JATT: Large Vessel Occlusion Prediction

Chu Duc Thang
Department of Computing Science
University of Alberta
116 St & 85 Ave, Edmonton, AB T6G 2R3
thang@ualberta.ca

Akrash Sharma
Department of Computing Science
University of Alberta
116 St & 85 Ave, Edmonton, AB T6G 2R3
akrash@ualberta.ca

Tanyaradzwa Gerald Gozhora
Department of Computing Science
University of Alberta
116 St & 85 Ave, Edmonton, AB T6G 2R3
gozhora@ualberta.ca

James Le
Department of Computing Science
University of Alberta
116 St & 85 Ave, Edmonton, AB T6G 2R3
quanganh@ualberta.ca

Abstract

Large vessel occlusions (LVO) are ischemic strokes resulting from a blockage in one of the brain's major arteries. Since LVO develops quickly and causes irreversible damage to patients' brains, doctors need a quick but accurate system to predict whether the patients have this disease. Currently, doctors use the Los Angeles Motor Scale (LAMS) score to predict and assess stroke severity. However, these predictions are sometimes inaccurate. Therefore, we use machine learning algorithms to build a classifier that can predict whether the patients have LVO or not. Specifically, we propose an approach to use both the Electroencephalography (EEG) and clinical data, then try various techniques, such as Support Vector Machine (SVM) and EEGNet. To evaluate the model, we use a custom evaluation metric that penalizes heavily when the model makes false-negative predictions. Our experiments shows that EEGNet is currently the best models for this task. Furthermore, our experiments indicate that using both EEG data and clinical data performs better than using only clinical data for the task in the case of SVM and EEGNet.

1 Introduction

A stroke is a medical condition in which poor blood flow to the brain causes cell death (21). Stroke is the leading cause of adult disability in Canada and the third leading cause of death. Every year, nearly 14,000 Canadians die from stroke, and about 300,000 Canadians are living with the effects of stroke (16). There are two types of strokes: ischemic, due to lack of blood flow, and hemorrhagic, due to bleeding (17). Both cause parts of the brain to stop functioning. LVO is a type of ischemic stroke in which there is a blockage in a vessel that supplies blood to the brain. When a person has a stroke, the paramedics need to determine whether this particular patient has LVO or not to decide whether to transfer to an appropriate medical facility.

EEG is a method to record an electrogram of the electrical activity on the scalp. Specifically, the medical practitioners will paste electrodes consisting of small metal discs with thin wires onto the scalp. The electrodes then detect tiny electrical charges resulting from the brain cells' activity. The charges are amplified and appear as a graph on a computer screen, which the healthcare provider then interprets (15). Since the paramedics can quickly obtain EEG recording in an ambulance, they are exploring the use of EEG as a diagnostic and prognostic tool in stroke (9).

Currently, the paramedics use clinical data (age, gender) and LAMS scores to estimate whether a patient has LVO or not. Specifically, the LAMS score is one of several validated prehospital stroke scales. The paramedics will assign the score from one to five based on the patients' current conditions. A higher LAMS score could indicate a more likely patient to have LVO (18). However, limitations of LAMS score can make it inaccurate(31). Therefore, we propose to use Machine Learning (ML) and Deep Learning (DL) to address this issue. Specifically, since EEG is easy to obtain, we suggest combining EEG data with clinical data and LAMS score as input to the model. Therefore, our task is to use ML and DL to build a classifier that can predict whether a patient has LVO or not by using EEG, clinical data, and LAMS score. To evaluate the model, we will compare the model that uses only clinical data and LAMS score with the model that uses all three data types. Our goal is to have the second model perform better than the first one. Besides building a classifier, our paper also suggests multiple ways to reduce the EEG data noise.

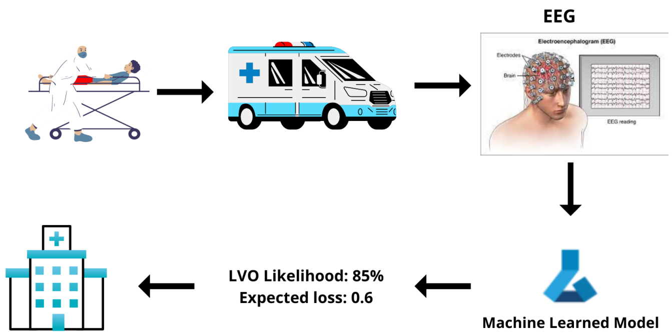

Figure 1: The potential application of the model

Figure 1 illustrates an example of how to use our model. Suppose a patient has a stroke, and someone calls an ambulance. The paramedics will have to determine whether a patient has LVO or not. The paramedics will hook the patient up with a headset with EEG sensors to record the EEG data. Then, the paramedics perform a LAMS assessment and gather some other basic clinical data, namely age, and gender. After that, the paramedics will input EEG, LAMS, age, and gender into our classifier to output the probability of a patient having LVO. The ambulance then carries the patient to an appropriate medical facility that can treat LVO patients. Making an accurate decision for these situations can help save many people's lives and reduce the load on the medical facilities.

In the remaining of this paper, section 2 will go over the related works and previous related research. Next, section 3 will go over different methods implemented in the research. Then, section 4 will dicuss the evaluation metrics, and summarize the result of different experiments. Lastly, section 5 will conclude our paper with some afterthoughts.

2 Related Work

Previous research studies have focused on predicting large acute ischemic stroke outcomes (9). Additionally, they only used either EEG data (4) or clinical data (5). Since the task of predicting strokes and predicting LVO are similar, we will learn from previous research and adapt to our current task. Moreover, we will combine EEG data, clinical data and LAMS score to use in our model.

