ConVnet BiLSTM for ASD Classification on EEG Brain Signal

https://doi.org/10.3991/ijoe.v18i11.30415

N. A. Ali[1], A. R. Syafeeza(✉)[1], A. S. Jaafar[1], Norazlin Kamal Nor[2]
[1] Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2] Hospital Canselor Tuanku Mukhriz, Kuala Lumpur, Malaysia
syafeeza@utem.edu.my

Abstract—As a neurodevelopmental disability, Autism Spectrum Disorder (ASD) is classified as a continuum disorder. The availability of an automated technology system to classify the ASD trait would have a significant impact on paediatricians, as it would assist them in diagnosing ASD in children using a quantifiable method. In this paper, we propose a novel autism diagnosis method that is based on a hybrid of the deep learning algorithms. This hybrid consists of a convolutional neural network (ConVnet) architecture that merges two LSTM blocks (BiLSTM) with the other direction of propagation to classify the output state on the brain signal data from electroencephalogram (EEG) on individuals; typically development (TD) and autism (ASD) obtained from the Simon Foundation Autism Research Initiative (SFARI) database to classify the output state. For a 70:30 data distribution, an accuracy of 97.7 percent was achieved. Proposed methods outperformed the current state-of-the art in terms of autism classification efficiency and have the potential to make a significant contribution to neuroscience research, as demonstrated by the results.

Keywords—state-of-art architecture, autism, classification, brain signal

1 Introduction

Developmental disability Autism Spectrum Disorder (ASD), is defined by its behavioural symptoms in children. In most cases, the symptoms of an ASD appear in the first year of life. The first signs of autism are a lack of social interaction and a lack of language development in early childhood. In Malaysia and around the world, the number of children with ASD continues to rise. According to a Malaysian survey, one in every 625 children has ASD [1].

ASD affects one in every 68 children in the United States and one in every 100 children in the United Kingdom, according to the Centers for Disease Control and Prevention (CDC) [2]. A clinical specialist or paediatrician may take a long time to confirm an ASD diagnosis. According to a recorded paediatrics study by [3], a two-year-old autistic child is still unable to produce two meaningful words that do not need to be imitated or repeated. Research into this disorder has been extensive, however, we have yet to discover what lies beneath it. Autistic Spectrum Disorders are hard to

diagnose for the simple reason that the ASD traits are not passed down through the generations. The ASD is defined as a pattern of repetitive, unusual, and abnormal behaviour. Therefore, early detection prior to the age of three is extremely unlikely [4]. ASD can be mistaken for a variety of other neurodevelopmental symptoms. Because of the lack of obvious differences in the primary neural conditions, it may be difficult to identify this atypical behaviour.

The Electroencephalogram (EEG), a representation of the state of the brain's electromagnetic signals, can be used to measure brain activity. According to prior research by [5][6], autistic children's brain waves become less complex and synchronised as their frequency increases. The EEG signals of autistic and non-autistic brains show a correlation [7]–[11]. Previous research in the field [12], [13] scientists in fields such as neuroscience could benefit from the higher temporal resolution, lower complexity, lower cost and acceptable technique of EEG recording over functional Magnetic Resonance Imaging (fMRI). The EEG dataset is more accessible than the Magnetoencephalogram (MEG) dataset because of the greater availability of EEG data [14] [15].

In the last few years, deep learning (DL) models for EEG analysis have become increasingly common [16]–[20]. Different from traditional machine learning methods, which use a variety of feature extraction and classification techniques, DL takes an intelligent and holistic approach to feature extraction and classification [21]–[23]. In several applications, it has been demonstrated that ASD classification can be significantly improved using deep learning neural networks rather than traditional methods of signal processing. To classify ASD, we use four state-of-the-art models based on specific brain regions in this work. As a result of our proposed ConVnet BiLSTM method's superior performance over other deep learning approaches, clinically relevant subgroups of ASD patients were identified using the vectorized classification outputs of the ConVnetBiLSTM classifier.

This paper organizational structure is summarized in the following paragraphs. Autism classification methodology is discussed in Section 2 of this document. Detailed explanations of the findings and conclusions appear in Section 3. The final section of the report will summarise the findings and make recommendations for future research.

2 Methodology

2.1 Dataset

SFARI, more specifically the Simons Searchlight initiative, which was formerly known as the Simons Variation in Individuals Project, provided the data for this analysis (SVIP - Simons VIP consortium, 2012). The Simons Searchlight datasets, which include 16p11.2 copy number variant-CNV data, are made available to researchers who have passed the data request process and have been approved. A healthy control is one who has been identified as having a typically developing (TD) personality and is therefore considered to be a normal participant, free of any signs of autism. Detailed descrip-

tions of the Simons Searchlight study's participant identification, recruitment, and inclusion/exclusion criteria have been provided in the study description within [24], [25]. Until recently, the CNV region's deletion or duplication was a condition of eligibility. Exclusion criteria included the presence of any known genetic syndromes or infective CNV. There were no Simon's VIP (SVIP) evaluations performed on any of the study's TD subjects. The comparison group was drawn from the Boston Children's Hospital participant database on its own [26]. The group was made up of people without any kind of neurological or developmental disorder. An ethical review board from UTeM's Centre for Research & Innovation Management (CRiM) has approved the request to obtain access to SFARI's EEG data. A researcher distribution agreement and all the necessary paperwork had been completed, and it was approved. The SFARI gave the SVIP its initial ethical green light for operation. An approved researcher was given permission to collect and share data from SVIP participants.

