

Variations of the LSTM Architecture for the Classification of Cognitive Stimuli-Based EEG Signals for BCI Applications

Prashant Srinivasan Sarkar ¹, E. Grace Mary Kanaga ², M. Bhuvaneshwari ³

^{1, 2, 3} Karunya Institute of Technology and Sciences

Abstract:- Deep learning is a branch of artificial intelligence most closely linked with the functioning of an actual human brain. Deep neural networks perform millions of simple computations that, when combined, allow for incredibly complex problem-solving capability. With recent breakthroughs in computational technology, neural networks are now larger than ever, capable of performing complex tasks that were hitherto believed to be science fiction. As modern technology grows closer to being capable of mimicking the human brain, researchers around the globe ponder the question: can this technology understand the brain? This experimental analysis of LSTM Deep Neural Networks proposes multiple variations of the traditional LSTM architecture as an optimized method for the classification of electroencephalogram (EEG) brain signals. It explores 5 variations of the LSTM architecture: Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, Stacked Bidirectional LSTM and LSTM with Attention Mechanism. This experimental analysis proposes a hybrid LSTM with Attention mechanism as a potential solution to create efficient, accurate and lightweight brain computer interfaces. In this experimental analysis, the proposed architecture achieved an F1 score of 94% for the classification of cognitive stimuli-based EEG signal data. These architectures show promise in the realm of brain computer interfacing as potential solutions for the future of human-computer interaction.

Keywords: Deep Learning, Deep Neural Network, LSTM, EEG Signals, Attention Mechanism, Bidirectional LSTM, Brain-computer Interface.

1. Introduction

A brain-computer interface (BCI), as the name implies, is a device capable of interpreting brain signals and converting them into machine action. Most BCIs function in relatively the same way, they use electrical impulses from the brain which correspond to different kinds of brain activity.

An Electroencephalogram (EEG) is a method of recording the spontaneous electrical activity of the brain. These signals represent the synaptic activity of the brain and can help doctors better understand the function of an individual's brain. Similarly, these signals can be passed to a computer to identify thought patterns and classify the intent of the subject.

The applications of a brain-computer interface range from security and authentication to IOT devices to aid the differently abled. Deep learning provides the most promising method to decode the immense complexity of the brain to provide a variety of new ways in which we can interact with computers.

The primary challenge faced by developers of BCI technology is the complexity of EEG signals. This complexity is not related to the data itself, but in the complex patterns created by the synaptic activity of trillions of neurons. Moreover, since every individual has a unique brain, these patterns created by the EEG signals can vary. Therefore, any BCI application must be tailored to the individual. Therefore, not only does a BCI application need to be capable of understanding the complex patterns of EEG signals, it also needs to be able to adapt to the unique brain of the individual.

The advent of deep learning has brought about a technological revolution that has extended into virtually every domain of academia and industry. Advances in computational capabilities of modern processors have allowed us to build AI systems capable of very complex learning behaviour. These powerful AI models could provide the necessary capabilities to produce a practical and accurate brain computer interface.

2. Literature Survey

An evaluation of classical and modern methods of brain-computer interface technology conducted by Hossain et al. delves into the current landscape of BCI research and the applications of artificial intelligence in augmenting BCI methodologies [1]. This study explored numerous public research databases such as PubMed, ERIC, JSTOR, IEEE Xplore and Google Scholar. 110 articles and nearly 20 years of research were reviewed to gain a thorough understanding of the timeline, history and progression of BCI research.

This study explored a variety of AI-based approaches to BCI applications including more traditional AI algorithms such as KNN clustering, Bayes Classification, and SVM regression as well as more modern and advanced deep neural networks such as LSTM and CNN. Finally, it highlights potential areas of exploration for future research such as Graph CNNs and Generative Deep Learning.

This study demonstrated the growing popularity and tremendous advancements in BCI technology as well as the impact of AI on its development. Furthermore, it demonstrates a strong correlation between the improvement of AI technology and the enhancement of BCI application accuracy and performance.

Research conducted at the Vellore Institute of Technology utilized the masking empirical mode decomposition (MEMD) to perform feature extraction on EEG signal data. MEMD performs a sequence of AM-FM frequency modulations on the data using a method called the Hilbert transform. After extraction, a min-max method was utilized to normalize the data. This extraction method, augmented by the min-max normalization allowed the researchers' feed forward back propagation deep neural network to achieve an incredibly high degree of accuracy [2].

Along the same vein of research, Alzahab et al. experimented with hybrid deep learning (HDL) networks based on Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Deep Belief Network (DBN) architectures to construct more powerful and accurate BCI applications [3]. This research was unable to prove any tangible benefit to applying these hybrid networks as they were not able to out-perform the plain RNN architecture. However, the results themselves were quite promising as the RNN architecture was able to exceed an accuracy of 95%.