Since EEG data is noisy (19), we researched different methods to preprocess the EEG data. Previous papers (8; 9) use Independent Component Analysis (ICA) to decompose EEG signals into independent components for artifact removal. Artifacts are unwanted components that arise from other sources and mislead the actual cerebral activity of present recorded EEG data. Furthermore, a previous research paper (8) suggests analyzing the EEG data using statistical features such as Mean Absolute Value, Standard Deviation, and Variance. However, comparing statistical features among different patients can lead to misinterpretation since each patient has an unique condition. Therefore, Ravikiran Mane (6) proposes to apply a discrete Fast Fourier Transform (FFT) feature extraction to obtain a delta band (measured in the 1–3 Hz) and a low beta band (measured in the 13–19 Hz.) after performing ICA. Re-examining the results using a Laplacian transform, they noticed that no EEG measure correlated with infarct volume or with NIHSS score ($p>0.1$). The research paper (10) provides a reason to explain previous behavior that there is information loss in the time domain and it gives only spectral information in the frequency domain. The paper (10) shows that wavelet transform is more effective than FFT since wavelet is a type of time-frequency analysis, which provides information about both frequency and time within signals. Therefore, we tried to apply Wavelet transform for feature extraction on EEG data.

Besides preprocessing the data, we also research on the algorithms to use in our project. Recent approaches (11) have used SVM, RF, and stacking with logistic regression to detect large artery atherosclerosis using clinical data. Since LVO is an ischemic stroke and ischemic stroke is caused by atherosclerosis, these three algorithms can be applied in our task. Another study (7) used Gaussian Process (GP) to predict LVO using clinical data. Specifically, the paper suggested that the model using GP algorithm (0.874 ± 0.025) outperforms the model using NIHSS score, which is another prehospital stroke scale, (0.819 ± 0.024) in terms of AUC score. Since this task is similar to our task, we will try to implement a GP algorithm as one of our models.

For DL models, recent research studies (12; 13) proposes EEGNET and DeepConvNet in the task of Brain-Computer Interfaces (BCI) for classifying the stroke patients motor imagery EEG data. While classification performance for motor imagery tasks BCI of DeepConvNet and EEGNet were similar across all cross-subject analyses, DeepConvNet performance was lower across nearly all within-subject analyses. In a within-subject design, researchers compare related measures from the same participants between different conditions whereas in a cross-subject analysis the source domain contains the EEG data of the subjects, which are not seen in target domain and source domain and target domain contain completely different subjects. This result is due to the difference in the amount of training data used to train the model. Specifically, the size of training data is about 10-15 times larger than that of within-subject analyses. This led them to conclude that DeepConvNet is more data-intensive compared to EEGNet. Since our dataset is limited (115 patients), we choose to use EEGNet on the EEG data to predict the likelihood of LVO.

3 Methods

Our goal is to build a classifier to predict whether patients have LVO or not. To achieve this, we first consider standard machine learning techniques. Previous work suggested that Support Vector Machine (SVM), Random Forest (RF), and Gaussian Processes (GP) are reasonable methods to consider for our task (5; 6; 7; 10). We also modify some model details for each algorithm to suit our task.

3.1 Preprocessing

3.1.1 Clinical

Our dataset contains 115 patients. Each patient has 51 features: age, gender, hemorrhage location, stroke onset time, and EEG recording time. Since some features require time, such as hemorrhage location, we only selected features that are available onsite, such as age, gender, and LAMS score. We expressed gender as a binary digit and age as an integer. LAMS is an integer between zero and five assigned by emergency health practitioners to estimate the severity of a stroke. Moreover, we will use stroke onset time and EEG recording time from the original data to calculate a new feature called time elapsed. Thus, our clinical data will include four features, namely age, gender, LAMS score, and time elapsed.

3.1.2 EEG

Our EEG data consisted of three-minute recordings from a Muse two headset. First, we read the data channels from the respective positions, namely TP9 (left ear), TP10 (right ear), AF7 (left forehead), and AF8 (right forehead) see fig 2. Since our device had a sampling rate of 256 Hz, we had approximately 46,000-time points for each channel, see Fig 13. Moreover, we had electrocardiogram (ECG) signals from a Right Auxiliary Channel. Since we have two approaches to building a classifier, namely machine learning and deep learning, we introduced two ways to preprocess EEG data for ML and one way for DL, with two variations to how we do it for ML.