Prior to the previous study [26], SVIP collaborators collected EEG data and performed some pre-processing. NetAmps 300 amplifier and NetStation at 0.3-30Hz Bandpass filter were used for the 128 channel HydroCel Net (Electrical Geodesics Inc., Eugene, OR) at 500 Hz sampling rate. While the EEG was being recorded, participants sat back and watched silent videos on a monitor for 2–12 minutes. During the experiment, infants were placed on the laps of their primary caregivers. Pre-processing was carried out after obtaining the data. Each participant was disqualified from further analysis if more than 10% of their data points fell outside of the predefined range of -150 V to 150 V. There were eleven channels that were not used for analysis or as input for interpolation because of this criterion. It was decided to conduct further research on the remaining 117 channels. The channel map of the EEG device placed on participant scalp as shown in Figure 1. The grey colored indicates the deleted channels electrode that were excluded from analysis.

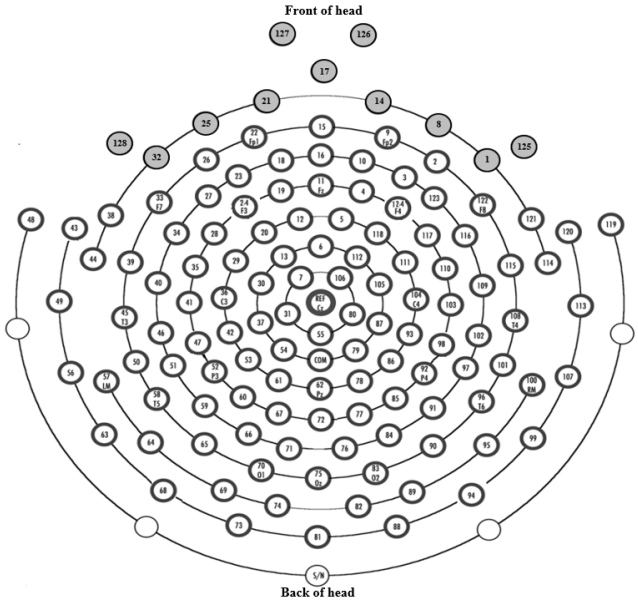

Fig. 1. The hydrocel geodesic sensor net-128 channel map

Initially, there were fifty-seven data points (TD=16, ASD=41) that were collected. Two individuals were excluded from the TD and ASD classes because they had not participated or had their EEGs recorded, and new data from 53 participants (TD = 14 controls, ASD = 39) were obtained. Summary of the database is shown in Table 1. Even though there is a disparity in the number of participant in these two classes, the augmentation of the signal samples is done in order to balance these two classes.

Table 1. Summary of the EEG database

Genotype Case	Gender	Categorical			Total
		Infant	***Toddler***	***Children***	
Typically development (TD)	Male-6 Female-8	- -	3 5	3 3	14
Autism	Male-22 Female-17	1 -	12 10	9 7	39

2.2 Data preparation

It is necessary to segment and enhance the signals that have been gathered. Figure 2 depicts the data preparation process flow diagram.

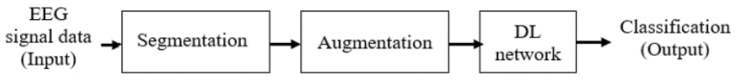

Fig. 2. Flow process on preparation of data

In this work, the extraction process from a .csv file begins with segmentation [27]. It was then saved as a Matlab .mat file. Splitting up data into two or more subsets and grouping them based on the parameters that have specified, allows to analyse, and use the information gathered more effectively. It's possible to label the data after segmentation, using a similar group that was assigned to the removed channel as discussed earlier. Here is the pseudocode for the labelling of the deleted channel, as shown.

```
Data labelling <dataset>
1) Loading the data load SfariData
2) Assign array b for new b={}
3) The data taken from initial to max subject for (i =
   1:53)
4) Assign old signal to all subject a=Signals (i : )
5) Delete 11 channels for all subject
   a([1,8,14,17,21,25,32,125,126,127,128], ; ) = [ ]
6) Update new signal b (i) =a;
7) end for
8) Compute new assigned signal b signals = b'
9) Accomplish saving of removed channels for all
subjects save. Autismchnldelete Signals Labels
```

Segmentation also transmitted the signal along with their assigned labels of TD and autism (ASD). Despite the dataset's apparent imbalance, data augmentation can restore the sample signal balance for both data training and data testing. It appears that most researchers [28][29] use this data distribution of data splitting and this is being tested and proven in their research. Figure 3 shows the segmented data being distributed for 70:30 data distribution. Primary data distribution between typically development (TD and autism (ASD) is shown in Table 2. While Figure 4 depicts the allocation of EEG data used to train and test the algorithm proposed for further augmentation.

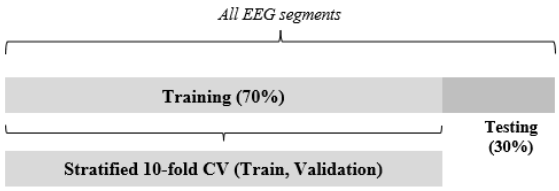

Fig. 3. Stratified CV method on data distribution of 70:30

Table 2. Primary data distribution of signal sample between 2 classes

	70% Data Train	**30% Data Test**
Typically Development (TD)	367	157
Autism (ASD)	1096	470

The stratified of 10-fold cross validation is used for 70% of the original EEG segment signal. The type of k-fold cross-validation uses fold selection to ensure that the mean response value is approximately the same across all folds; this ensures that roughly equal proportions of TD and ASD class labels are present across all folds. As a result, the train and validation accuracy have been determined to be accurate. Then, a new network with the updated weight will be created for the remaining 30% of data to be tested on. The training of the model is the first step. The model was validated and the parameters were adjusted. It is only at this point that an unbiased estimate of the model's performance can be made. Figure 4 depicts the process of data augmentation between these categories; typically developmental disorders (TD) and autism spectrum disorders (ASD).