A critical review of research and publications carried by Mansoor et al. has explored a wide variety of applications of deep neural networks towards EEG classification [4]. This study was carried out to assess the capabilities of these models to classify signals specifically for application in a BCI. The results conclude that deep neural networks have tremendous potential for present and future applications of BCI technology.

An editorial by Ahn et al. analysed two reviews and seven research papers about the applications of deep learning in BCI technology. It took into account the results of the research as well as the authors' perspectives and opinions on the outcome of the research [5]. This editorial not only demonstrates the capability of deep learning for augmenting BCI technology, but also highlights the number of methodologies and approaches towards applying deep learning as well as the number of researches pursuing this field of research.

An evaluation of current literature regarding the field of BCI technology and the applications of deep learning within the field reveals the potential for very promising exploration. It is evident that deep neural networks are capable of classifying very complex data to a high degree of accuracy. This makes them ideal for decoding the intricate patterns of EEG signals and interpreting the unique impulses generated by a subject's brain.

Research conducted by Ingolfsson et al. led to the development of a temporal convolutional neural network for BCI application. This model achieved an accuracy of 0.84 and is considered a fairly accurate model. This architecture utilizes the popular CNN model that is often used for BCI applications while also analysing the temporal dimension to improve the accuracy of time-series data [9].

Research by Li et al. describes their attempt to develop an accurate EEG signal classification model using a Bidirectional LSTM model with an attention mechanism. This study yielded exceptionally promising results with an accuracy of 90.04%, or 0.90 [14]. This model utilized a CNN submodule in parallel to the LSTM/attention model which was then concatenated with a SoftMax classification. This model performed very well on binary emotional stimuli-based EEG data, but was a fairly large model as it contained 2 deep neural networks running in parallel.

3. Proposed Methodology

This paper explores the LSTM architecture and its capability to classify EEG signals. The purpose for this task is for application within a brain-computer interface. This will be achieved by constructing multiple such models of several sizes, exploring the numerous variations of the LSTM architecture. These variations are: Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, and LSTM with Attention.

Data collection

This paper sought to utilize cognitive stimuli-based EEG signals to train the classification model. These signals were gathered using a GUSBamp brain cap. Data was gathered by having subjects perform various cognitive challenges while wearing the cap. These challenges were designed to stimulate logic and reasoning as well as image recognition and language. The goal of these challenges was to stimulate the brain across multiple different cognitive tasks to facilitate training on the numerous processes of the brain.

The resultant dataset consisted of over 200 thousand rows and 17 columns. The input data contains 16 columns representing the electrical signal picked up by one of the 16 sensors of the GUSBamp brain cap.

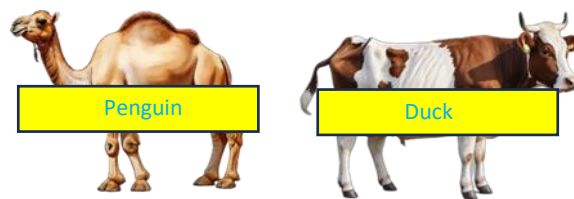


Fig. 1. Example cognitive stimuli used in data collection

The cognitive stimulus shown above is designed to trigger cognitive dissonance. The stimuli present a visual component that is in direct contradiction to the linguistic component shown in the text box. This contradictory information stimulates the brain and allows for a large amount of data to be collected [17].

LSTM

Long Short-Term Memory (LSTM) is a type of artificial neuron based off of the Recurrent Neural Network (RNN) neuron. This subset of neural network takes inputs at regular time intervals and is capable of retaining information from previous time steps. This past information is used to make more informed inferences for data that is dependent on past data. Such data includes audio, weather and signal data. However, one of the larger issues with the RNN architecture is that it struggles to maintain long-term dependencies. After a short amount of time, the RNN is less capable of retaining relationships between data. The LSTM cell was developed to overcome the issues of vanishing and exploding gradient that the RNN architecture faced [6].

The LSTM architecture is based on 3 'gates', namely the input, output and forget gate. It also has a cell state that is transferred to the next cell along with the cell's output. The forget gate is the first gate in the cell. The input is passed into this gate along with a weight and bias as a sigmoid function. This is then multiplied with the cell state. Essentially, if the result is 0, the information is forgotten, and if the result is 1, the information is remembered and passed to the next cell state.

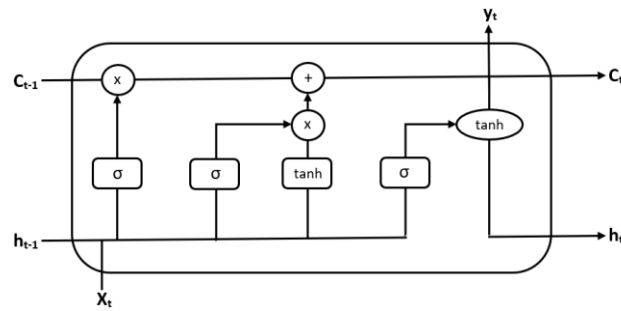


Fig. 2. LSTM Architecture

The input gate adds any new, useful information from the current timestep to the cell state. First, in a similar way to the forget gate, a sigmoid function is applied to the input to filter the values that are to be remembered. Then, a tanh function is applied to the input to form a vector. This vector is multiplied with the filtered information to generate useful information for the cell state to retain.