EEG Preprocessing for standard ML models

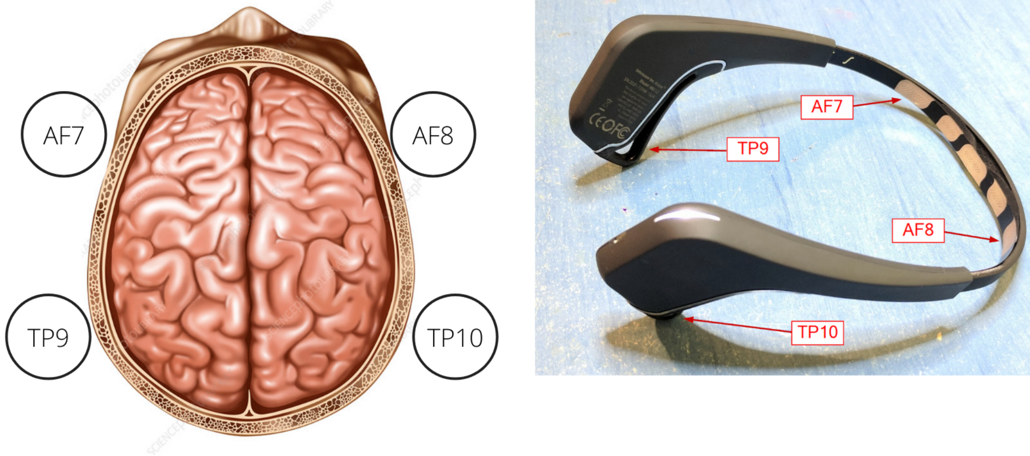

Figure 2: EEG Data being read from headset

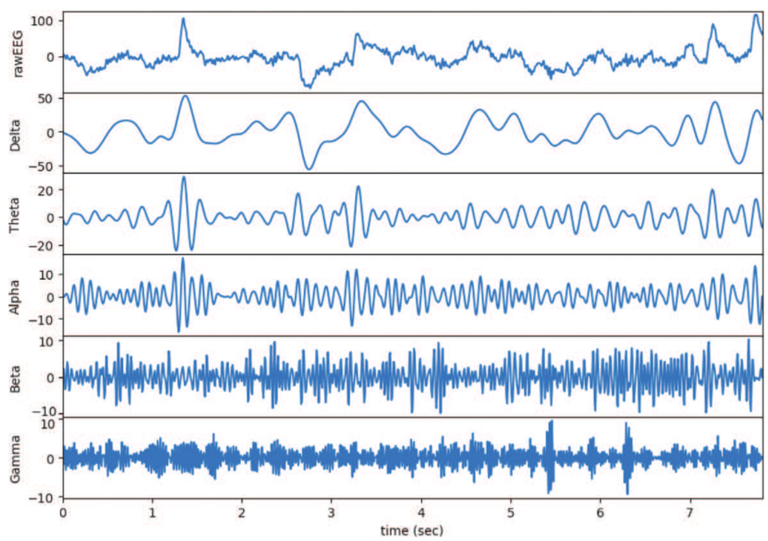

Figure 3: Division of EEG signal into the 5 signals

Since our EEG data is in high dimension, we perform dimensionality reduction and feature extraction for our standard ML models such as RF, GP, and SVM. We try two different methods: wavelet transformation, Higuchi Fractal Dimension, and Detrended Fluctuation Analysis (HFD DFA). Since wavelet transformation is a standard method (30), we perform the method on EEG data to find waves that make up a given signal see fig 3 and calculate statistics from those waves. First, since the EEG data of one patient contains four channels, we performed a wavelet transformation into five waves for each channel. We choose five waves for each channel as EEG signals are made of five waves that have different frequency bands, namely gamma, alpha, beta, theta, and delta, note fig 3. Therefore, our data can be more interpretable and valuable. After obtaining twenty (five times four) total waves, we calculate each wave's Shannon Entropy and Wave Energy, which give us real scalars. For each patient, we reduce the original dimension of the EEG data from 200000 values (50000 times four) to forty values (two times twenty). We repeat the whole process for all 115 patients in the dataset. See fig 4. Finally, we normalize the data using the Z-score. Whereby we replace each value x by $z = \frac{x-\mu}{\sigma}$., where σ is the standard deviation of that particular value, the given Shannon Entropy or Wave Energy value, across participants and μ the mean for the value.