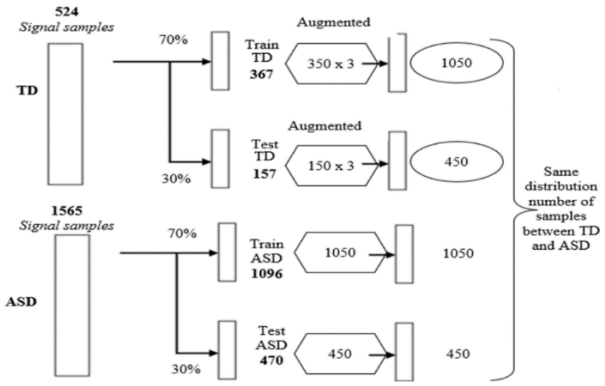

Fig. 4. Data augmention of segmented sampled data

Increase the amount of data by adding slightly modified copies of previously existing information or by creating new synthetic information from previously existing information through the process of data augmentation [27]. There were originally 53 subjects and 128 channels in the original data set. Removed the 11 channels, there were 117 channels. There were 524 samples of typical development and 1565 samples of autism (ASD) after segmentation. TD and ASD data were out of balance, necessitating the use of data augmentation.

2.3 Proposed DL method

An integrated CNN-BiLSTM model and hyperparameter grid search are shown in Figure 5 of this proposed work. Hyperparameters are variables that have the power to influence how well a system learns. Similar to machine learning algorithms may have different types of learning rates, weights, and constrains for generalizing data patterns. These measures are called hyperparameters. It is necessary to tune hyperparameters so that the problem can be solved optimally. The process of optimization involves finding a tuple that provides an optimal model and minimize loss function. There are several approaches to hyperparameter tuning. For this study work relied on the method of grid

search, which consists of exhaustive searching through a subset of hyperparameter space of the algorithm, followed by a performance metric.

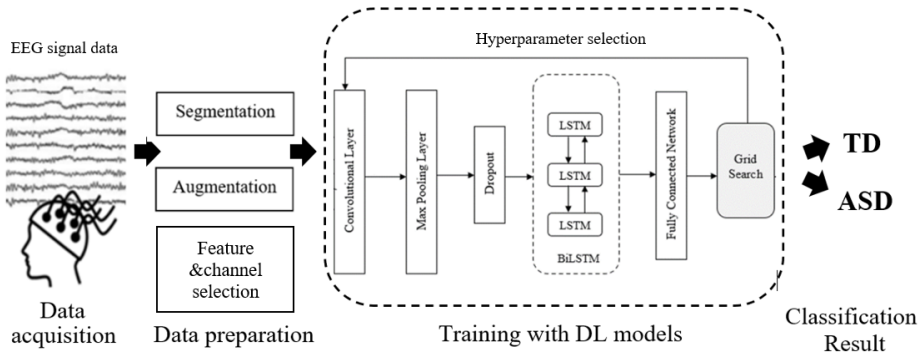

Fig. 5. The proposed method structure diagram

Image recognition has made extensive use of convolutional neural networks (CNNs), but little work has been done on how these networks might be applied to physiological signal. The CNN also referred to as covnets [28], are neural networks that may have common parameters. A covnet is made up of multiple layers, each of which has a differentiable function that allows it to transform one volume into another. A convolutional layer, a pooled layer, and an all-connected layer are all part of the system's architecture. Filters (kernels) slide across the EEG signal in the convolutional layer. Input EEG signals are convolved with a kernel matrix, and stride controls the amount of filtering that occurs. The feature map is another term for the convolution's output. A down-sampling layer, is another name for this pooling layer. The convolutional layer's output neurons are shrunk by the pooling operation, which lowers the computational load and prevents overfitting. The fully connected layer, on the other hand, has complete access to all the activations in the layer prior to that layer.

A Long-Short-Term Memory is what the term LSTM refers to. In the context of an artificial intelligence (AI), it is from a deep learning family system. Recurrent Neural Network (RNN) model issues can be addressed by using LSTM mechanism of gates and memory cells to replace the updates of hidden layers. Its networks can acquire dependencies of a long-term sequences data between the interval of time [29]. Inside a single LSTM block as shown in Figure 6 consists of three gates called input, forget and output gates.

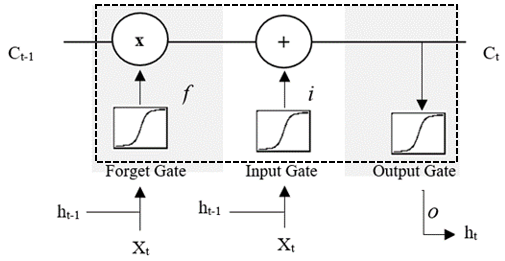

Fig. 6. LSTM internal block diagram [29]

