LSTM-Looker, a Predictor of Eye Position

Owen Webb.
Department of Computer Science
Swarthmore College
Swarthmore, PA 19081
owebb1@swarthmore.edu

Emily Hapgood
Department of Computer Science
Swarthmore College
Swarthmore, PA 19081
ehapgoo1@swarthmore.edu

Abstract

EEG signals are a complex source of training data, yet the importance of being able to make accurate predictions is immense. EEGEyeNet created a baseline of machine and deep learning models which were trained and tested on the large dataset of EEG and Eye Tracking information provided by the authors. While many different models were tested in that paper, it was not exhaustive, so we would like to add a machine learning model and deep learning model to the list of models tested on EEGEyeNet in order to improve overall accuracy of prediction. Least Angle Regressors (LARS) have been shown to fit well to high dimensional data such as EEG data, but we found otherwise. LARS showed an average rms distance of 119.12 mm which is consistent to what is shown for other machine learning models in EEGEyeNet. Long Short-Term Memory neural networks (LSTM) have previously been shown to improve motor imagery classification from EEG data, however when we implemented this model it performed worse than CNNs when training on EEG data to predict absolute eye position.

1 Introduction

1.1 Problem statement

The ability to make predictions about an individual from their EEG data would improve multiple fields including behavioral science, assistive technology, or user experience (Hammon u. a. (2008), Bamdad u. a. (2015)). However, the main difficulty of learning EEG data is that it is time series data, which can make most traditional machine learning models and deep learning models inefficient at fully learning features and predictive traits of the data during training (Nagabushanam u. a. (2020), Roy u. a. (2019)). Given the high frequency of artifacts in the data, it is possible for these models to overfit to unimportant details rather than have a full understanding of patterns in signal that would be possible if time and order were accounted for in training Saqib u. a. (2020). Another shortcoming of using large EEG datasets is that many models take long periods of time to train and test the models Praveena u. a. (2020). This time commitment is inefficient and potentially detrimental in several applications that would require the instant prediction of behavior from EEG data.

In "EEGEyeNet: a Simultaneous Electroencephalography and Eye-tracking Dataset and Benchmark for Eye Movement Prediction", Kastrati u. a. (2021) were one of the first groups to create a large dataset tracking both eye movement as well as EEG data. This published dataset is an important contribution to the field because it allows the validation of predictions made from deep learning models and machine learning algorithms created by others. One shortcoming of their research is that they tested several simple machine learning algorithms and deep learning was less of a focus, in which they only tested five separate variations of Convolutional Neural Networks (CNN). We propose that there are other machine learning algorithms and neural network models that could be used on this dataset to establish even more accurate results, especially in predicting absolute position of an

Preprint. Under review.

individual's gaze. We predict that using Least-angle Regression may outperform several machine learning models used by Kastrati u. a. (2021), as it is designed to work best with high-dimentional data. In previous EEG research, Long Short-Term Memory (LSTM) Neural Networks have shown positive results for improving classification accuracy. We believe that LSTM models could improve analysis of the EEGEyeNet dataset. Specifically on EEG eye movement signals, absolute position regression is an issue that previous machine learning models have struggled with. One issue is that EEG signals are time series data, which can cause trouble for prediction because in most neural networks, time is not a variable and therefore the previous patterns do not influence what will come next. The problem with this is that time should influence prediction with this dataset because the EEG signal is continuous and will not make significant jumps in a period of zero seconds. Given this, we know that we should be able to use the last place the EEG signal was as a jumping point, while Convolutional Neural Nets are not as accurate at processing this type of data. We see this conflict as a possible problem that was not explored during the research in the EEGEyeNet paper done by Kastrati u. a. (2021).

Addressing these concerns, will Least-angle Regression or Long Short-Term Memory Neural Networks improve the accuracy of estimating gaze location from EEG data?

1.2 Literature review

Previous work done to estimate gaze location by Kastrati u. a. (2021) compared accuracy of several Convolutional Neural Networks and machine learning algorithms trained on EEG data. They found that the Convolutional Neural Nets outperformed the machine learning algorithms consistently, but only had close accuracy at predicting direction (left versus right) of gaze. There is significant room for improvement in gaze estimation of angle/amplitude as well as absolute position of gaze.