The output gate is responsible for extracting useful information from the cell state and presenting it as output. First, a tanh function is applied to the cell state value to create a vector. Then, a sigmoid function is applied to the input to once again generate the values to be remembered. These 2 values are multiplied together to generate output. The output for that timestep finally exits the cell as a processed inference. The output and cell state are also passed to the next cell in the network.

This more complex architecture provides the cell with more precision in terms of what information to retain and what information is unnecessary and might produce noise. This helps the overall network better retain information over longer periods of time while reducing the propagation of error through time due to noisy, unnecessary or inaccurate data.

These advantages to using LSTM networks makes them a powerful tool for language modelling, image processing, speech/handwriting recognition and even music generation. However, there are some drawbacks to using LSTM networks. First of all, due to the increased complexity of the neurons, LSTM networks can take longer to train. Moreover, as a result of the increased complexity, they are more memory intensive. It is therefore important, when using LSTM networks, to find an ideal balance between size and accuracy [7].

There are many variations to the LSTM architecture that improve upon the various capabilities of these networks to further augment the power of LSTM.

The most basic LSTM network is called the Vanilla LSTM. This network consists of a single layer of LSTM cells of a variable number. For classification using Vanilla LSTM, a SoftMax layer is added to classify into any number of target classes. If multiple such LSTM layers are connected sequentially, this is called a Stacked LSTM. The goal of a Stacked LSTM, as with any multi-layered neural network, is to extract better and more useful features to use for classification. Unlike with a Convolutional Neural Network, these features can also be oriented along the dimension of time. Stacking multiple layers of LSTM cells results in a longer training time, but an overall more accurate model in most cases.

LSTM architectures are designed to predict patterns based on sequential information. The standard LSTM model assumes that the sequence flows in the direction of time and predicts the sequence in a single direction. However, it is possible that more information might be inferred based on the backward flow of information. This is where the Bidirectional LSTM comes in. This architecture has a forward and backward layer connected parallelly to each other. This allows for features to be extracted based on the flow of sequential data in both directions. While this adds even more complexity to the network and a higher number of parameters, it allows for superior feature extraction for data where patterns can be observed in the reverse direction.

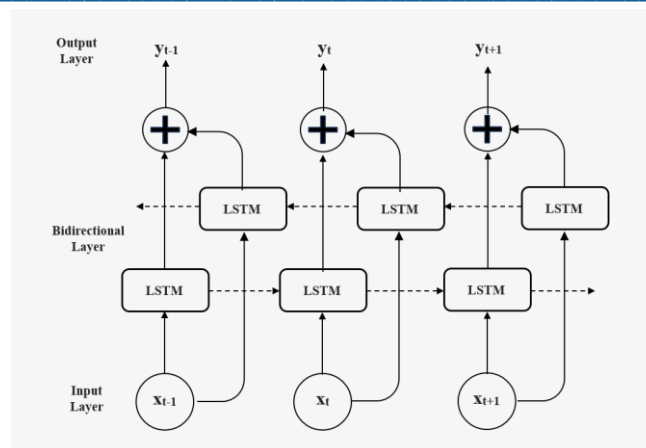


Fig. 3. Bidirectional LSTM Architecture

While the aforementioned architectures help with feature extraction and provide a more powerful model overall, the trade-off is that the model will be much larger. With EEG signal data, it is not necessarily complex features that make data difficult to classify. Instead, it could be the length of the sequence itself. And when it comes to long and complex sequences, the Attention Mechanism is perhaps the most powerful tool available. The Attention Mechanism is a relatively new neural network design that frees the encoder-decoder architecture from the fixed length internal representation. This means, in essence, that the model can focus or “pay attention” to specific inputs and relate them to the output regardless of the internal fixed length that the network is supposed to analyse. This is achieved by passing all past hidden states to the decoder or output so that the model has more context than if only the most recent hidden state is passed. Hidden states are scored with matrix multiplication before a SoftMax function is applied to determine the information that needs to be focused on from the past hidden states [8].