Using a sequence of inputs, the input gate determines whether or not the block's current state is altered. By keeping track of information output and deciding whether the current hidden state should be moved to the next restatement, LSTM blocks use forget gates to hold or delete previous conditions. In order to determine if a hidden state can be restated, the output gate monitors the information output. A sigmoid activation function is used by these three gates; it generates values between 'zero' and 'one,' with a value of '0' indicating a closed gate that passes no data and a value of '1' indicating the opposite. Ct is cell memory, Xt is the input vector while ht is the cell output. During the training phase, both bias (b) and weight matrix (w) are updated. The LSTM gates equation are expressed as:

$$i_t = \sigma(w_i[h_{t-1}, x_t]) + b_i \tag{1}$$

$$f_t = \sigma\big(w_f[h_{t-1}, x_t]\big) + b_f \tag{2}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t]) + b_o \tag{3}$$

where i_t, f_t and o_t denotes the gate's input, forget and the output. Only previous data is collected by the LSTM network, and no data from the future is used. Instead of only transmitting information from front to back, the bidirectional-LSTM (BiLSTM) network also computes the output value and the output value is computed simultaneously in two separate blocks. With the pre-and-post dependencies in the EEG data, this deep model can achieve excellent results. Figure 7 demonstrates a deep network of bidirectional-LSTM network is designed as in previous work [30]. The ($h_t\rightarrow$) in the forward layer portrays the first LSTM block with front-to-back passed information, while the ($h_t\leftarrow$) represents the backward layer of final LSTM block with back-to-front information transferred. There are 117 channels of EEG data in the input layer, and the output layer Yt specifies the last sequence. The LSTM block sequence is shown in the rectangular box with a dotted line in both forward and reverse directions. Specifically, this BiLSTM network with a hidden layer of 100 was specified, as well as the output of the final element in the sequence.

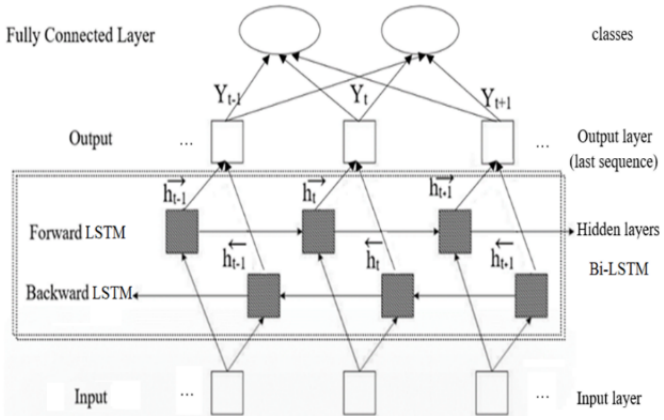

Fig. 7. Bidirectional-LSTM network structure

The proposed CNN-BiLSTM (ConVnet BiLSTM) architecture, which is a combination of CNN and BiLSTM, includes an initial convolutional layer that is responsible for receiving EEG signal input. Given that the brain signal from the dataset indicates that each channel is a two-dimensional (2D) array, the input size has been specified as size of two. The output of ConVnet should then be passed to the maxpooling layer. The number of output neurons is reduced by using the max-pooling operation, which only uses the highest value in each feature map. In the next stage, the Dropout stage which is a training technique in which neurons selected at random are ignored. On the forward pass, the neurons temporarily stop contributing to activating the cells below them by dropping-out randomly. There are no more weight updates for the neuron when it goes backwards. In order to prevent overfitting, the dropout technique is used. Hidden units' outgoing edges are randomly assigned zero values during each training update. Neurons are the building blocks of these hidden units. Finally, the bidirectional LSTM layer receives an output that has been pooled to a smaller dimension. It is assumed that the convolutional later will extract local features, such as that the BiLSTM will order features to understand the signal input's order.

This network's architecture makes it possible to store data for a long period of time. It is also capable of processing data in both directions, making full use of data that has come before and after it. This proposed network model will map the input data into 100 features, and then generate output for the fully connected layer. The fully connected layer, on the other hand, has complete access to all of the activations from the layer before it. Final layer employs softmax activation function to get probabilities of input being in a specific class after passing through fully connected layers (classification). Finally, a fully connected layer of size 2, connected by a classification and softmax layers as its activation function, was used to define two classes labelled as normal and autism. Figure 8 show the entire process in action.

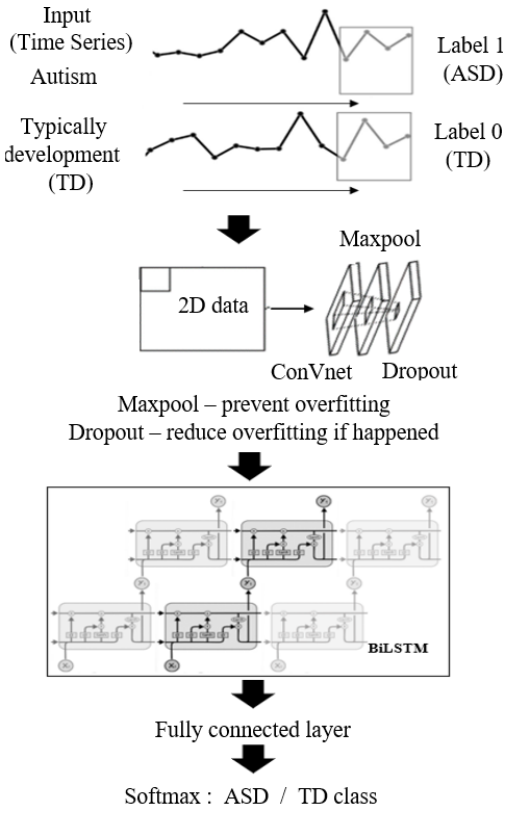

Fig. 8. The proposed method for autism classification

Figure 9 represents the flow testing methodology with other state-of-the-art Deep Learning models. The first step is data exploration (using the SFARI dataset), followed by data transformation, and finally by the division of the data into training and testing sets. Five parameter tests are being used to evaluate four state-of-the-art deep learning algorithms. Based on its highest degree of accuracy, the classification can finally be determined.