We predict training a LARS model may improve the accuracy of absolute gaze prediction, as the models Kastrati u. a. (2021) tested performed poorly (about at the accuracy of a naive baseline). LARS Efron u. a. (2004) is best at fitting high-dimensional data with significantly high predictors making it well suited for EEG data. When LARS is applied to sparse coding of ECoG signals Deng u. a. (2018), it was found to help prevent some overfitting that was seen in other algorithms. It also lead to a higher classification accuracy when tested on this motor imagery ECoG dataset with sparse representation-based classification when compared to just SVM or KNN Deng u. a. (2018). An addition benefit to LARS is that it is more computationally effective as shown in Figure 1 than other linear machine learning algorithms Khan u. a. (2007). It uses an additional mathematical formula to accelerate the computations so that only m steps are required for the full set of solutions, where m is the number of covariates Efron u. a. (2004).

In a review by Roy u. a. (2019), one of the main challenges in using deep learning models trained on EEG data is that EEG is a non-stationary signal. This can lead to poor generalizability between time points in data. Our solution to this problem is to use a long short-term memory (LSTM) model. LSTM models add benefit to time series data, and may improve absolute gaze prediction over traditional convolutional neural networks given a better understanding of EEG data over time.

In research done by Nagabushanam u. a. (2020), LSTM neural networks were found to improve accuracy of EEG motor imagery classification from other machine learning models. LSTM models are based on Recurrent Neural Networks that do not use external memories to store previous outputs Nagabushanam u. a. (2020). This research is significant because they were able to use 2-layer LSTM neural networks to improve the classification accuracy of motor imagery data which points us in the direction that LSTM can be used to improve the accuracy of predictions for the EEGEyeNet dataset as well. Their LSTM model had higher accuracy, precision, and recall than their traditional Neural Net.

Another example of LSTM improving classification accuracy on EEG data is in Alhagry u. a. (2017). They found that in analysis of EEG data to classify emotion recognition, an LSTM model was very powerful as a tool because of its ability to learn features from the raw data directly. In EEGEyeNet Kastrati u. a. (2021), the minimally preprocessed data as input to the CNN models performed the best, so if we are able to train an LSTM on the raw data, we would expect that to succeed. We would like to improve the accuracy predicting a participant's absolute eye position from their EEG data by taking advantage of memory as a factor of the dataset Siami-Namini u. a. (2018) by implementing an LSTM as shown in Figure 2 Hochreiter und Schmidhuber (1997). An LSTM model is able to use past patterns in data to assume future steps to be dependent on previous steps. We predict that our LSTM model may take significantly longer to run, and to address this we plan to parallelize some aspects of our training Ouyang u. a. (2017)

1.3 Purpose of study

The purpose of our study is to improve the accuracy of gaze prediction from EEG data using Least-angle Regression and a Long Short-Term Memory Neutral Net. We predict that using LARS may outperform several machine learning models used by Kastrati u. a. (2021). We also predict using a LSTM model will improve absolute position estimation when compared to the CNN models in EEGEyeNet Kastrati u. a. (2021). This research could drastically improve the accuracy of deep learning in eye movement detection and encourage future research to be done to improve our model using LSTMs. We hope to establish LSTM as the tool to be used when classifying EEG signals.

1.4 Research Questions

Due to limiting factors of Convolutional Neural Networks, can Long Short-Term Memory Neural Networks improve the ability of estimating gaze location from EEG data, specifically the EEGEyeNet Dataset Kastrati u. a. (2021)? Can Least-angle Regression outperform other machine learnign algorithms in this same task?

