation/ desynchronisation of brain activity which involves mu rhythm amplitude change happens. This relevant characteristic makes rhythm based BCI suitable to be used. Recently, some research work has been done to develop many applications of BCI for wheelchairs. The main function of BCI is to convert and transmit human intentions into appropriate motion commands for the wheelchairs, robots, devices, and so forth. BCI allows improving the quality of life of disabled patients and letting them interact with their environment. The goal of this project is to outline the best classifier to be used on electroencephalography data, specifically for the purposes of controlling a wheelchair.

2.0 Table of Contents

3.0 Literature Review

These brain computer interfaces have the opportunity to impact a lot of people in the physically disabled community, but the robustness of the classification algorithm is a reason for the lack of implementation. Only in recent years has the fusion between machine learning classifiers and brain computer interfaces become a focus in the machine learning community. There are few research papers comparing the strength of different machine learning classifiers on EEG data. Additionally, the data sets that some of these research papers incorporate do not include a primary stakeholder to the accessibility community: children and individuals who face traumatic brain injury (TBI). There is also very little documentation to present whether or not there is an actual difference in the EEG signals between individuals who have suffered from TBI with respect to how it translates into motor control systems.

4.0 Problem/ Opportunity

4.1 Research Design and Methodology

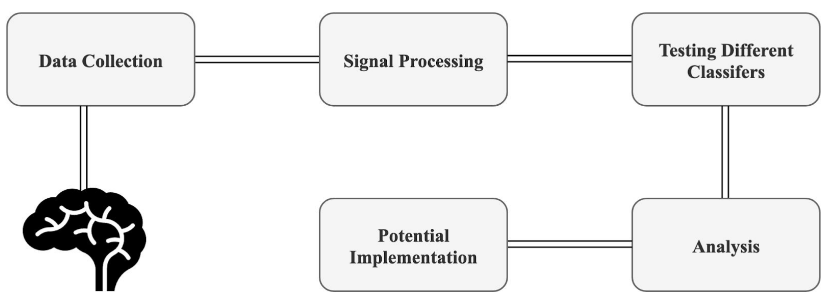

Image 1. *This image is a visual representation of the project pipeline that I anticipate to follow for this project.*

4.1.1 Data Collection

Some of the requirements I am looking for in a data-set is that the data-collection process used some of the company OpenBCI hardware as that is the hardware and framework that I plan to work with moving forward in this project.

1) **Use existing data-sets that are available online:**
   There are a variety of open source EEG data-sets available online. Examples of a few that I have considered using include:

a) The University of San Diego has created a public EEG dataset on a variety of different participants in many different scenarios outlined on their webpage: https://sccn.ucsd.edu/~arno/fam2data/publicly_available_EEG_data.html
b) An individual on GitHub has referenced a wide array of different publicly available EEG datasets to use.
https://github.com/meagmohit/EEG-Datasets
c) McGill NeuroTech Team's EEG dataset that they collected from anonymized and consented participants.
https://github.com/NTX-McGill/NeuroTechX-McGill-2019/tree/master/offline/data

2) Collect and build by my own data-set:

Building my own data-set is a stronger approach to the project as I believe it will yield the best results to confirm the mission of this study: which machine learning classifier is the best for EEG data, specifically pertaining towards individuals that face physical disabilities. Thus, building a dataset that includes that major stakeholders: children and individuals that have faced TBI. There is minimal literature supporting differences in EEG signals for individuals who have faced TBI and thus it would be worth exploring.

In order to collect my own data-set, it would require a lot more work, however, I believe that this would yield better results for the project as I the dataset I would be using would not be biased and include all involved stakeholders. From a logistics perspective, I have outlined below the process I would take in order to collect my own data-set:

- Gathering participants for the study. Referencing the ethical compliance section of this report, I would be abiding by all of the University of Toronto's research study guidelines. In order to get these participants I plan on reaching out to a variety of different hospitals in Toronto to work with individuals who have faced TBI as well as Sick Kids to work with children. If I am unable to work with either of these institutions, I plan to approach schools so I would at minimum be able to include children in this dataset.
- Equipment necessary for this dataset gathering is outlined further in the resources section of this proposal.

4.1.2 Signal Processing

One of the main issues with EEG data is the fact that it is very noisy, non-stationary and complex. Thus specific filters will need to be tested and used on the EEG data that has been collected. Some of the filters that I plan on exploring to use include, but are not limited to: Chebyshev Type I, Chebyshev Type II, Butterworth, and Elliptic.

4.1.3 Testing

In order to determine the best classification method for the EEG data, I will have to have some predefined metrics that I am going to be comparing them against. The metrics that I will base the success of the classifier on include: % of correctly classified instances, mean absolute error, and root mean squared error. The classifiers that I will explore include: SVM, linear regression, logistic regression, naive bayesian, random tree, random forest, linear discriminant analysis, K-nearest neighbors, and LSTM.