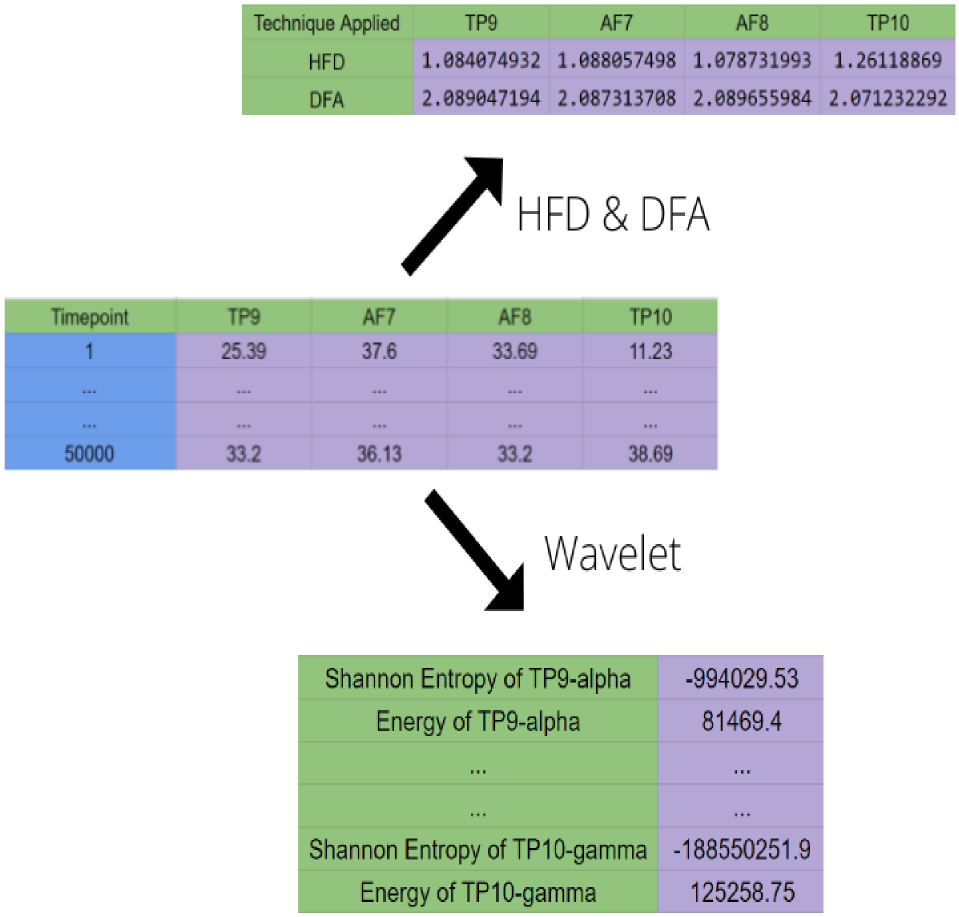

Figure 4: HFD and DFA versus Wavelet transformation

Additionally, we try HFD and DFA since they are promising techniques for emotional detection and analysis of Alzheimer's (24). HFD is "an approximate value for the box-counting dimension of a real-valued function or time series" (22), and DFA is a "method for determining the statistical self affinity of a signal" (23). First, we pass the raw EEG data from the headset through a low and high pass filter from 0.5 to 30 Hertz (Hz). Then, we calculate the HFD and DFA values for each channel; again, these are real scalars, so we have two values. Specifically, since each patient has four channels in EEG data, the method will produce eight values per participant (two times four). Therefore, we reduce the dimension of the original data from 20000 (5000 times four) values per participant to eight values (two times four) per participant. We repeat the same process for 115 patients in the dataset. We also normalize the current data using the z-score technique, just as we did for wavelet transformation. Although our task is different from the previous applications, we want to attempt these techniques for preprocessing EEG data to understand which method works best in our case. Fig 4 shows the mechanics of the two pipelines, and see fig 4 shows how HFD and DFA work with specific data.

EEG Preprocessing for DL and Neural Network Models

Since deep-learning models can learn a complex representation of the data in a high dimensional space (29), we will not perform the same procedure as above. Instead, we perform a simpler preprocess on the EEG data to reduce the noise. First, we passed the raw EEG data through a low and high pass filter of 0.5 and 30 Hz Fig 9. Then, since we had access to the right Auxiliary channel with an ECG signal, we opted for ICA rejection using ECG (14) to remove components of the signal that contains noise. Finally, we projected back to the space and returned the data to our models.

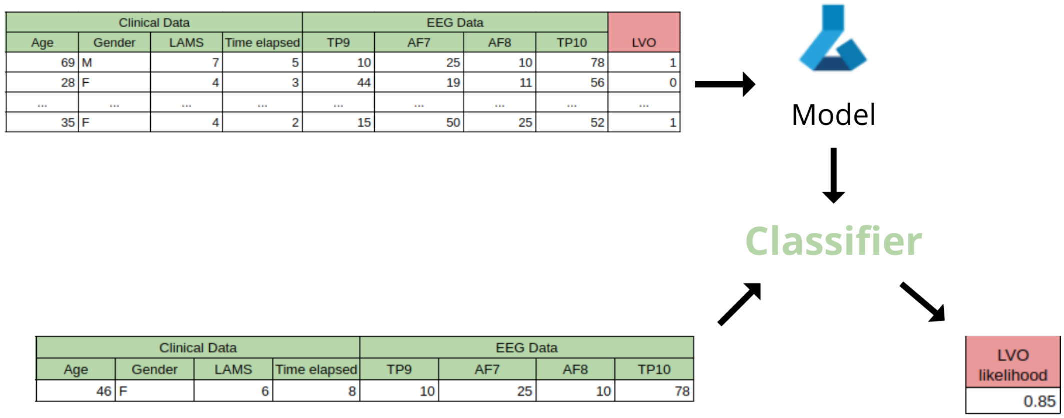

Figure 5: Development of our Machine Learning and Deep Learning models

3.2 Support Vector Machine

SVM is a supervised machine learning algorithm used in binary classification problems (25). Given clinical and EEG data, we use SVM to build a classifier to predict whether the patient has LVO or not. In our task, we apply SVM in two different ways. Our first approach was to train two SVM classifiers, one on EEG data and one on just clinical data. Each classifier outputs a number between 0 and 1. Then, we used a soft voting classifier to output the probability that the patient has an LVO. Soft voting classifier averages the probabilities of the predictions made by the other classifiers. Similar to the previous research study (11), we tried stacking with Logistic Regression. A stacking classifier takes the predictions of two base classifiers (SVM in our case) and uses those as new features to train the meta-classifier, namely Logistic regression. Our second approach was to use SVM on the combination of EEG and clinical data to build a single classifier. We used Grid Search Cross-validation to select the hyper-parameters.

3.3 Gaussian Processes

GP is a supervised, non-parametric machine learning algorithm. It is based on Bayesian methodology that assumes some prior distribution on the underlying probability densities (27). To apply this learning algorithm, we used the preprocessed EEG data and clinical data as our training data to build a classifier. Similar to SVM, we apply Grid Search Cross-validation to find the best parameters for the model. Given a new data point, we will output the probability that the patient has LVO or not.

3.4 Random Forest

Different from the previous two methods, RF is an ensemble learning method that constructs many decision trees at the training time. Then, it outputs the predicted class selected by most trees (28). In our task, we used this learning algorithm with the combination of EEG and clinical data to build a classifier. Then, we use the trained classifier to predict the probability of a new patient having LVO or not. As the previous approaches, we used Grid Search Cross-validation and Randomized Search Cross-validation to optimize the model's parameters.

3.5 EEGNet

EEGNet is currently the best model using EEG data for the Brain-Computer Interfaces task (6). Since our task uses EEG data, we decided to use this model with slight modifications to the architecture. EEGNet is essentially a convolution neural network, so we do not need to perform feature extraction on the data before using it to build a classifier. Instead, we only standardize the data before use. As the convolution layers require the input to have the same dimension, we only select the first 5000 time points of the EEG data. Besides using the EEGNet for the EEG data, we also have a separate neural network for the clinical data. We train these two models simultaneously and output the four-by-one vector for each network before concatenating it into one eight-by-one feature vector. We choose the size of output to be four-by-one after trying different numbers for the dimension. After that, we have another neural network that inputs the previous feature vector and outputs the probability of having LVO or not. After building the classifier, we take a new instance of data to the classifier to output the probability that the patient has LVO or not.