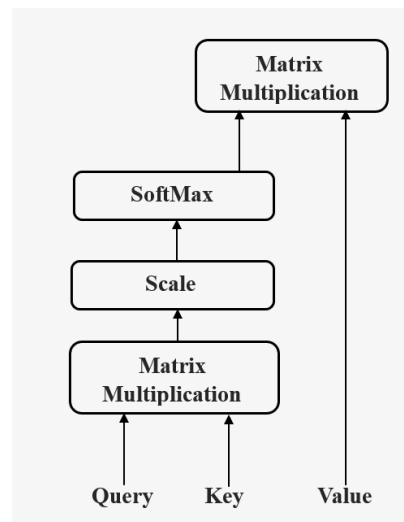


Fig. 4. Attention Mechanism

In the context of EEG signal classification, an attention mechanism would be applied in a slightly different way. Within a conventional encoder-decoder model, the Attention mechanism serves as a means to provide more context via a context vector to the decoder layer rather than simply passing the cell state. This provides more accurate context retention when generating sequential data. In this context however, the same principle cannot be applied as the purpose of the model is to classify. Therefore, the decoder layer was removed and replaced with a dense layer and softmax classifier that used the context vector to gain a better understanding of the EEG signal and therefore perform more accurate classification.

The experimental analysis outlined in this paper applied all of the aforementioned LSTM-based models to attempt to train a classifier for cognitive-stimuli based EEG signals. 5 models of varying sizes were trained for each LSTM variation. These sizes were based on the number of LSTM cells in each layer: 16, 32, 64, 128 and 256.

For the Stacked LSTM architecture, a 4-layer architecture was used. The decision to use 4 layers for the stacked LSTM architecture was based on the significant increase in parameters with the addition of more layers. As the 4-layer architecture is approximately 7 times larger than the Vanilla LSTM, the addition of any more layers would be counterproductive. This decision is further supported by prior research in cognitive stimuli-based EEG signal classification in which the impact of the number of layers was thoroughly explored and it was concluded that a 4-layer architecture performed exceedingly well, and the addition of further layers would not likely yield results much greater than those achieved by the 4-layer model [16].

For the Bidirectional LSTM model, a 1-layer and 2-layer model were both tested to determine whether the benefits of a Stacked-LSTM architecture could be carried over to a bidirectional architecture as well. For the LSTM with Attention, a single attention layer was applied to both a Vanilla and Stacked LSTM architecture.

The Keras library was utilized to facilitate the model construction while NVIDIA T4 GPUs via Google Colab were utilized to facilitate the training of these models. The data was then consolidated and analysed with respect to accuracy and model size to determine the ideal model for use in a BCI application.

4. Results

As per the methodology presented above, the LSTM models were constructed and trained on 200,000 rows of EEG signal data. Each row consists of 16 input variables and a single target output. The results for each model were compared against the LSTM with Attention mechanism created by Kim et al. which is considered to be fairly accurate with an accuracy of 0.90 [14]. This value is represented by the orange lines in fig 1 – 6.

For the Vanilla LSTM architecture, the results are as follows:

Table 1. Vanilla LSTM Performance Metrics

Size	No. Params	Size	Accuracy	F1 Score	Precision	Recall
16	1271	4.96	0.45	0.49	0.54	0.45
32	4583	17.9	0.56	0.60	0.64	0.56
64	17351	67.78	0.6	0.63	0.67	0.6
128	67463	263.53	0.61	0.64	0.68	0.61
256	265991	1034.24	0.63	0.65	0.68	0.63

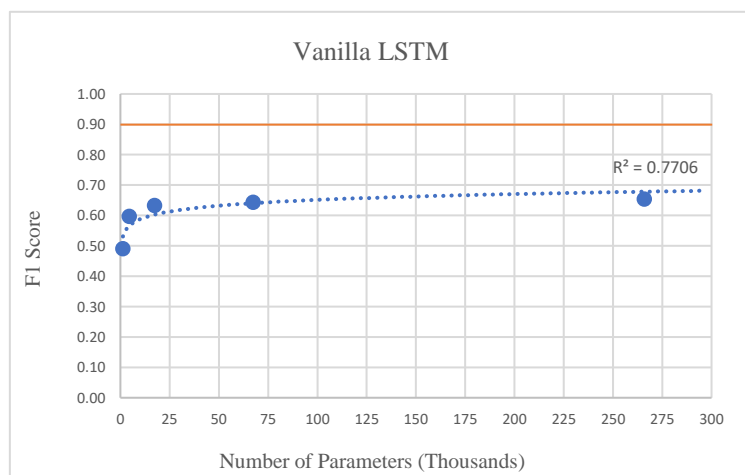


Fig. 5. Vanilla LSTM F1-Score vs Number of Parameters

As seen in the table and graph above, the Vanilla LSTM has a moderate classification capability that scales as expected as the number of LSTM cells in the layer (and the number of parameters) increase. From the above trend, it is also evident that this architecture is unlikely to be able to achieve an F1-Score above 0.7. Therefore, while we can observe the capability of the LSTM mechanism, it is clear that the Vanilla architecture is not complex enough for this classification task.