After determining which classifier is best on the dataset that is chosen to be used, I will then be working on prototyping a wheelchair (for this summer, a small scale 3-D printed/hand built) to test the findings in real life. Finally, I will be writing a paper on my findings and the process I decided to take when approaching this project and new findings that I come across during the process.

4.2 ML4H Vector Institute Benefits

In our last meeting, we discussed the possible work of a variety of different projects which I took the time to look into and explore. However, I believe that the opportunity that this project has, despite being a re-implementation to some degree, will lead to new ideas and new solutions to problems faced in this area of machine intelligence. There are very few projects at Vector currently being explored in the BCI space, and despite my limited work experience, I believe that this is a very important area of machine learning to continue exploring and there is no better supervisor for this project other than Dr. Ghassemi who leads the machine learning for health project group.

5.0 Practicalities

In this section, the logistics of the project proposal will be outlined. This includes the project lens and timeline, resource plan/budget, risk management as well as ethical compliance.

Timeline of Events:

TIME	**TASK**
February	**Apply for all necessary research grants**: NSERC, ESROP, UTEA, and engineering fellowship grant. I will also be reaching out, provided approval, to corporate sponsors. (*Resource plan is below)
March	**Clean and gather data:** Work on either gathering a new dataset (going to the hospital and working with patients who have had TBI, working with children, as well as able bodied individuals). The goal is to gather a more robust set of data. (*See data section for options of datasets that are currently available). **Reach Out:** Continue to reach out to sponsors for product support so I will have the necessary equipment for the summer.
April	**Continue to attend lectures and wrap up first year engineering science exams.**
May - July	**Start working full-time:** Test all of the different models outlined earlier in this document on the data-set either collected or gathered. Do thorough analysis on results to ensure that the selected classifier is the best and start to test findings on a prototype (3D-print a small wheelchair with arduino etc).
August	Write a paper on findings and in parallel potentially build a prototype of a working wheelchair.
Future	Depending on where I am in my research in August, I would like to expand my commitment to research in the health and machine learning space and whether that is a continuation on this project or others I am motivated to continue to build skills and make a valuable impact where I can.

Table 1. This table describes the timeline of events I anticipate to take place over the course of this year.

5.1 Location of Work/Placement of Resources:

As I do plan to have the equipment funded/sponsored, they will be stored in safe-keeping at Vector Institute or at the discretion of Dr.Ghassemis choice. I will be working at the Vector Institute of Technology occasionally throughout the school year and will be coming into the lab full time in the summer to work on this project. I will be starting this project part-time throughout the school year, but take it on full-time from May 1st-August 16th. A minimum of 16 weeks will be allocated to meet the requirements for the research grants, but out of personal interest, can see this project extending longer or expanding into new projects in the future.

5.2 Resource Plan/Budget:

I will be applying to a variety of different research scholarships. **All of the scholarships below require a description of the research project as well as confirmation of support of partial funding from the supervisor.** After the selected project is chosen, I will be submitting my application to each of the scholarships discussed below.

These grants include:

1) **NSERC**: NSERC provides $4500 and then the **supervisor is to provide $1500** to make the total scholarship $6000**.** The deadline for this scholarship is March 10th.
2) **UTEA:** UTEA provides $4875 and then the **supervisor is to provide $1125** to make the total scholarship $6000. The deadline for this scholarship is March 1st.
3) **ESROP:** ESROP provides $3000 and the **supervisor is required to provide equal or greater scholarship** to total the scholarship to a *minimum* of $6000. The deadline for this scholarship is March 6th.
4) **FY Fellowship (Engineering):** This fellowship provides $7000 and the **supervisor is at liberty to provide funding at their own discretion.** This Scholarship is due March 11th.

Through being the Vice-President of the Machine Intelligence team at UofT, directing Major League Hacking (MLH) Hackathon events and leading accessibility tech hackathons and workshops in the past, my background with respect to developing connections for sponsors in this space is strong. Prior to starting full time in the summer, I will be reaching out to a variety of organizations that would be of interest to sponsor this project **(with your approval).** Some of the organizations that I am interested in contacting, but are not nearly limited to, include: Invictus Games, ParaOlympic Committee, Deepmind, OpenBCI, Muse, IBM, Neuralink, Kernel, OpenAI, Halo Neuroscience, Mindstrong, and CTRL-Labs. I believe that through leveraging these partnerships for this project, I will be able to develop a better understanding of the potential applications that this project can have in industry as well as gather more data from potential stakeholders with this product.

Below I have completed an estimated cost of the equipment required to conduct this project. As I have not yet started the project and am still learning more daily about the opportunities this is simply a template and not a final say for what this projects execute cost will be. The goal is to have these expenses covered by the sponsors.