3.6 LSTM

LSTM is a Recurrent Neural Network (RNN) technique specialized in dealing with time series problems (26). Since EEG is a time-series data, we consider LSTM a good choice in our case. Unlike other feedforward neural networks, LSTM can process not only a single data point but also an entire sequence of data. Since there can be unknown duration delays between important EEG data events, LSTM can capture these events with its input gate, output gate, and forget gate. Furthermore, unlike EEGNet, the LSTM model does not constrain the number of time points we input. Therefore, the model can have more information on EEG data to train. While we use LSTM to learn about the EEG data, we also have a separate neural network for the clinical data. Since we have more information on EEG data, we decided to output a sixteen-by-one vector from the LSTM network while remaining a four-by-one vector from the other network. Then, we concatenate these two vectors into a twenty-by-one vector and feed it to another neural network to predict the probability of having LVO or not. Finally, we test new data through the learned classifier to output the probability of having LVO.

4 Experiment

4.1 Expected loss

$$\text{Expected loss} = \frac{4FN + FP}{4(TP + FN) + FP + TN}$$

After training the classifier, we will input the test dataset to evaluate our classifier. For each data, the classifier will output the probability of having LVO. We decided to choose the common threshold of 0.5 so that if the probability is greater than 0.5, the classifier predicts that the patient has LVO and vice versa. Then, we compare the predicted label of the test data with their true label to obtain the confusion matrix. From the matrix, we are interested in False Negative (FN), which occurs when the true label of the data is positive (have LVO), but the predicted label of the data is negative (not have LVO). Since we consider the cost of having FN is higher than that of having FP, we penalize the model more when it makes FN. After discussing with the domain expert, we decided to evaluate our model performance by calculating the above equation using values from the confusion matrix. We notice that the smaller the expected loss, the better the model's performance is.

4.2 Results

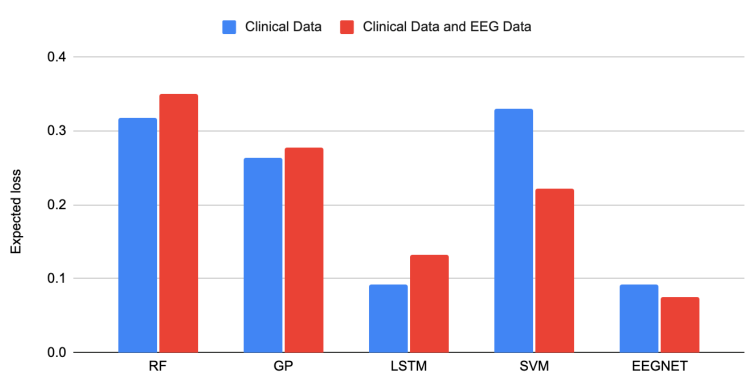

Figure 6: Expected Loss on Clinical Data versus Clinical and EEG Data on the models

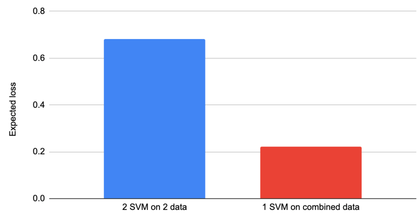

Figure 7: Performance of stacking two SVM models using logistic regression versus SVM on combined EEG and Clinical Data

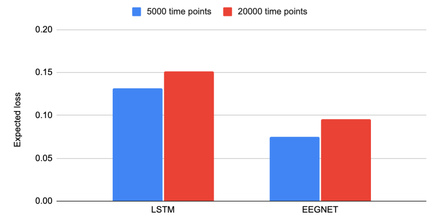

Figure 8: Performance of LSTM, EEG-Net with 5000 time points versus 20000 time points

From Figure 6 we can see that while EEGNet achieves the best performance (0.0754), RF shows the worst performance in our custom evaluation metric (0.3498). One possible explanation is that deep learning can learn a more complex representation of the data in the high dimensional space. Since SVM works well with high dimensional data, it has the best performance (0.2212) among standard machine learning techniques. However, although LSTM is theoretically good with time-series data, our experiment shows that it is not always the case. Specifically, our experiment indicates that the performance of the LSTM model is worse than EEGNet. It can happen when there is noise in our data and the current dataset is small.

Moreover, we wanted to understand how the model's performance will change if we combine the EEG and the clinical data. While RF, GP, and LSTM show that incorporating EEG data degrades the models' performance, the opposite trend occurs for SVM and EEGNet. It can happen as RF and GP models are suitable for low dimensional data, and the LSTM model is sensitive to perturbations.

During experiment with SVM, we also explored the performance difference between having two distinct SVM classifiers with a soft voting classifier on separated data and having one SVM classifier on the combined data. Initially, we anticipate that the first solution will perform better than the second solution since two SVM classifiers can learn to specialize on separate data to produce a whole optimal classifier. However, from Figure 7 we can see that our second solution (0.2212) outperformed the first one (0.6805) in the experiment. This result can occur when these two classifiers do not complement each other.

Lastly, we tried to select the number of time points in EEG data. Furthermore, our experiment shows that increasing time points of EEG data decreases the performance of LSTM and EEGNet. Since the data's dimension increases, the model needs more parameters to learn efficiently. However, it can lead to an overfitting problem as our training data is limited.

5 Conclusion and Discussion

We will consider using Accelerometer (ACC), which measures the non-gravitational acceleration, and Gyroscope datasets, which determine the orientation. These data can provide more information for models to learn and increase their performance. Lastly, we are also interested in how the classifier makes predictions. Therefore, we plan to apply Grad-Cam and SHAP methods on EEGNet and other methods to see how each layer contributes to the prediction.