2 Dataset

2.1 Dataset Description

EEG datasets are traditionally very small and difficult to train deep learning models on successfully without overfitting to specific points within the dataset. In order to best train our LSTM model we are using the dataset from "EEGEyeNet: a Simultaneous Electroencephalography and Eye-tracking Dataset and Benchmark for Eye Movement Prediction" Kastrati u. a. (2021) due to its large size and clear direction. The code is well documented to allow for custom implementation of both machine and deep learning models. This dataset provides synchronized EEG data and Eye Tracking validation. The processed data is provided in the form of a MATLAB structure containing all electrodes and channels (time, x-coordinate of eye, y-coordinate of eye, pupil size) with their corresponding experimental events. Experimental data is provided for each of the benchmark tasks referenced in the paper; the antisaccade paradigm, large grid paradigm, and visual symbol search paradigm Kastrati u. a. (2021). The authors provide raw, minimally, and maximally processed data, however the goal for our study is to use the minimally processed data provided to determine if LSTM or LARS performs better than the originally tested algorithms. We are using only the large grid paradigm, so this is what is used to test absolute gaze prediction. In the data downloaded from Kastrati u. a. (2021), there are two separate datasets. One has used the Hilbert transform, provided by Kastrati u. a. (2021), as a feature extraction method to improve the performance of out LARS algorithm. The Hilbert transform is a feature selection process that results in a complex time series from which phase and amplitude are extracted. The other set is the minimally band passed data. Our LSTM model ended up performing better without the feature extraction provided. Due to the complexity of EEG data, it is extremely difficult to identify important features during training without overfitting, so predicting absolute gaze position is a task that will not easily be performed accurately.

3 Experiment

3.1 Experiment Design

As shown in the dataset description, the EEGEyeNet dataset contains raw, minimally preprocessed and maximally preprocessed data as well as several different types of tasks users were asked to complete. For each type of data, Kastrati u. a. (2021) established a baseline using their basic alorithms. Our experiment will first compare LARS to all of the regression models presented by Kastrati u. a. (2021). The experiment we would like to replicate would be to train and test our LSTM and LARS models on the minimally processed data for the most difficult task discussed in the paper, the prediction of absolute position. Each classification accuracy of our models would be compared to the baseline established in Kastrati u. a. (2021). Given that LSTM is known for having better accuracy on raw EEG data, and LARS works well with high-dimensional data, this experiment may produce the most significant results using minimally preprocessed data and comparing the classification accuracy to the baseline scores. While the first task, right vs left eye movement, had a baseline set, it is unlikely that this algorithm will obtain higher accuracy as the baseline accuracy was above 99% and the LSTM model may produce similar results at the cost of an increase in speed so we are not testing our models on this task.

The two tasks where an LSTM neural network is likely to improve error are angle/amplitude estimation, and absolute position estimation. Due to the large size of the data, we will be training and testing our model on the minimally processed absolute position data as we believe this task has the most room for improvement. As explained in Kastrati u. a. (2021) in reference to absolute position estimation, "In order to estimate the current gaze position we expect past information, e.g., the previous gaze position, to be helpful". Because Nagabushanam u. a. (2020) states that LSTM takes into account time based prediction, our experiment is designed to highlight the importance of time in the estimation of absolute position when training a neural network.

3.2 Comparison of Algorithms

We will compare the accuracy of our Least-angle regression and LSTM models to the models trained by Kastrati u. a. (2021). The comparisons consist of determining if our models have lower accuracy of error when predicting absolute position than the previously explored machine learning models and CNNs.

LARS is designed to work best with high dimensional data, which is becoming an increasingly common type of data as it is now relatively simple to collect and store large quantities of data. With high dimensional data, the number of learnable features may be higher than the sample size, so it is important to be able to make strategic variable selections as well as perform accurate regression analysis. It is likely that this EEG data will have a significantly high number of learnable features, and it is difficult as well as computationally expensive to fit a linear model to such data.

LARS offers benefits that other types of regression to not in that it is computationally efficient while also providing a full piecewise linear solution path. At each step, it finds the feature most correlated with the target value. If multiple features have equal correlation, instead of continuing along the same feature, it will move in a direction that is equiangular between the features as shown in Figure 1. Because it works well with high dimensional data, it is designed to accommodate contexts where the number of predictors is large. This algorithm also works just as fast as forward selection Pedregosa u. a. (2011). However, LARS is increasingly susceptible to noise (like EEG artifacts), indicating that the selected variables may not be significant to the datasource because of the relation between residual fitting and the high correlation of variables Efron u. a. (2004).