Expenditure Description	Expenditure Cost	Justification for Expenditure
OpenBCI Cyton Biosensing Board [Open BCI] (8-channels) *only required if I do my own data collection	$499 USD + taxes + shipping	In order to gather data for this project, I will need to utilize the OpenBCI board.
Medical grade Ten20 EEG conductive paste	$20 USD + taxes + shipping	Along with the ultracortex headset, the medical grade EEG conductive paste would be needed as well.
Ultracortex "Mark IV" EEG Headset [OpenBCI]	$800 USD + taxes + shipping	In order to gather the data, the headset is the most important part of the hardware.
Prototyped Small Wheelchair	$100 USD	This can be used to demonstrate/prototype a wheelchair being used live by an individual.

Table2. *This is a budget estimate of the project.*

5.3 Ethical Compliance:

In line with all of the ethics protocols outlined by the University of Toronto Delegated Ethics Review Committee (DERC) or Research Ethics Board (REB), there is very minimal risk associated with this project. Should I decided to move forward with in-person data collection, the probability and magnitude of harm to any participants that are participating in the research gathering process are not subject to: physiological or health issues (ie. clinical diagnoses or side effects), cognitive/emotional factors (ie. stress/anxiety) during data collection, and finally there are no socio-economic or legal ramifications such as stigma, loss of employment, deportation, or criminal investigation (eg, in the event of duty to report intent to cause serious harm, subpoena, or breach of confidentiality. By leveraging the University of Toronto risk matrix [University of Toronto], this project has minimal risk. Additionally, should I move forward with the data collection process, I will be sure to submit the "undergraduate ethics review protocol form", prior to moving forward as well as taking any other precautions outlined by advisors.

6.0 References

Abiyev, Rahib & Akkaya, Nurullah & Aytac, Ersin & Günsel, Irfan & Çağman, Ahmet. (2016). Brain-Computer Interface for Control of Wheelchair Using Fuzzy Neural Networks. BioMed Research International. 2016. 1-9. 10.1155/2016/9359868.

Decoding Upper Limb Movement Attempt From EEG Measurements of the Contralesional Motor Cortex in Chronic Stroke Patients

Ianof, J. N., & Anghinah, R. (2017). Traumatic brain injury: An EEG point of view. Retrieved January 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5619208/

J. M. Antelis, L. Montesano, A. Ramos-Murguialday, N. Birbaumer and J. Minguez, "Decoding Upper Limb Movement Attempt From EEG Measurements of the Contralesional Motor Cortex in Chronic Stroke Patients," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 1, pp. 99-111, Jan. 2017.

Kim, J., Lee, B., Lee, H. S., Shin, K. H., Kim, M. J., & Son, E. (2014). Differences in Brain Waves of Normal Persons and Stroke Patients during Action Observation and Motor Imagery. Journal of physical therapy science, 26(2), 215–218. https://doi.org/10.1589/jpts.26.215

Louis, E. K. S. (1970, January 1). The Normal EEG. Retrieved January 2020, from https://www.ncbi.nlm.nih.gov/books/NBK390343/

Open BCI Featured Products. (2020, January). Retrieved January 2020, from https://shop.openbci.com/collections/frontpage

Palo, N. D. (2018, August 5). From brain waves to robot movements with deep learning: an introduction. Retrieved January 2020, from https://towardsdatascience.com/from-brain-waves-to-arm-movements-with-deep-learning-an-introduction-3c2a8b535ece

Rapp PE, Keyser DO, Albano A, Hernandez R, Gibson DB, Zambon RA, Hairston WD, Hughes JD, Krystal A and Nichols AS (2015) Traumatic brain injury detection using electrophysiological methods. Front. Hum. Neurosci. 9:11. doi: 10.3389/fnhum.2015.00011

Siswoyo, Agus & Arief, Zainal & Sulistijono, Indra Adji. (2017). Application of Artificial Neural Networks in Modeling Direction Wheelchairs Using Neurosky Mindset Mobile (EEG) Device. EMITTER International Journal of Engineering Technology. 5. 10.24003/emitter.v5i1.165.

Sun, Shiliang & Zhang, Changshui & Zhang, Dan. (2007). An experimental evaluation of ensemble methods for EEG signal classification. Pattern Recognition Letters. 28. 2157-2163. 10.1016/j.patrec.2007.06.018.

University of Toronto. (n.d.). Determine If a Protocol Qualifies for Delegated Review. Retrieved February 24, 2020, from https://research.utoronto.ca/ethics-human-research/determine-if-protocol-qualifies-delegated-review

Velasco-Álvarez F., Ron-Angevin R. (2009) Asynchronous Brain-Computer Interface to Navigate in Virtual Environments Using One Motor Imagery. In: Cabestany J., Sandoval F., Prieto A., Corchado J.M. (eds) Bio-Inspired Systems: Computational and Ambient Intelligence. IWANN 2009. Lecture Notes in Computer Science, vol 5517. Springer, Berlin, Heidelberg