Overall, our project is to build a classifier to predict whether each patient has LVO or not. Our study suggests that EEGNet is currently the best model among all our tried techniques. This result is consistent with the fact that deep learning techniques can capture the complex representation of the data without performing feature extraction (29). We noticed that SVM has higher performance than RF and GP among standard machine learning techniques since SVM works well with high dimensional data. We noted that different machine learning techniques captured different aspects of the data during the research. For instance, wavelet transform works well with RF, while HFD DFA with SVM and GP achieves a lower expected loss. Combining two SVM models did not perform better than one SVM model as stacking classifiers can give poor performance if those classifiers are not complementing each other. Finally, stacking classifiers works better with a more significant number of input models than with a smaller number of ensembled models.

References

[1] Alexander, J.A. & Mozer, M.C. (1995). Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995). *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.

[4] E. P. Giri & M. I. Fanany & A. M. Arymurthy & S. K. Wijaya. (2016). Ischemic stroke identification based on EEG and EOG using ID convolutional neural network and batch normalization *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 484-491

[5] Hidehisa Nishi & Naoya Oishi & Akira Ishii & Isao Ono & Takenori Ogura & Tadashi Sunohara & Hideo Chihara & Ryu Fukumitsu & Masakazu Okawa & Norikazu Yamana & Hirotoshi Imamura & Nobutake Sadamasa & Taketo Hatano & Ichiro Nakahara & Nobuyuki Sakai & Susumu Miyamoto. (2019). Predicting Clinical Outcomes of Large Vessel Occlusion Before Mechanical Thrombectomy Using Machine Learning *Stroke. 2019;50:2379–2388*

[6] Mane Ravikiran & Chew Effie & Chua Karen & Ang Kai Keng & Robinson Neethu & Vinod A. P. & Lee Seong-Whan & Guan Cuntai. (2021). FBCNet: A Multi-view Convolutional Neural Network for Brain-Computer Interface *arXiv:2104.01233*

[7] Huo D & Leppert M & Pollard R et al.Huo. (2021). Large Vessel Occlusion Prediction in the Emergency Department with National Institutes of Health Stroke Scale Components: A Machine Learning Approach *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association vol. 30,10 (2021): 106030.*

[8] R. D. Kusumastuti & A. D. Wibawa & M. H. Purnomo. (2021). Stroke Severity Classification based on EEG Statistical Features *1st International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)* pp. 138-142

[9] Shreve L & Kaur A & Vo C & Wu J & Cassidy JM & Nguyen A & Zhou RJ & Tran TB & Yang DZ & Medizade AI & Chakravarthy B & Hoonpongsimanont W & Barton E & Yu W & Srinivasan R & Cramer SC. (2019). Electroencephalography Measures are Useful for Identifying Large Acute Ischemic Stroke in the Emergency Department *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association vol. 28,8 (2019): 2280-2286.*

[10] Kumar N & Alam K & Siddiqi A. H. (2017). Wavelet Transform for Classification of EEG Signal using SVM and ANN *Biomed Pharmacol J 2017;10(4).*

[11] Garg R & Oh E & Naidech A & Kording K & Prabhakaran S. (2019). Automating Ischemic Stroke Subtype Classification Using Machine Learning and Natural Language Processing. *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association vol. 28,7 (2019): 2045-2051.*

[12] Raza Haider & Chowdhury Anirban & Bhattacharyya Saugat. (2020). Deep Learning based Prediction of EEG Motor Imagery of Stroke Patients' for Neuro-Rehabilitation Application. *10.1109/IJCNN48605.2020.9206884.*

[13] Vernon J Lawhern & Amelia J Solon & Nicholas R Waytowich & Stephen M Gordon & Chou P Hung & Brent J Lance. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces *Journal of Neural Engineering* `https://doi.org/10.1088%2F1741-2552%2Faace8c`

[14] Alexandre Gramfort & Martin Luessi & Eric Larson & Denis A. Engemann & Daniel Strohmeier & Christian Brodbeck, Roman Goj & Mainak Jas & Teon Brooks & Lauri Parkkonen & Matti S. Hämäläinen.(2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience, 7(267):1–13*

[15] Johns Hopkins Medicine Electroencephalogram (EEG) https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg%20-

[16] CorHealth Ontario, Fact Sheet: Stroke Statistics http://www.ontariostrokenetwork.ca/pdf/Final_Fact_Sheet_Stroke_Stats_3.pdf

[17] Centers for Disease Control and Prevention, About Stroke. https://www.cdc.gov/stroke/about.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fstroke%2Ftypes_of_stroke.htm

[18] Alberta Health Services, Lams Reference Tool. https://www.albertahealthservices.ca/assets/about/scn/ahs-scn-cvs-era-lams-reference.pdf

[19] Repovš G. Dealing with Noise in EEG Recording and Data Analysis. https://ims.mf.uni-lj.si/ims_archive/15-01/15b-21.pdf

[20] Shailaja Kotte & J R K Kumar Dabbakuti. (2020). Methods for removal of artifacts from EEG signal: A review *J. Phys.: Conf. Ser.1706 012093*

[21] Mayo Clinic, Stroke. https://www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113#:~:text=An%20ischemic%20stroke%20occurs%20when,brain%20damage%20and%20other%20complications.

[22] Harne, Bhavna P. Higuchi fractal dimension analysis of EEG signal before and after OM chanting to observe overall effect on brain *International Journal of Electrical and Computer Engineering 4.4 (2014): 585*

[23] Márton, L. F., et al. Detrended Fluctuation Analysis of EEG Signals *Procedia Technology 12 (2014): 125-132*

[24] Atypical temporal-scale-specific fractal changes in Alzheimer's disease EEG and their relevance to cognitive decline National Library of Medicine *Cognitive neurodynamics vol. 13,1*

[25] Scikit-Learn, Support Vector Machines. https://scikit-learn.org/stable/modules/svm.html#:~:text=Support%20vector%20machines%20(SVMs)%20are,than%20the%20number%20of%20samples.