Long Short Term Memory (LSTM) is a version of the Recurrent Neural Network(RNN) that is able to hold onto longer term states, keeping a history of the sequence. LSTMs use a series of 'gates' to achieve this. Each of which control how the information in a sequence of data comes into, is stored in, and leaves the network. There are three gates in a typical LSTM; forget gate, input gate and output gate. The forget gate determines which pieces of information are useful given the past history states and the new input data. The input gate is simply a filter that processes the information coming into the network. The output gate acts as a filter that will process the information and therefor limit the information expelled.

Using a LSTM model, Figure 2, on EEG data will provide increasing benefits that CNNs or other RNNs cannot provide. CNNs train in batches, rather than the sequential training of RNNs, so the predictions made are not influenced by long term previous patterns in data or by the data preceding a given timestep, which would hinder predictions made from EEG data. An additional benefit of LSTMs over other types of RNNs is that RNNs are unable to learn relevant input information over large spans of data, whereas the addition of 'gates' to LSTMs allows for the retention of important sections of data, and more importantly, the forgetting of non-important sections Yu u. a. (2019). Given that we trained on an extremely large EEG dataset, it is unlikely that we would be able to successfully train on this dataset with a RNN model without running into the exploding gradient problem. While CNNs are computationally more effective, and are able to train on this dataset, they do not offer the same promise for long-term sequential learning that LSTM does.

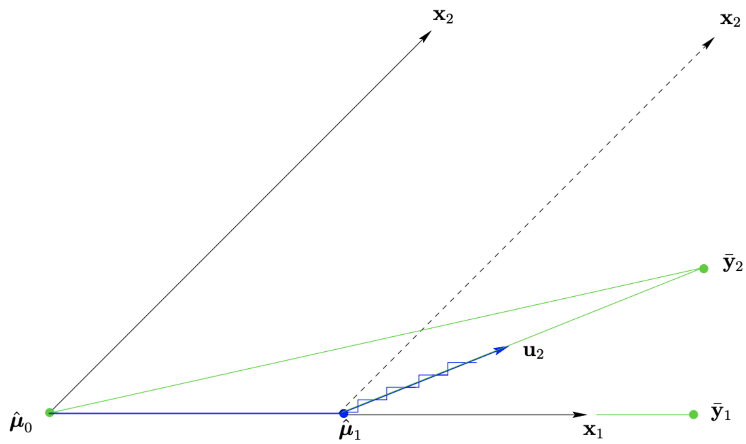

Figure 1: The LARS algorithm in the case of m = 2 covariates Efron u. a. (2004)

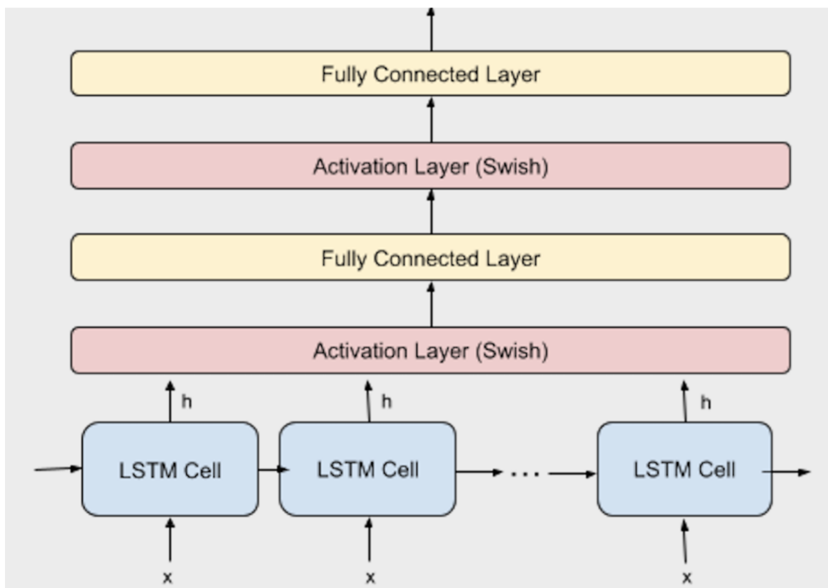

Figure 2: Architecture of our LSTM-Looker used in the methods.