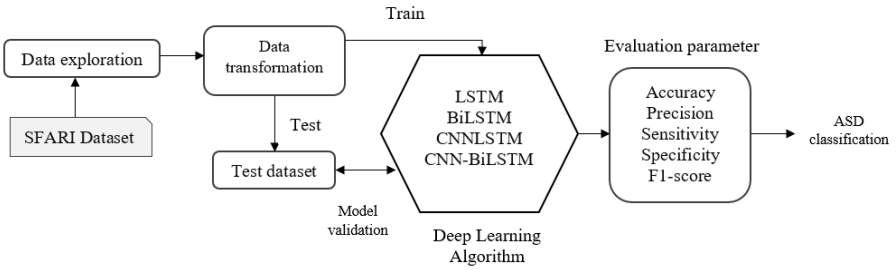

Fig. 9. Methodology flow of the 4 state-of-art models

2.4 Performance measure

An efficient classification system is one that is highly accurate, sensitive, and precise in terms of a few key performance indicators. Classification problems make use of confusion matrices to show the results of a classifier on a set of data that has been identified as having the true values. While the Predicted Class serves as a reference for what the signal actually is, the Actual Class represents what the network decided it should be. These cases are divided into two categories: those considered 'Typically development' for the normal case, and those considered to have 'Autism'. The classification performance is visualized in confusion matrix.

To outline the performance of implemented DL algorithms, seven parameters namely accuracy (Acc), precision (Pre), sensitivity (Sen), specificity (Spe), F-measure (F1-Score), Matthew Correlation Coefficient (MCC) and Area Under Curve (AUC) are utilized for classifying ASD. Eq. (4-10) represents the expressions for performance parameter.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Pre = \frac{TP}{TP+FP} \tag{5}$$

$$Sen = \frac{TP}{TP+FN} \tag{6}$$

$$Spe = \frac{TN}{TN+FP} \tag{7}$$

$$F1-Score = \frac{2TP}{2TP+FP+FN} \tag{8}$$

$$MCC = \frac{TN \times TP - FN \times FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{9}$$

$$AUC = \frac{\left(1+\frac{TP}{TP+FN}-\frac{FP}{FP+TN}\right)}{2} \tag{10}$$

3 Result and discussion

This section discusses the performance of the proposed system that has been simulated. the systems runs on MATLAB 2020b environment on an Intel(R) Core(TM) i7-7700HQ CPU with 2.8GHz clock speed and 20GB RAM. Figure 10 shows the extracted data brain signal of a single person in both the TD and ASD classes after the segmentation and augmentation. For a total of five thousands signal samples, the brain signals of autistic people are dispersed irregularly while those signal of TD child are in regular pattern.

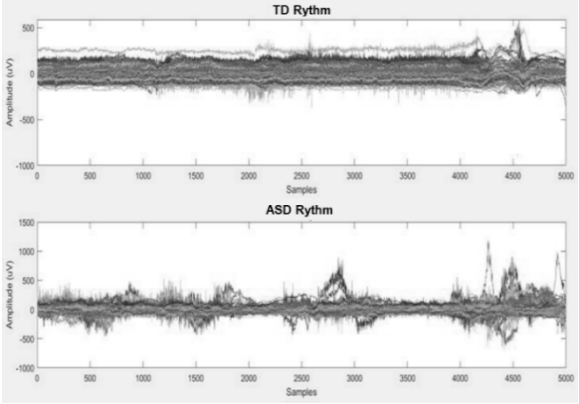

Fig. 10. The plotted figure represent brain rythms between TD and ASD

3.1 Fine-tuning hyperparameter

Grid searching is one of the technique involving parameter tuning and is used throughout deep learning to find the best model parameters, and it is not limited to a single type of model. This type of searching builds a model based on every possible combination of parameters, resulting in multiple iterations of the process. The ten-fold of stratified cross-validation and 70:30 data distribution is used for the whole process. There were several hyperparameters that needed to be adjusted in the hidden layer to get the best results which involved batch size, epochs, optimization algorithm, dropout regularization and learning rates neurons in the hidden layer, respectively. Table 3 show the best tuned-parameter based on the highest accuracy. Table 3 yielded a list of optimum hyperparameters, and these parameters were used in the development of the deep learning model of Convnets BiLSTM involved and listed in table are the best number of each parameter based on the highest accuracy obtained. Using a grid search technique for fine-tuning parameters, we were able to determine the best values for each category. As for the data distribution, Table 4 show the accuracy obtained for three types of data distribution against train and test data.

Table 3. Fine-tuning hyperparameter

Parameter	Number of parameters	Accuracy (%)
Batch size	50	88.6
Epoch	20	89.3
Dropout value	0.5	92.58
Learning rate	0.01	91.6
Optimization algorithm	Adam	89.5

Table 4. Number of data distribution

Data distribution	90:10	80:20	70:30
Accuracy (%)	98.53	99.15	99.48

3.2 Deep Learning model development

As the hybrid model of combination CNN and BiLSTM (ConVnet BiLSTM) is the first ASD classification using a 117 channel of EEG time series data that employs neural network model, there are no recognized comparison partners. Thus it can be compared with some widely used single LSTM, single BiLSTM and hybrid CNN combined with LSTM (CNNLSTM) classification methods. LSTM is an artificial RNN used in the field of deep learning that has feedback connection which can process not only single data point but also the entire sequences of data.