The next model that was tested was the Stacked LSTM. The results for this architecture are given below:

Table 2. Stacked LSTM Performance Metrics

Size	No. Params	Size	Accuracy	F1 Score	Precision	Recall
16	7607	29.71	0.74	0.75	0.76	0.74
32	29543	115.4	0.79	0.79	0.8	0.79
64	116423	454.78	0.91	0.91	0.91	0.91
128	462215	1802.24	0.94	0.94	0.94	0.94
256	1841927	7198.72	0.91	0.91	0.91	0.91

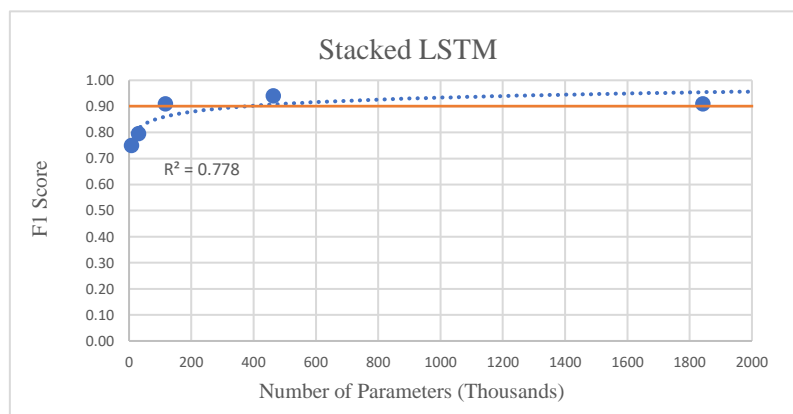


Fig. 6. Stacked LSTM F1-Score vs Number of Parameters

As seen in the above table and graph, the F1-Score increases along the same logarithmic trend as for the Vanilla architecture. Moreover, the proportion of the number of parameters also increases at the same rate as in the case of the Vanilla model. However, the accuracy is significantly higher for this architecture as it is capable of extracting more complex features from the data.

The Stacked LSTM architecture achieves a very high accuracy, though it is worth noting that the number of parameters is between 6 and 7 times higher than the Vanilla model. Other variations of the architecture have been explored to attempt to reduce this size constraint.

The Bidirectional LSTM was explored to determine whether more features could be extracted by passing the data in reverse, to supplement fewer layers and LSTM cells.

Table 3. 1-Layer Bidirectional LSTM Performance Metrics

Size	No. Params	Size	Accuracy	F1 Score	Precision	Recall
16	5895	23.03	0.73	0.74	0.76	0.73
32	15879	62.03	0.8	0.81	0.82	0.8
64	48135	188.03	0.82	0.83	0.84	0.82
128	161799	632.03	0.9	0.90	0.9	0.9
256	585735	2283.52	0.9	0.90	0.91	0.9

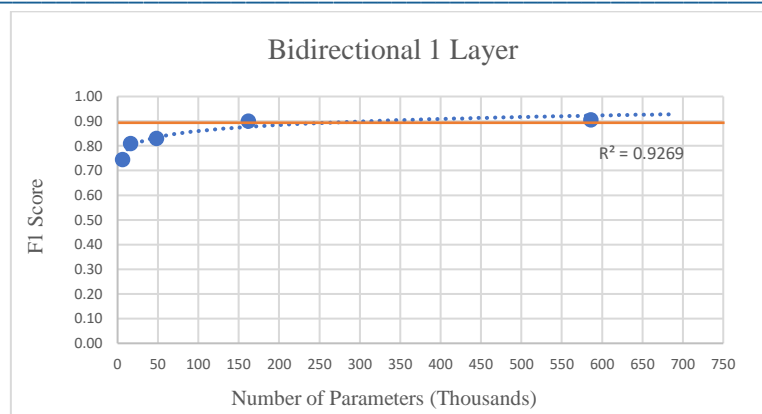


Fig. 7. 1-Layer Bidirectional LSTM F1-Score vs Number of Parameters

From the results of the 1-Layer Bidirectional LSTM model training, one can clearly observe the increased classification power of the Bidirectional LSTM architecture. This model is achieving results comparable to the stacked LSTM model with only double the number of parameters as the Vanilla LSTM model. In fact, the 128 model achieved the same score as the 256 model and is smaller than the best performing Vanilla LSTM by 100 thousand parameters.

