# Classification of Adults with Autism Spectrum Disorder using Deep Neural Network

Muhammad Faiz Misman, Azurah A. Samah, Farah Aqilah Ezudin, Hairuddin Abu Majid, Zuraini Ali Shah, Haslina Hashim, Muhamad Farhin Harun
Artificial Intelligent and Bioinformatics Group (AIBIG), Department of Information System,
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia,
81310, Johor Bahru, Johor, Malaysia
faizmisman@gmail.com, azurah@utm.my, farahezudin@gmail.com, hairudin@utm.my, aszuraini@utm.my, haslinah@utm.my, farhin1993@gmail.com

***Abstract*—Autism Spectrum Disorder (ASD) is a developmental brain disorder that causes deficits in linguistic, communicative, and cognitive skills as well as social skills. Various application of Machine Learning has been applied apart from the clinical tests available, which has increased the performance in the diagnosis of this disorder. In this study, we applied the Deep Neural Network (DNN) architecture, which has been a popular method in recent years and proved to improve classification accuracy. This study aims to analyse the performance of DNN model in the diagnosis of ASD in terms of classification accuracy by using two datasets of adult ASD screening data. The results are then compared with the previous Machine Learning method by another researcher, which is Support Vector Machine (SVM). The accuracy achieved by the DNN model in the classification of ASD diagnosis is 99.40% on the first dataset and achieved 96.08% on the second dataset. Meanwhile, the SVM model achieved an accuracy of 95.24% and 95.08% using the first and second data, respectively. The results show that ASD cases can be accurately identified by implementing the DNN classification method using ASD adult screening data.**

**Keywords—Classification, Autism Spectrum Disorder, Deep Neural Network, Support Vector Machine component**

## I. Introduction

Autism Spectrum Disorder (ASD) is a developmental brain disorder which can be identified by repetitive, restricted behaviour [1]. The disorder may cause an individual to have deficits in linguistic, communicative and cognitive skills as well as social skills [2]. Statistically, Redfield et al. (2018) in their report published by the Centers for Disease Control and Prevention (2018) estimated that there is about 1 in 59 children identified with ASD, with the number of identified cases is four times higher in males than that of females [3]. The diagnosis of ASD usually starts around the age of two, but in some cases, the diagnosis can be very difficult because the symptoms are only noticeable in older children.

Unfortunately, the cause of this mental disorder is yet to be known by scientists, since there have been differences and causes of ASD, but most are yet to be discovered. However, in most of the cases, there are combinations of genetic risk factors which can interact with environmental risk factors. Due to the overlapping symptoms and complexity of the disorder, which makes each autistic individual unique, it is difficult to examine the exact cause that contributes to the matter [4].

Currently, there is no medical test to analyse the disorder. Therefore, observation of the behavioural characteristics and intellectual properties as brain abnormalities has been the primary diagnosis of ASD, which will be evaluated by the medical experts. However, only a few clinical diagnostic tests have been used for the diagnosis of ASD, such as the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule-Revised (ADOS-R) [5]. These tests were applied to review the basic skills and the behaviour of the child.

Work on ASD diagnosis prediction is considered active research conducted by various researchers using various techniques such as Machine Learning techniques and statistical analysis techniques. According to [5], Decision Tree and Random Forest has produced accuracy level of 83.94% and 91.74%, respectively as these methods were implemented on ASD screening adult dataset for ASD diagnosis classification [6]. Furthermore, Support Vector Machine and K-Nearest Neighbors are 2 classification techniques used in a study [7]. As for the results, SVM has produced an accuracy of 95%, while K-Nearest Neighbors only produced an accuracy of 89% in the classification of ASD diagnosis using ASD adult screening dataset.

In this study, the proposed classification method to classify ASD diagnosis using ASD screening dataset is the Deep Neural Network (DNN). The performance of DNN classifier is then analysed and evaluated to see whether it can produce a better performance compared to the method from the previous study. The performance comparison is evaluated in term of classification accuracy.

## II. Materials And Method

### A. Materials

The data used in this study are two adult ASD screening datasets from two different sources. The first dataset is obtained from a webpage publicised by a Professor of Data Analytics. Meanwhile, the second dataset is from the UCI Machine Learning Repository. Both datasets are published by [8]. Generally, the first dataset is the latest version made of the second dataset, with three additional features summing a total of 24 features with 1118 samples, while the second dataset contains 21 features with 704 samples. Both datasets described ASD screening results, which obtained from a mobile application developed by Dr.Fadi Thabtah, called "ASDQuiz" which is available on both Android and iOS devices.