In other hand, the Bidirectional LSTM also known as BiLSTM is a sequence processing model that consists of two LSTM blocks in both directions (forward and backward) and also effectively increase the amount of information available to network, improving the context available to the algorithm. A combination of CNN LSTM is an LSTM architecture specifically designed for sequence prediction problem with spatial input signal such EEG signal.

Both single and hybrid form of BiLSTM has stronger ability of generalization in comparison with other existing deep learning algorithms. Firstly, data set divided in the form of labeled data with 70:30 portion of data splitting, the input to these models were two-dimension data (the signal of samples, labelled signal). Then cross validation parameter search to tune hyper parameter and made feature work. By contrast, the ConVnet BiLSTM model outperformed the other three models as shown in Table 5 and Figure 11.

Table 5. Result of the deep learning models

Performance Measure	Single model		Hybrid Model	
	LSTM	***BiLSTM***	***CNNLSTM***	***ConVnet BiLSTM***
Acc	88.04	94.84	91.01	97.72
Pre	0.79	0.85	0.81	0.89
Sen	0.65	0.66	0.59	0.85
Spe	0.94	0.96	0.95	0.96
F1-Score	0.71	0.75	0.69	0.87
MCC	0.63	0.68	0.61	0.83
AUC	0.92	0.94	0.91	0.97
Computational Time	134.64	132.70	140.57	136.72

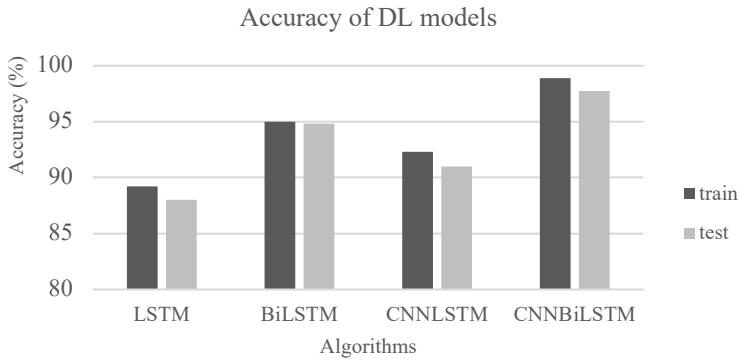

Fig. 11. Accuracy of different classification models

Based on plotted boxplot in Figure 12, the distribution accuracy on tenfold stratified cross validation for the different deep learning models. The middle line indicates the distribution between the typical development and autism classes. These plots have been made to the same scale. Seems the CNNBiLSTM or so called ConVnet BiLSTM achieved highest accuracy among other three state-of-art deep learning models. Table 6 show the previous work that using the hybrid of CNN bidirectional-LSTM on different approaches such classification, detection, recognition, prediction and etc.

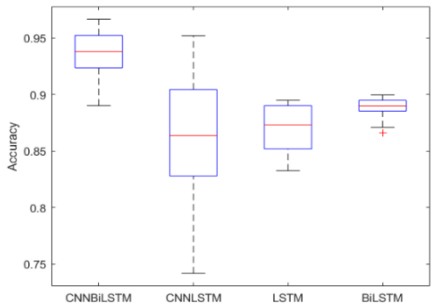

Fig. 12. Boxplot for four different models versus accuracy

Table 6. CNNBiLSTM based on past research works

Authors	Case Study	Highest accuracy
Sakirin Tam et al., 2021	Twitter sentiment classification	91%
W-L Mao et al., 2020	Seizure detection	92.6%
Amit Rai, 2020	Solar radiation prediction	94.23%
Yimin Hou et al., 2020	Human motor recognition	94.6%
Hao Zhen et al., 2020	Wind power forecast	95%
Ji-Hoon Jeong et al., 2020	Brain-controlled robotic arm	96.3%
Our proposed algorithm	Autism Classification on EEG	97.72%

4 Conclusions

Based on a more expensive image FMRI dataset, earlier work by AbdulRazak [36] used a convolutional neural network strategy to categorize ASD patients with non-ASDs. The primary objective of this study is to use the deep learning algorithm to classify ASD based on the rapid, inexpensive, and non-invasive measurable way. The LSTM-based blocks of deep learning model was offered in this study as a novel technique to diagnosing autism. Segmentation and augmentation of EEG data from the SFARI database were used to create a sample signal that was evenly distributed between the two groups. Nonlinear and non-stationary difficulties in the brain signal data were well-suited to this method.

This study presented a deep learning ConVnet BilSTM algorithm network constructed for both EEG classification; autism and normally development (obtained from healthy control participants), and the results were spectacular. Over 97.7 percent accuracy was achieved for 70:30 stratified cross validation of data distribution, which showed that the approach proposed has an efficient classification capability and performance improvements over a single model and hybrid CNN LSTM method. This study is currently essential and potentially give a significant contribution to neuroscience research. When the highest levels of precision are attained, this will allow pediatricians or clinicians to employ a measurable approach to diagnose ASD in children rather than relying entirely on behavioral observation of the suspected autistic child (according to the standard manual used by pediatricians on DSM-V). A decent rule of thumb is to use statistical hypothesis testing with expert analysis as well. To categorize ASD and non-ASD, deep learning is the preferable method in this study.

5 Acknowledgement

Authors acknowledge the support from the Centre for Telecommunication Research and Innovation (CeTRI), Universiti Teknikal Malaysia Melaka (UTeM). This research was supported by UTEM and the Ministry of Higher Education Malaysia (MOHE) through a fundamental research grant (FRGS/2018/FKEKK-CERIA/F00363). Authors are grateful and sincerely appreciate the access given to phenotypic and EEG data on SFARI Base from Simon VIP Foundation.