As the 1-Layer Bidirectional demonstrated such excellent results, the 2-Layer bidirectional model was also explored in the hopes that it would yield the similar results:

Table 4. 2-layer Bidirectional LSTM Performance Metrics

Size	No. Params	Size	Accuracy	F1 Score	Precision	Recall
16	12167	47.53	0.78	0.78	0.78	0.78
32	40711	156.03	0.93	0.93	0.93	0.93
64	146951	574.03	0.94	0.94	0.94	0.94
128	556039	2170.88	0.94	0.94	0.94	0.94
256	2160647	8437.76	0.95	0.95	0.95	0.95

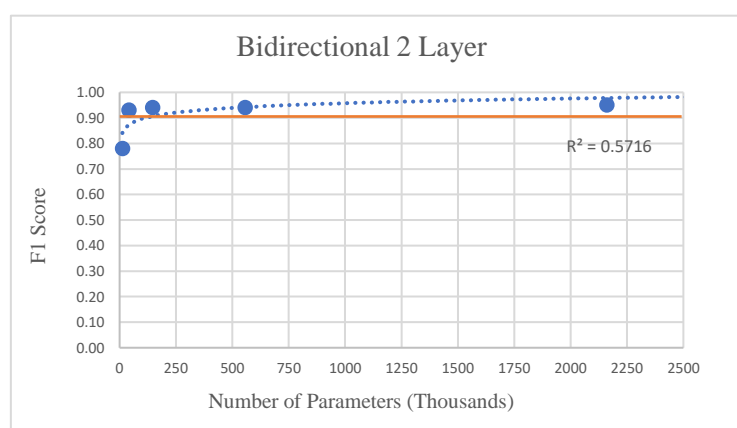


Fig. 8. 2-Layer Bidirectional LSTM F1-Score vs Number of Parameters

As expected, the 2-Layer Bidirectional model performed better than the 1-Layer variant in the same way that the Stacked LSTM performed better than the Vanilla LSTM. While the highest performance of 0.95 is only a small increment in relation to the much higher number of parameters, the size 64 model was able to achieve an F1-Score of 0.94 with far fewer parameters. The results from the Bidirectional LSTM models shows that a neural network can classify EEG data very effectively when it is passed in both directions.

The last variant of LSTM that was tested in this investigation was the LSTM with Attention. This variation was divided into 2 separate architectures: Vanilla LSTM with an Attention Mechanism and Stacked LSTM with an Attention Mechanism. The results for the Vanilla LSTM Attention are as follows:

Table 5. Vanilla LSTM with Attention Mechanism Performance Metrics

Size	No. Params	Size	Accuracy	F1 Score	Precision	Recall
16	1303	5.09	0.68	0.69	0.7	0.68
32	4631	18.09	0.68	0.69	0.71	0.68
64	17431	68.09	0.77	0.77	0.77	0.77
128	67607	264.09	0.84	0.85	0.87	0.84
256	266263	1044.48	0.9	0.90	0.9	0.9

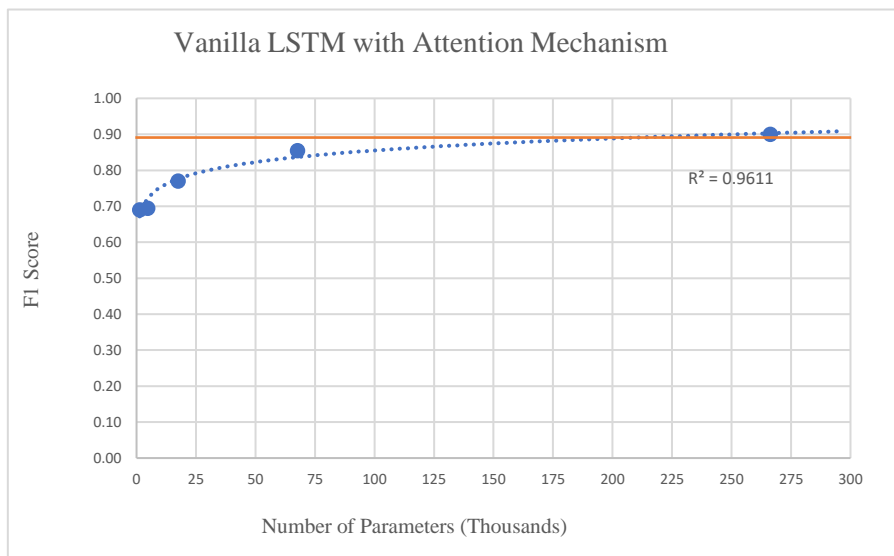


Fig. 9. Vanilla LSTM with Attention F1-Score vs Number of Parameters

As seen in the results above, the Attention Mechanism is able to significantly augment the results without significantly increasing the size of the model. By adding an Attention Mechanism to the Vanilla LSTM network, the F1-Score is increased until it is comparable to the stacked model while maintaining the size of the Vanilla LSTM model.