4 Results

As shown in Table 1, the rms distance of LARS remained similar to the distance from the absolute position found in EEGEyeNet on the machine learning models providied, K-Nearest Neighbors (KNN)Fix und Hodges (1989), Linear Regression (LinearReg)A.M. Legendre (1805), Ridge RegressionHoerl und Kennard (1970), Lasso RegressionTibshirani (1996), Elastic NetZou und Hastie (2005), Radial-Basis Function Support Vector Regressor (RBF-SV) Cortes und Vapnik (2009), Random ForestBreiman (2001), Gradient BoostFriedman (2001), Ada Boost Freund u. a. (1999), and XGB Boost Chen und Guestrin (2016).

The addition of Least Angle Regression (LARS) to the research helps to confirm that while many of the machine learning algorithms were very fast, they are not accurate in predicting absolute position. While hoping for a better result, the RMS distance(mm) of 119.12 mm was expected from a simple machine learning algorithm. All of the EEGEyeNet implemented algorithms showed results similar to Kastrati u. a. (2021). However, the training time of LARS was 0.64 seconds, notably quicker than Random Forest or Gradient Boost.

Model	rms distance (mm)	Runtime (s)
KNN	119.68	4.92
LinearReg	118.37	0.38
Ridge	118.25	0.10
Lasso	118.04	1.90
ElasticNet	118.13	2.02
RBF-SVR	123.01	16.79
Random Forest	116.75	1,297.49
Gradient Boost	117.53	1,422.02
Ada Boost	119.42	948.35
XGB Boost	117.98	138.09
LARS	**119.12**	**0.64**

Table 1: rms distance from absolute position data and the runtime of each Machine Learning algorithm for training and testing on the EEGEyeNet absolute position minimally preprocessed dataset

Model	rms distance (mm)	Runtime (s)
CNN	70.67	667.0
PyramidalCNN	71.56	448.6
EEGNet	77.45	4396.6
InceptionTime	69.78	2492.1
Xception	76.75	1230.3
LSTM	**89.73**	**760.46**

Table 2: rms distance from absolute position data and the runtime of each of the Deep Learning Algorithms. Bolded is the LSTM implemented for this innovation

Our LSTM recorded an average rms distance of 89.73mm in 760.46s. The tested CNN LeCun u. a. (1998) models still beat our LSTM as seen in Table 2, but the accuracy of our LSTM was still much better than traditional ML algorithms. Table 2 shows multiple versions of the CNN that performed similarly to eachother, all slightly beating our LSTM accuracy.

5 Discussion

The highest performing machine learning algorithm on the minimally processed EEG data was Random forest shortly followed by several boosting algorithms: Gradient Boost and XGB Boost. Boosting algorithms tend to perform better on data similar to EEG data because they are ensemble algorithms that reduce bias and variance in data as shown by Verbaeten und Assche (2003). One of the innovations implemented in this paper, Least-angle regression, performed just behind these models in terms of accuracy, but did so in nearly 1/2000th of the time. Despite the slight improvement in accuracy among machine learning models, none performed substantially better than the naive baseline of 123.3 mm. The task the models were trained to perform is a difficult task: predicting absolute position of an individual's gaze based on minimally processed EEG data. One reason they may be performing poorly is due to the non-stationary and order dependent nature of EEG data. Least-angle regression is designed for high-dimensional data, however it is susceptible to high amounts of noise Weisberg (2004). It is likely that the variables selected in the regression are not the actual variables that are determining absolute eye position, therefor behaving similarly to the naive baseline.

The LSTM did not perform as well as the theory suggests, but it is not abundantly clear as to why it performed worse than the CNN models. Kastrati u. a. (2021) claimed that the top CNN model performed with 70mm accuracy. Recurrent Neural Networks in general are resistant to noise, which may be beneficial when training on the minimally processed data. An LSTM model is designed to learn sequence prediction problems, indicating that it may be ideally suited for EEG datasets. However, our LSTM model was not able to outperform other deep learning models discussed in Kastrati u. a. (2021). Running our LSTM model with the additional use of CUDA decreased the runtime, but would also decrease the runtime of the original CNN models. The prediction that these models are making is a difficult task, the EEG data may contain many artifacts and is not straightforward to detect the correct features that would lead to an accurate prediction of a participants' gaze.