[26] Machine Learning Mastery, A Gentle Introduction to Long Short-Term Memory Networks by the Experts. https://machinelearningmastery.com/promise-recurrent-neural-networks-time-series-forecasting/

[27] Scikit-Learn, Gaussian Processes. https://scikit-learn.org/stable/modules/gaussian_process.html

[28] Leo Breiman & Adele Cutler. Random Forests. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

[29] Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN COMPUT. SCI. 2, 420 (2021). https://doi.org/10.1007/s42979-021-00815-1

[30] Oliver Faust & U. Rajendra Acharya & Hojjat Adeli & Amir Adeli. Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis (2015). https://doi.org/10.1016/j.seizure.2015.01.012

[31] Panichpisal, Kessarin et al Validation of Stroke Network of Wisconsin Scale at Aurora Health Care System. *Journal of vascular and interventional neurology vol. 10,2 (2018): 69-73.*

A Data Pre-processing Procedure

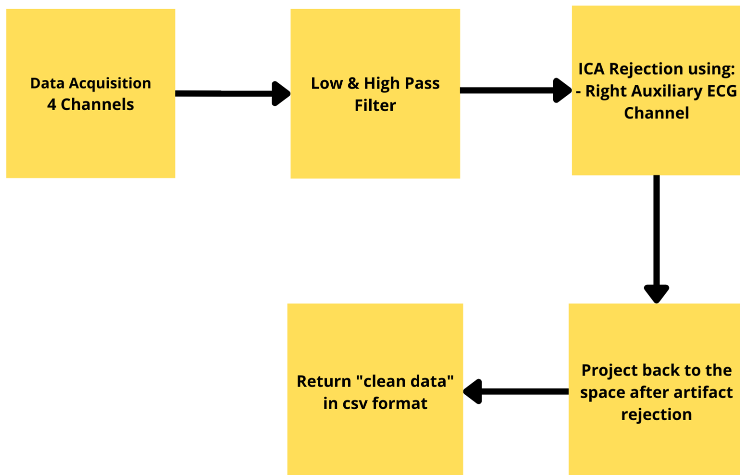

Figure 9: Preprocessing Procedure for Neural Network and Deep Learning based models

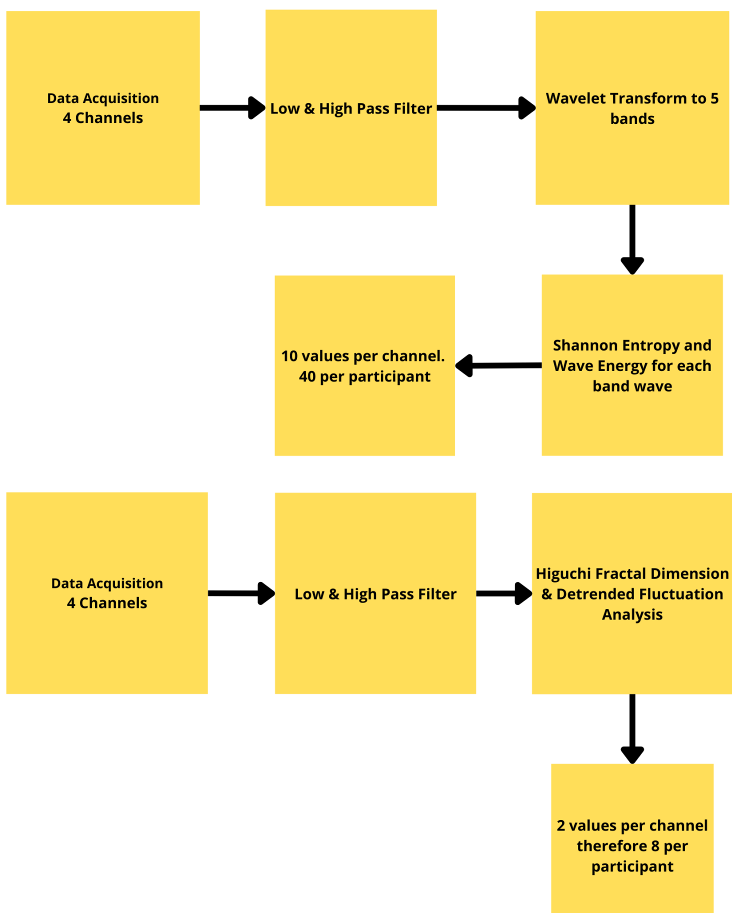

Figure 10: Preprocessing Procedure for Standard Machine Learning models

B Tables

	Clinical	Clinical & EEG
SVM	0.3305	**0.2212**
RF	**0.3182**	0.3498
GP	**0.2639**	0.2781
EEGNET	0.0928	**0.0754**
LSTM	**0.0928**	0.1316

Table 1: Performance of 1. SVM, 2. Random Forest (RF), 3. Gaussian Process (GP), 4. Neural Network with EEGNet (, 5. Neural Network with LSTM in terms of Evaluation metric

	Seperated EEG and Clinical data	Combined EEG and Clinical data
SVM	0.6805	**0.2212**

Table 2: The performance of the combinations of SVM versus single SVM on combined data in terms of evaluation metric

	Linear	RBF	Quadratic
SVM	**0.2212**	0.6669	0.6805
GP	**0.2781**	0.6517	0.6714

Table 3: The performance of 1. GP (Gaussian Process) 2. SVM (Support Vector Machine) with Linear, RBF, Quadratic kernels in terms of evaluation metric

	Seperated EEG and Clinical data	Combined EEG and Clinical data
SVM	0.6805	**0.2212**

Table 4: The performance of the combinations of SVM versus single SVM on combined data in terms of evaluation metric

	5000 time points	20000 time points
EEGNET	**0.0754**	0.0954
LSTM	**0.1316**	0.1516

Table 5: The performance of EEGNet and LSTM with 5000 time points vs 20000 time points