To leverage the power of the Attention Mechanism on more complex features, an Attention layer was added to the stacked LSTM architecture as well. The results of this architecture are as follows:

Table 6. Stacked LSTM with Attention Mechanism Performance Metrics

Size	No. Params	Size	Accuracy	F1 Score	Precision	Recall
16	7639	29.84	0.76	0.77	0.78	0.76
32	29591	115.59	0.8	0.81	0.82	0.8
64	116503	455.09	0.94	0.94	0.94	0.94
128	462359	1802.24	0.95	0.95	0.95	0.95
256	1842199	7198.72	0.95	0.95	0.95	0.95

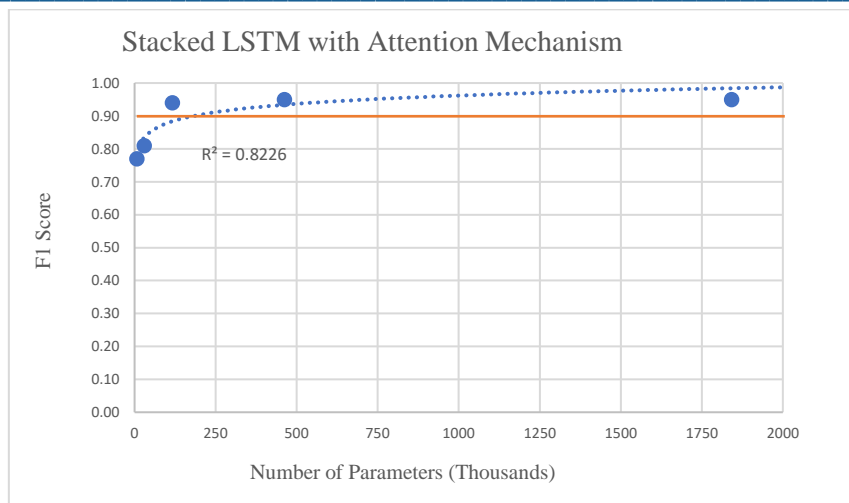


Fig. 10. Stacked LSTM with Attention F1-Score vs Number of Parameters

As seen in the data above, the Attention Mechanism is able to squeeze out a little more accuracy from the model with very little increase to the number of parameters. This allows for a noticeably more capable model with little to no compromise in terms of model size.

Looking at the data gathered from all the models, one can observe a number of interesting trends. The first is that each model follows roughly the same trend in terms of the increase in F1-Score in relation to number of parameters. A clear logarithmic relationship is present in all of the variations that were tested. From this trend, one can infer that regardless of the number of layers, there is a limit to how accurate the model can be based on the type of architecture used. Therefore, to achieve an effective EEG signal classifier, it is not size that is the primary constraint, but the neural network architecture.

Based on the capability of these various LSTM-based networks in relation to their size, one can plot a graphical representation of F1-Score against the number of parameters for each model:

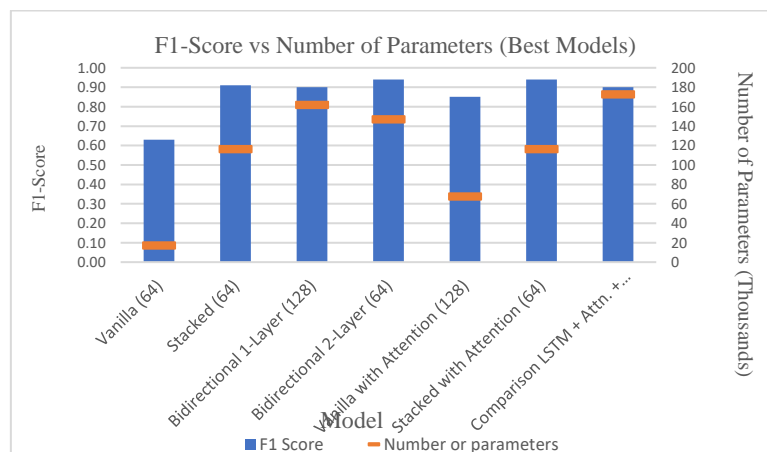


Fig. 11. F1 Score vs Number of Parameters for best architectures of all models

In the above figure, all F1 scores are compared to the LSTM with attention + CNN hybrid model developed by Li et al

From the consolidated data above, one can see that the LSTM architecture, in general, is very effective at classifying EEG signal data. If the LSTM architecture is further augmented by modifying the way in which features are learned, one can further enhance the classification capability of this architecture to construct a very powerful classification model.

Comparing all the best trained models, and evaluating them based on accuracy and size, it is evident that the Stacked LSTM with Attention provides the architecture most appropriate for use in a BCI.

5. Conclusion

This experimental analysis sought to build and test multiple variations of the LSTM architecture on cognitive stimuli-based EEG signals. Upon implementation, the results demonstrate without any reasonable doubt that the LSMT is an entirely viable means to interpret and classify EEG signal data.

The results of the best model, the Stacked LSTM with Attention Mechanism proved to provide an accuracy and F1 score of 94% while maintaining a comparatively small number of parameters (116 thousand). This is not only a testament to the capability of the LSTM architecture, but also to the remarkable power of the attention mechanism that has so heavily impacted the domain of deep learning.