There are multiple papers that have found similar results where the LSTM alone did not perform well enough. Hasib u. a. (2018) shows that a hierarchical LSTM gained an improvement of 12% over the original LSTM on EEG data. The LSTM in this paper shows the same accuracy as simpler machine learning models such as Support Vector Machines (SVM)Cortes und Vapnik (1995). This result shows that a more complex LSTM implementation may be needed to improve accuracy further. Li u. a. (2019) created a review of all deep learning used in EEG classification where it found that a combination hybrid model between CNN and LSTM was the most promising at detecting disease from EEG data. This again shows that while LSTM can be used for classification, but there are improvements to be made. Zhang u. a. (2020) created a model tested on EEG hand classification signals using an attention based LSTM. The attention based LSTM is an interesting direction that this paper could explore. The attention based LSTM performed 8% better than the traditional LSTM. In this paper, if those results were achieved, the attention based LSTM would achieve roughly a 81mm accuracy.

5.1 Future Directions

To attempt to improve the overall accuracy of prediction of absolute gaze position from EEG data we implemented a Long Short-Term Memory (LSTM) model, which did not out perform traditional CNNs. It is possible that future work that uses combined LSTM networks with other deep learning methods may lead to higher accuracy of prediction. As discussed earlier, this combination approach had yielded improved results on other types of datasets, and is likely to improve accuracy of prediction here. Using both models may allow for the generalized features to be identified with early CNN layers, allowing a high level initial identification, followed by LSTM layers to remember long term dependencies. Another aspect of our training process that we think could be improved with more time is feature detection, our models, as well as the CNNs tested from Kastrati u. a. (2021), all performed worse with feature detection. If we were able to accurately identify a few important EEG features used in training with, then the overall accuracy may improve. Using methods developed by Truong u. a. (2021) may be helpful to identify weaknesses in our current feature detection that would allude to better strategy. In order to train and test a substantial amount of future LSTM or combination models, it would be helpful to use a high performance computer cluster to decrease overall runtime.

6 Conclusion

We determined that both algorithms that were newly tested on this large EEG dataset, LARS and LSTM, did not outperform models that have been previously shown to predict absolute gaze position. While each did have certain overall benefits, neither increased the accuracy of being able to learn from EEG data. Being able to make accurate predictions from EEG data quickly would have a large impact on the development of BCIs used in assistive technology devices, and this paper adds light as to what future directions should be explored so that these devices and others can be more helpful to those who need them.

References

[Alhagry u. a. 2017] ALHAGRY, Salma ; FAHMY, Aly A. ; EL-KHORIBI, Reda A.: Emotion Recognition based on EEG using LSTM Recurrent Neural Network. In: *International Journal of Advanced Computer Science and Applications* 8 (2017), Nr. 10. – URL http://dx.doi.org/10.14569/IJACSA.2017.081046

[A.M. Legendre 1805] A.M. LEGENDRE, F. D.: Nouvelles methodes pour la determination des orbites des cometes [microform]. (1805)

[Bamdad u. a. 2015] BAMDAD, Mahdi ; ZARSHENAS, Homayoon ; AUAIS, Mohammad A.: Application of BCI systems in neurorehabilitation: a scoping review. In: *Disability and Rehabilitation: Assistive Technology* 10 (2015), Nr. 5, S. 355–364

[Breiman 2001] BREIMAN, Leo: Random forests. In: *Machine learning* 45 (2001), Nr. 1, S. 5–32

[Chen und Guestrin 2016] CHEN, Tianqi ; GUESTRIN, Carlos: XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, aug 2016. – URL `https://doi.org/10.1145%2F2939672.2939785`

[Cortes und Vapnik 1995] CORTES, Corinna ; VAPNIK, Vladimir: Support-vector networks. In: *Machine learning* 20 (1995), Nr. 3, S. 273–297