C Extra Figures

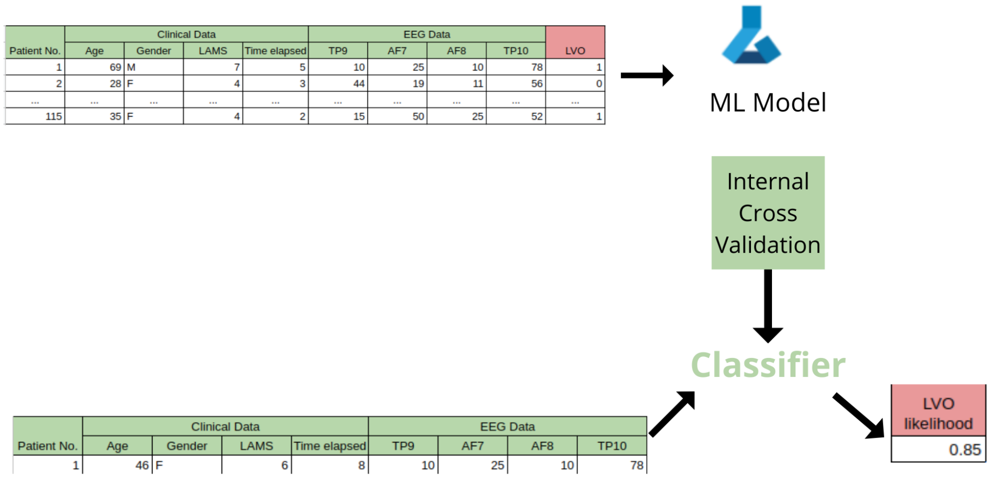

Figure 11: Development of our ML models

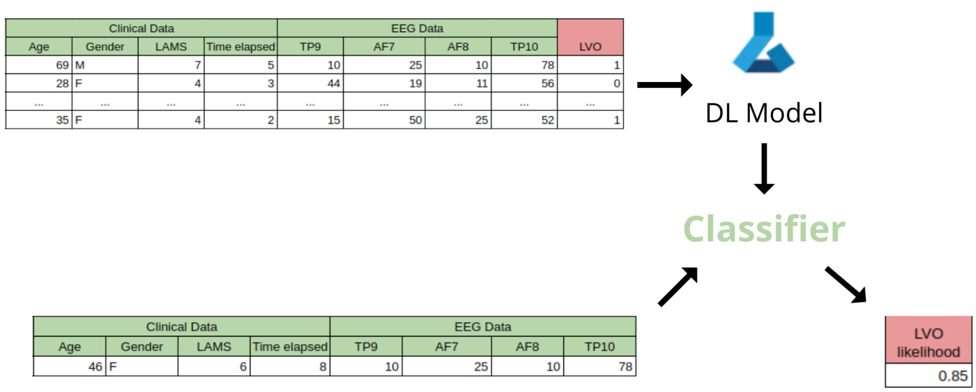

Figure 12: Development of our DL models

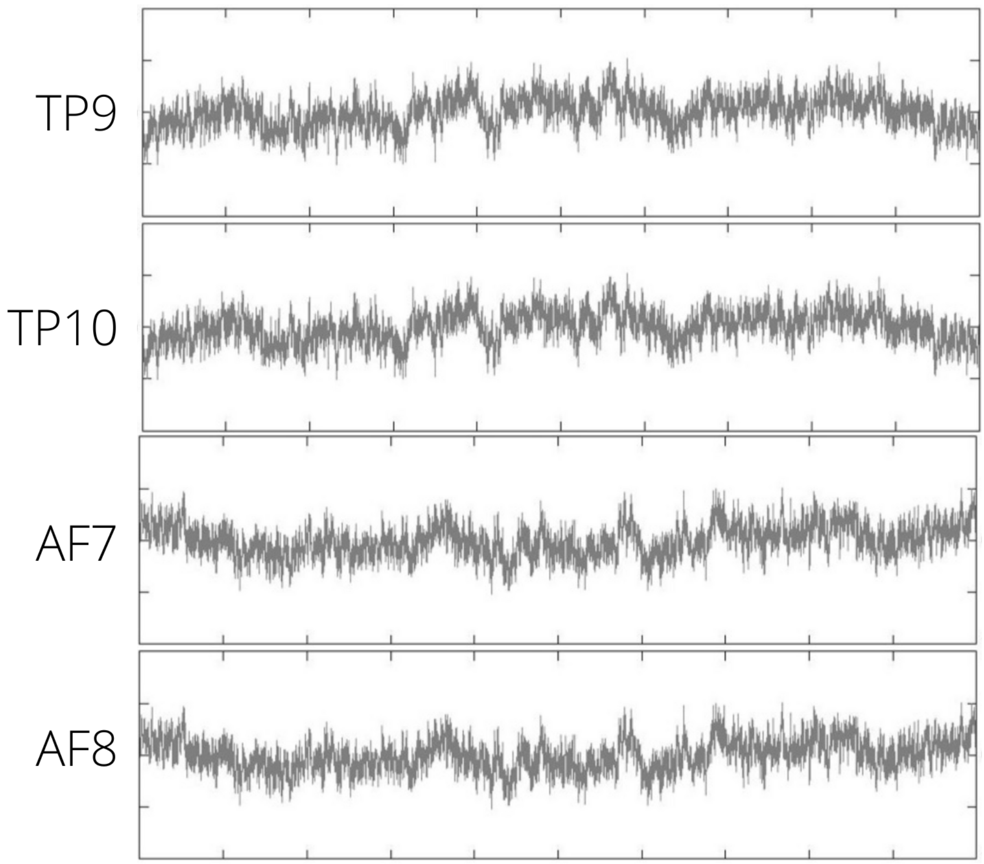

Figure 13: Signals collected from different channels