6 References

[1] N. Mohd Salleh, N. Mat Noor, and J. Samsudin, "A Survey of Knowledge of Autism Spectrum Disorder Among Malaysia Polytechnic Communities," *Int. J. Stud. Child. Women, Elder. Disabl.*, vol. 5, pp. 65–70, 2018, [Online]. Available: http://www.ijcwed.com/wp-content/uploads/2018/07/IJCWED5_25.pdf

[2] E. Jon Baio, "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network , 11 Sites , United States , 2010," *MMWR Surveill. Summ. - Centres Dis. Control Prev.*, vol. 63, no. 2, pp. 1–21, 2014.

[3] C. A. D. Kevin P. Brady, "Navigating the Current Legal Landscape of the Identification and Eligibility of Children with Autism Spectrum Disorder," *Leg. Front. Educ. Complex Law Issues Leaders, Policymakers Policy Implementers*, pp. 119–145, 2015. https://doi.org/10.1108/S1479-366020150000024046

[4] M. A. A. Y. F. M Dai, A. W. Nei, A. A. Samad, "Management of Autism Spectrum Disorder In Children And Adolescents," *Clin. Pract. Guidel.*, pp. 1–49, 2015.

[5] J. Kang, H. Chen, X. Li, and X. Li, "EEG entropy analysis in autistic children," *J. Clin. Neurosci.*, vol. 62, no. xxxx, pp. 199–206, 2019. https://doi.org/10.1016/j.jocn.2018.11.027

[6] M. Ahmadlou, H. Adeli, and A. Adeli, "Fuzzy Synchronization Likelihood-wavelet methodology for diagnosis of autism spectrum disorder," *J. Neurosci. Methods*, vol. 211, no. 2, pp. 203–209, 2012. https://doi.org/10.1016/j.jneumeth.2012.08.020

[7] D. Abdolzadegan, M. H. Moattar, and M. Ghoshuni, "A robust method for early diagnosis of autism spectrum disorder from EEG signals based on feature selection and DBSCAN method," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 482–493, 2020. https://doi.org/10.1016/j.bbe.2020.01.008

[8] W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, "EEG complexity as a biomarker for autism spectrum disorder risk," *BMC Med.*, vol. 9, 2011. https://doi.org/10.1186/1741-7015-9-18

[9] S. Schwartz, R. Kessler, T. Gaughan, and A. W. Buckley, "Electroencephalogram Coherence Patterns in Autism: An Updated Review," *Pediatr. Neurol.*, vol. 67, pp. 7–22, 2017. https://doi.org/10.1016/j.pediatrneurol.2016.10.018

[10] L. Billeci *et al.*, "On the application of Quantitative EEG for characterizing autistic brain: A systematic review," *Front. Hum. Neurosci.*, vol. 7, no. JUL, pp. 1–15, 2013. https://doi.org/10.3389/fnhum.2013.00442

[11] T. Liu, Y. Chen, D. Chen, C. Li, Y. Qiu, and J. Wang, "Altered electroencephalogram complexity in autistic children shown by the multiscale entropy approach," *Neuroreport*, vol. 28, no. 3, pp. 169–173, 2017. https://doi.org/10.1097/WNR.0000000000000724

[12] B. A. Cociu *et al.*, "Multimodal functional and structural brain connectivity analysis in autism: A preliminary integrated approach with EEG, fMRI, and DTI," *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 2, pp. 213–226, 2018. https://doi.org/10.1109/TCDS.2017.2680408

[13] F. H. Duffy, A. Shankardass, G. B. McAnulty, and H. Als, "A unique pattern of cortical connectivity characterizes patients with attention deficit disorders: A large electroencephalographic coherence study," *BMC Med.*, vol. 15, no. 1, pp. 1–19, 2017. https://doi.org/10.1186/s12916-017-0805-9

[14] A. Puce and M. S. Hämäläinen, "A review of issues related to data acquisition and analysis in EEG/MEG studies," *Brain Sci.*, vol. 7, no. 6, 2017. https://doi.org/10.3390/brainsci7060058

[15] E. Grossi, C. Olivieri, and M. Buscema, "Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study," *Comput. Methods Programs Biomed.*, vol. 142, pp. 73–79, 2017. https://doi.org/10.1016/j.cmpb.2017.02.002

[16] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "EEG classification of driver mental states by deep learning," *Cogn. Neurodyn.*, vol. 12, no. 6, pp. 597–606, 2018. https://doi.org/10.1007/s11571-018-9496-y

[17] A. Craik, "Deep learning for EEG classification tasks: A review," *J. Neural Eng.*, pp. 0–12, 2019.

[18] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, 2018. https://doi.org/10.1016/j.compbiomed.2017.09.017

[19] J. H. Jeong, K. H. Shim, D. J. Kim, and S. W. Lee, "Brain-Controlled Robotic Arm System Based on Multi-Directional CNN-BiLSTM Network Using EEG Signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1226–1238, 2020. https://doi.org/10.1109/TNSRE.2020.2981659

[20] A. Antoniades *et al.*, "Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2285–2294, 2017. https://doi.org/10.1109/TNSRE.2017.2755770

[21] N. Varuna Shree and T. N. R. Kumar, "Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network," *Brain Informatics*, vol. 5, no. 1, pp. 23–30, 2018. https://doi.org/10.1007/s40708-017-0075-5

[22] B. Tang *et al.*, "DeepChart: Combining deep convolutional networks and deep belief networks in chart classification," *Signal Processing*, vol. 124, pp. 156–161, 2016. https://doi.org/10.1016/j.sigpro.2015.09.027