[Cortes und Vapnik 2009] CORTES, Corinna ; VAPNIK, Vladimir: Support-vector networks. In: *Chem. Biol. Drug Des.* 297 (2009), 01, S. 273–297

[Deng u. a. 2018] DENG, Xin ; LI, Danni ; MI, Jianxun ; GAO, Fengxing ; CHEN, Qiaosong ; WANG, Jin ; LIU, Rui: Motor Imagery ECoG Signal Classification Using Sparse Representation with Elastic Net Constraint. In: *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)* (2018), S. 44–49

[Efron u. a. 2004] EFRON, Bradley ; HASTIE, Trevor ; JOHNSTONE, Iain ; TIBSHIRANI, Robert: Least Angle Regression. In: *The Annals of Statistics* 32 (2004), Nr. 2, S. 407–499

[Fix und Hodges 1989] FIX, Evelyn ; HODGES, J. L.: Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. In: *International Statistical Review / Revue Internationale de Statistique* 57 (1989), Nr. 3, S. 238–247. – URL `http://www.jstor.org/stable/1403797`. – Zugriffsdatum: 2022-05-09. – ISSN 03067734, 17515823

[Freund u. a. 1999] FREUND, Yoav ; SCHAPIRE, Robert ; ABE, Naoki: A short introduction to boosting. In: *Journal-Japanese Society For Artificial Intelligence* 14 (1999), Nr. 771-780, S. 1612

[Friedman 2001] FRIEDMAN, Jerome H.: Greedy function approximation: a gradient boosting machine. In: *Annals of statistics* (2001), S. 1189–1232

[Hammon u. a. 2008] HAMMON, Paul S. ; MAKEIG, Scott ; POIZNER, Howard ; TODOROV, Emanuel ; DE SA, Virginia R.: Predicting Reaching Targets from Human EEG. In: *IEEE Signal Processing Magazine* 25 (2008), Nr. 1, S. 69–77

[Hasib u. a. 2018] HASIB, Md M. ; NAYAK, Tapsya ; HUANG, Yufei: A hierarchical LSTM model with attention for modeling EEG non-stationarity for human decision prediction. In: *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2018, S. 104–107

[Hochreiter und Schmidhuber 1997] HOCHREITER, Sepp ; SCHMIDHUBER, Jürgen: Long Short-term Memory. In: *Neural computation* 9 (1997), 12, S. 1735–80

[Hoerl und Kennard 1970] HOERL, Arthur E. ; KENNARD, Robert W.: Ridge regression: Biased estimation for nonorthogonal problems. In: *Technometrics* 12 (1970), Nr. 1, S. 55–67

[Kastrati u. a. 2021] KASTRATI, Ard ; PŁOMECKA, Martyna Martyna B. ; PASCUAL, Damián ; WOLF, Lukas ; GILLIOZ, Victor ; WATTENHOFER, Roger ; LANGER, Nicolas: EEGEyeNet: a Simultaneous Electroencephalography and Eye-tracking Dataset and Benchmark for Eye Movement Prediction. In: *arXiv preprint arXiv:2111.05100* (2021)

[Khan u. a. 2007] KHAN, Jafar A. ; AELST, Stefan V. ; ZAMAR, Ruben H.: Robust Linear Model Selection Based on Least Angle Regression. In: *Journal of the American Statistical Association* 102 (2007), Nr. 480, S. 1289–1299

[LeCun u. a. 1998] LECUN, Yann ; BOTTOU, Léon ; BENGIO, Yoshua ; HAFFNER, Patrick: Gradient-based learning applied to document recognition. In: *Proceedings of the IEEE* 86 (1998), Nr. 11, S. 2278–2324

[Li u. a. 2019] LI, Gen ; LEE, Chang H. ; JUNG, Jason J. ; YOUN, Young C. ; CAMACHO, David: Deep learning for EEG data analytics: A survey. In: *Concurrency and Computation: Practice and Experience* 32 (2019), Februar, Nr. 18. – URL `https://doi.org/10.1002/cpe.5199`