### B. Data Preparation and Preprocessing

Before the datasets are fitted into the model, they need to be pre-processed. In this study, data pre-processing performed on the datasets include handling of missing values, variable reduction, normalisation, and label encoding. In handling missing values, the datasets are first loaded into WEKA tools

[9] to identify missing values. The missing values in the datasets are then removed, and samples with a huge amount of missing values are dropped. In handling missing values, deleting the entire sample or record for missing entries may not be the best solution for data cleaning; thus, in this experiment, we decided to delete the entire sample for data cleaning. Then, by exploring further, we found that there are a few features in both datasets that are not relevant and insignificant for the experiment. So, the features are dropped, and therefore, the variables were reduced. After removal of missing values, dataset 1 was reduced from 1118 to 1117 samples while dataset 2 was only left with 609 samples after removing about 95 samples with missing values.

After removal of missing value, normalisation was done on both datasets and applied only on features with a numerical value. With features contained text value and categorical value, we converted into a numerical value that is understandable for the model. Finally, the final preparation is to partition the datasets into the training set and test set. Table I below shows the features and their descriptions.

TABLE I. VARIABLES AND THEIR DESCRIPTIONS

| Variables | Description |
|---|---|
| Questions (1-10) | 10 questions regarding behavioural features |
| Age | Age in years |
| Gender | Male or female |
| Ethnicity | List of common ethnicities |
| Jaundice | Whether born with jaundice or not |
| Autism | Whether family members have PDD |
| Residence | List of countries |
| Score | The final score of the screening algorithm |
| Class | ASD/Non-ASD |

### C. Development of Deep Neural Network Model

In this study, we developed a DNN model using Keras Sequential model API [10]. Keras is a high-level neural network API which is written in Python programming language.

DNN is designed based on the neurons in the human brain. It is an artificial neural network (ANN), but with deeper layers. An overview of the architecture of the DNN model is depicted in Fig. 1.

Next, in developing the proposed DNN model, it is important to determine the architecture of the model and determine the activation function that will be used in the development of the model DNN.

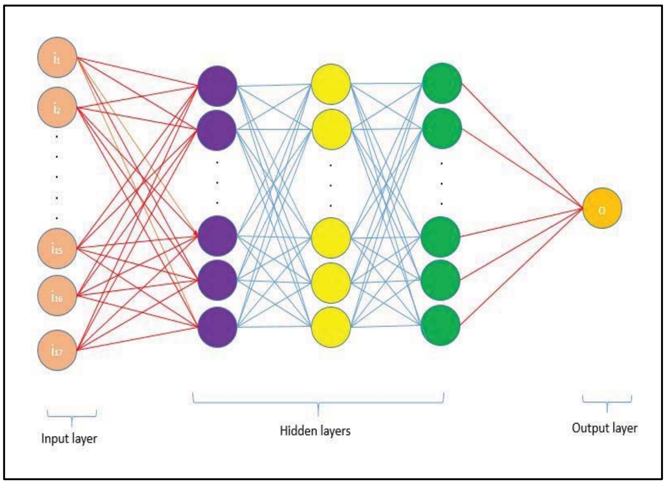



Fig. 1. DNN Model Architecture

There are two main hyperparameters in a neural network, which is the number of layers and the number of nodes. Both hyperparameters need to be tuned to find the model with the best accuracy [11]. Therefore, the number of nodes and the number of hidden layers were set by using an experimental method, where the number of nodes and layers were changed in each experiment with the same datasets, and the best result from the experiments will decide the number of layers as well as the number of nodes in each layer. Table II below shows the best network hyperparameters chosen in this experiment.

TABLE II. DNN ARCHITECTURE HYPERPARAMETERS

| Dataset | Hidden Layers | Input Node | Hidden Node | Output Node |
|---|---|---|---|---|
| 1 | 3 | 15 | 15 | 1 |
| 2 | 2 | 15 | 15 | 1 |

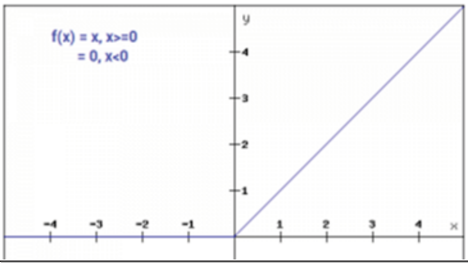
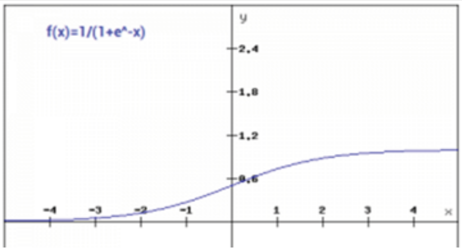
Other than the number of nodes and layers, there are also a few other hyperparameters applied in the model such as the activation function, drop-out rate, optimiser and batch size. These hyperparameters are similarly set for both datasets. These hyperparameters are set based on the best results obtained after each experiment. The following shows the list of hyperparameters setting for the model development.