The methodology applied provides a solid foundation for the justification of the focus of this research. The testing of numerous different layer sizes allowed for a more comprehensive representation of the capabilities of each of the tested models. The consideration of the F1 score and comparison to the number of parameters is appropriate for the evaluation of each model given the context in which these models would ideally be deployed. However, a more thorough analysis of the distribution of class-based performance and insight into potential biases within the model or data might have provided a more thorough study and analysis of these architectures.

Through the implementation of the methodology described in this paper, the capability of the LSTM architecture has been clearly demonstrated. The LSTM architecture, in all trials, has shown its capability to successfully classify EEG signal data to a high degree of accuracy. By further augmenting the LSTM architecture, an even stronger classifier can be designed which is even more adept at deciphering the complexities of EEG data. The results of this experimental analysis establishes the feasibility of the LSTM architecture as a candidate for developing BCI technology for a wide array of use cases.

References

- [1] Hossain, K. M., Islam, Md. A., Hossain, S., Nijholt, A., & Ahad, M. A. (2023). Status of deep learning for EEG-based brain-computer interface applications. *Frontiers in Computational Neuroscience*, 16. <https://doi.org/10.3389/fncom.2022.1006763>
- [2] Jaipriya, D., & Sriharipriya, K. C. (2022). A comparative analysis of masking empirical mode decomposition and a neural network with feed-forward and back propagation along with masking empirical mode decomposition to improve the classification performance for a reliable brain-computer interface. *Frontiers in Computational Neuroscience*, 16. <https://doi.org/10.3389/fncom.2022.1010770>
- [3] Alzahab, N. A., Apollonio, L., Di Iorio, A., Alshalak, M., Iarlori, S., Ferracuti, F., Moneriù, A., & Porcaro, C. (2021). Hybrid deep learning (hdl)-based brain-computer interface (BCI) systems: A systematic review. *Brain Sciences*, 11(1), 75. <https://doi.org/10.3390/brainsci11010075>
- [4] Mansoor, A., Usman, M. W., Jamil, N., & Naeem, M. A. (2020). Deep Learning Algorithm for Brain-Computer interface. *Scientific Programming*, 2020, 1–12. <https://doi.org/10.1155/2020/5762149>
- [5] Ahn, M., Jun, S. C., Yeom, H. G., & Cho, H. (2022). Editorial: Deep learning in Brain-Computer interface. *Frontiers in Human Neuroscience*, 16. <https://doi.org/10.3389/fnhum.2022.927567>
- [6] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [7] Hochreiter, S., & Schmidhuber, J. (1997a). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [8] Kang, Q., Chen, E. J., Li, Z.-C., Luo, H.-B., & Liu, Y. (2023). Attention-based LSTM predictive model for the attitude and position of shield machine in tunneling. *Underground Space*, 13, 335–350. <https://doi.org/10.1016/j.undsp.2023.05.006>
- [9] Ingolfsson, T. M., Hersche, M., Wang, X., Kobayashi, N., Cavigelli, L., & Benini, L. (2020). EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces. 2020

-
- IEEE International Conference on Systems, Man, and Cybernetics (SMC). <https://doi.org/10.1109/smc42975.2020.9283028>
- [10] Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of Deep Learning. *Neurocomputing*, 452, 48–62. <https://doi.org/10.1016/j.neucom.2021.03.091>
- [11] Lieskovská, E., Jakubec, M., Jarina, R., & Chmulík, M. (2021). A review on speech emotion recognition using Deep Learning and attention mechanism. *Electronics*, 10(10), 1163. <https://doi.org/10.3390/electronics10101163>
- [12] Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*, 337, 325–338. <https://doi.org/10.1016/j.neucom.2019.01.078>
- [13] Lieskovská, E., Jakubec, M., Jarina, R., & Chmulík, M. (2021). A review on speech emotion recognition using Deep Learning and attention mechanism. *Electronics*, 10(10), 1163. <https://doi.org/10.3390/electronics10101163>
- [14] Li, W., Qi, F., Tang, M., & Yu, Z. (2020). Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing*, 387, 63–77. <https://doi.org/10.1016/j.neucom.2020.01.006>
- [15] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Muller, K.-R. (2021). Explaining Deep Neural Networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278. <https://doi.org/10.1109/jproc.2021.3060483>
- [16] Sarkar, P. S., Mathew, J., Stephen, C., Kanaga, E. G. M., & Bhuvanewari, M. (n.d.). An Experimental Analysis of Various Deep Learning Architectures for the Classification of Cognitive Stimuli based EEG Signals. *GRENZE International Journal of Engineering and Technology*, 10(1), 250–255.
- [17] Bran, A., & Vaidis, D. C. (2020). On the characteristics of the cognitive dissonance state: Exploration within the pleasure arousal dominance model. *Psychologica Belgica*, 60(1), 86–102. <https://doi.org/10.5334/pb.517>