[23] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Syst. with Appl. X*, vol. 7, p. 100033, 2020. https://doi.org/10.1016/j.eswax.2020.100033

[24] J. Jenkins *et al.*, "Auditory Evoked M100 Response Latency is Delayed in Children with 16p11.2 Deletion but not 16p11.2 Duplication," *Cereb. Cortex*, vol. 26, no. 5, pp. 1957–1964, 2016. https://doi.org/10.1093/cercor/bhv008

[25] J. J. Leblanc and C. A. Nelson, "Deletion and duplication of 16p11.2 are associated with opposing effects on visual evoked potential amplitude," *Mol. Autism*, vol. 7, no. 1, pp. 1–7, 2016. https://doi.org/10.1186/s13229-016-0095-7

[26] R. Al-Jawahiri, M. Jones, and E. Milne, "Atypical neural variability in carriers of 16p11.2 copy number variants," *Autism Res.*, vol. 12, no. 9, pp. 1322–1333, 2019. https://doi.org/10.1002/aur.2166

[27] Iwana BK.; Uchida S., "An empirical survey of data augmentation for time series classification with neural networks," *PLoS One*, vol. 16, no. 7, pp. 1–32, 2021. https://doi.org/10.1371/journal.pone.0254841

[28] I. Priyadarshini and C. Cotton, "A novel LSTM – CNN – grid search - based deep neural network for sentiment analysis," *J. Supercomput.*, no. 0123456789, 2021. https://doi.org/10.1007/s11227-021-03838-w

[29] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, and Q. Yuan, "Scalp EEG classification using deep Bi-LSTM network for seizure detection," *Comput. Biol. Med.*, vol. 124, p. 103919, 2020. https://doi.org/10.1016/j.compbiomed.2020.103919

[30] N. A. Ali, A. R. Syafeeza, A. S. Jaafar, S. Shamsuddin, and N. K. Nor, "LSTM-based Electroencephalogram Classification on Autism Spectrum Disorder," *Int. J. Integr. Eng.*, vol. 13, no. 6, pp. 321–329, 2021. Retrieved from https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/8165

[31] S. Tam, R. Ben Said, and Ö. Tanriöver, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," *IEEE Access*, vol. 9, pp. 41283–41293, 2021. https://doi.org/10.1109/ACCESS.2021.3064830

[32] W. L. Mao, H. I. K. Fathurrahman, Y. Lee, and T. W. Chang, "EEG dataset classification using CNN method," *J. Phys. Conf. Ser.*, vol. 1456, no. 1, 2020. https://doi.org/10.1088/1742-6596/1456/1/012017

[33] A. Rai, "A CNN-BiLSTM based deep learning model for mid-term solar radiation prediction," *Int Trans Electr Energ Syst*, no. April, pp. 1–13, 2020. https://doi.org/10.1002/2050-7038.12664

[34] Y. Hou *et al.*, "Deep Feature Mining via Attention-based BiLSTM-GCN for Human Motor Imagery," *J. Latex Cl. File*, vol. 14, no. 8, pp. 1–8, 2020.

[35] H. Zhen, D. Niu, M. Yu, K. Wang, Y. Liang, and X. Xu, "A hybrid deep learning model and comparison for wind power forecasting considering temporal-spatial feature extraction," *Sustain. MDPI*, vol. 12, no. 22, pp. 1–23, 2020. https://doi.org/10.3390/su12229490

[36] A. Y. Saleh and L. H. Chern, "Autism Spectrum Disorder Classification Using Deep Learning," *Int. J. online Biomed. Eng.*, vol. 17, no. 08, pp. 103–114, 2021. https://doi.org/10.3991/ijoe.v17i08.24603

7 Authors

Nur Alisa Ali is currently a lecturer in Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia. She is in the stage of writing her thesis. Her research interest is in biomedical computer vision and artificial intelligent.

Assoc Prof Dr Syafeeza Ahmad Radzi received her B. Eng degree (2003) and M. Eng (2005) in Electrical – Electronic & Telecommunication Engineering from Universiti Teknologi Malaysia (UTM) Johor, Malaysia. She also received her Ph.D degree in Electrical Engineering from the same university in 2014. She is currently an associate professor at the Faculty of Electronic Engineering and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia. Her research fields include embedded system, pattern recognition, machine learning, deep learning, image processing, biometric, etc. She is currently supervising this research work and can be contacted at syafeeza@utem.edu.my.

Dr Abd Shukur Jaafar received both first and master degree from Universiti Teknologi Malaysia (UTM) in Bachelor of Electrical Engineering (2002) and Master of Engineering in Electronic and Telecommunication (2005). He joined Universiti Teknikal Malaysia Melaka (UTeM) as a lecturer in 2005 and received PhD in Communication System from Lancaster University, UK. Currently his research interest on RF, Microwave circuits, algorithm development for indoor positioning and navigation and artificial intelligent signal processing.

Assoc Prof Dr Norazlin Kamal Nor is an experienced lecturer with a demonstrated history of working in the higher education industry. Skilled in Developmental Paediatrics, Child Protection and Child Maltreatment Prevention, Genetic Epidemiology, Health Promotion, and Healthcare. Strong educational and research background with a Doctor of Philosophy (PhD) focused in Genetic Epidemiology from Johns Hopkins Bloomberg School of Public Health.

Article submitted 2022-02-25. Resubmitted 2022-04-27. Final acceptance 2022-04-29. Final version published as submitted by the authors.