[Nagabushanam u. a. 2020] NAGABUSHANAM, Perattur ; GEORGE, S. T. ; RADHA, Subramanyam: EEG signal classification using LSTM and improved neural network algorithms. In: *Soft Computing* 24 (2020), S. 9981–10003

[Ouyang u. a. 2017] OUYANG, Peng ; YIN, Shouyi ; WEI, Shaojun: A fast and power efficient architecture to parallelize LSTM based RNN for cognitive intelligence applications. (2017), S. 1–6

[Pedregosa u. a. 2011] PEDREGOSA, F. ; VAROQUAUX, G. ; GRAMFORT, A. ; MICHEL, V. ; THIRION, B. ; GRISEL, O. ; BLONDEL, M. ; PRETTENHOFER, P. ; WEISS, R. ; DUBOURG, V. ; VANDERPLAS, J. ; PASSOS, A. ; COURNAPEAU, D. ; BRUCHER, M. ; PERROT, M. ; DUCHESNAY, E.: Scikit-learn: Machine Learning in Python. In: *Journal of Machine Learning Research* 12 (2011), S. 2825–2830

[Praveena u. a. 2020] PRAVEENA, D. M. ; SARAH, D. A. ; GEORGE, S. T.: Deep Learning Techniques for EEG Signal Applications – a Review. In: *IETE Journal of Research* 0 (2020), Nr. 0, S. 1–8

[Roy u. a. 2019] ROY, Yannick ; BANVILLE, Hubert ; ALBUQUERQUE, Isabela ; GRAMFORT, Alexandre ; FALK, Tiago H. ; FAUBERT, Jocelyn: Deep learning-based electroencephalography analysis: a systematic review. In: *Journal of neural engineering* 16 (2019), Nr. 5, S. 051001

[Saqib u. a. 2020] SAQIB, Mohammed ; ZHU, Yuanda ; WANG, May ; BEAULIEU-JONES, Brett: Regularization of Deep Neural Networks for EEG Seizure Detection to Mitigate Overfitting. In: *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2020, S. 664–673

[Siami-Namini u. a. 2018] SIAMI-NAMINI, Sima ; TAVAKOLI, Neda ; SIAMI NAMIN, Akbar: A Comparison of ARIMA and LSTM in Forecasting Time Series. (2018), S. 1394–1401

[Tibshirani 1996] TIBSHIRANI, Robert: Regression shrinkage and selection via the lasso. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1996), Nr. 1, S. 267–288

[Truong u. a. 2021] TRUONG, Dung ; MAKEIG, Scott ; DELORME, Arnaud: Assessing learned features of Deep Learning applied to EEG. (2021). – URL `https://arxiv.org/abs/2111.04309`

[Verbaeten und Assche 2003] VERBAETEN, Sofie ; ASSCHE, Anneleen V.: Ensemble methods for noise elimination in classification problems. In: *International workshop on multiple classifier systems* Springer (Veranst.), 2003, S. 317–325

[Weisberg 2004] WEISBERG, Sanford: [Least Angle Regression]: Discussion. In: *The Annals of Statistics* 32 (2004), Nr. 2, S. 490–494. – URL `http://www.jstor.org/stable/3448473`. – Zugriffsdatum: 2022-04-12. – ISSN 00905364

[Yu u. a. 2019] YU, Yong ; SI, Xiaosheng ; HU, Changhua ; ZHANG, Jianxun: A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. In: *Neural Computation* 31 (2019), 07, Nr. 7, S. 1235–1270. – URL `https://doi.org/10.1162/neco_a_01199`. – ISSN 0899-7667

[Zhang u. a. 2020] ZHANG, Guangyi ; DAVOODNIA, Vandad ; SEPAS-MOGHADDAM, Alireza ; ZHANG, Yaoxue ; ETEMAD, Ali: Classification of Hand Movements From EEG Using a Deep Attention-Based LSTM Network. In: *IEEE Sensors Journal* 20 (2020), Nr. 6, S. 3113–3122

[Zou und Hastie 2005] ZOU, Hui ; HASTIE, Trevor: Regularization and variable selection via the elastic net. In: *Journal of the royal statistical society: series B (statistical methodology)* 67 (2005), Nr. 2, S. 301–320