- Activation function (hidden layers): ReLU
- Activation function (output layer): Sigmoid
- Dropout: 0.2 / 0.4
- Batch size: 10
- Optimiser: Adam
- Loss: binary crossentropy

For the activation function, ReLU is used because it has been the standard activation function for deep learning models. In this study as well, ReLU seemed to provide a good result for the model. Sigmoid, on the other hand, is a very suitable function for output layers because it exists between (0 to 1) and is suitable for binary classification. The following Table III shows the activation functions used in equation (1), where $\hat{Y}$ is the predicted value or output for each layer.

$$\hat{Y} = \text{Activation}\ (\sum\ (\text{weight} * \text{input}) + \text{bias} \qquad (1)$$

TABLE III. GRAPH OF ACTIVATION FUNCTION

| Activation Function | Graph |
|---|---|
| ReLU |   |
| Sigmoid |   |

Other than that, in order for the model to learn, each dataset is partitioned into two sets, which is used for data training and data testing. Dataset 1 is divided into 70:30 percent for all the dataset samples. Dataset 2, on the other hand, is divided into 75:25 percentage. We set a slightly higher percentage for training on dataset 2 because this dataset has a lower number of samples compared to dataset 1, and to avoid model overfitting we have to increase the number of the training set.

### D. *DNN Model Performance Assessment*

After implementing the proposed model of DNN, it is important to evaluate its performance to study whether it can outperform other techniques in classification problem. One of the techniques to evaluate the performance of our classifier model is the confusion matrix, that can show the number of the correct and incorrect prediction made by the classifier model. It is compared with the actual data and gives a brief description of the classification results. Table IV shows the confusion matrix that is used to calculate the performance of the classifier for both training and testing. Accuracy, Sensitivity and Specificity are the terms which are most commonly associated with a Binary classification test as they are capable of delivering statistical measurement regarding the performance of the test [12]. Equation (2), (3), (4) shows how we calculate the accuracy, sensitivity and specificity respectively.

Generally, the accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Meanwhile, the sensitivity of a test is its ability to determine the patient cases correctly. To estimate it, we calculate the proportion of true positive in patient cases. Last but not least, the specificity of a test is its ability to determine the healthy cases correctly. To estimate it, we calculate the proportion of true negative in healthy cases.

TABLE IV. CONFUSION MATRIX

| | True Condition | |
|---|---|---|
| **Predicted Condition** | True Negative (TN) | False Positive (FP) |
| | False Negative (FN) | True Positive (TP) |

$$\text{Accuracy} = ((TN + TP) / (TP + TN + FP + FN)) * 100 \quad (2)$$

$$\text{Sensitivity} = (TP / (TP + FN)) * 100 \quad (3)$$

$$\text{Specificity} = (TN / (TN + FP)) * 100 \quad (4)$$

Other than performance assessment based on the above metrics, the improvement percentage of the model is also calculated. The improvement is calculated by calculating the difference in the accuracy obtained by both DNN and SVM model.

## III. RESULTS

The DNN model is built to classify ASD diagnosis. Training data and validation data are provided and fitted into the model and the performance of the model is analysed and evaluated using the performance calculations as discussed previously. The performance of the model will be explained based on each dataset.

### A. *Performance of DNN Classifier using Dataset 1*

DNN model for the first dataset was run on 350 epochs. The accuracy of the classification model consistently increases over the epochs and continuously scores in a range of 95% to 100% for the last 10 epochs. The loss results, which indicates how well the algorithm models the data also shows a constant decrease for the last 10 epochs, which indicates a good classification performance.

After the results of epochs are obtained, the model assessment is also computed using the confusion matrix to further evaluate the classifier. The evaluation metrics carried out for the evaluation includes the accuracy, sensitivity and specificity. Before the evaluation metrics are computed, the values of predicted vs actual classes are executed. Table V and VI below shows the confusion matrix generated for DNN model using dataset 1.

TABLE V. CONFUSION MATRIX OF DATASET 1

| | | True Condition | |
|---|---|---|---|
| | | Actual: Non-ASD | Actual: ASD |
| **Predicted Condition** | Predicted: Non-ASD | TN = 241 | FP = 0 |
| | Predicted: ASD | FN = 2 | TP = 93 |

TABLE VI. RESULTS OF DNN CLASSIFIER FOR DATASET 1

| **Metrics** | **Accuracy (%)** | **Sensitivity (%)** | **Specificity (%)** |
|---|---|---|---|
| **Result** | 99.40 | 97.89 | 100 |

Based on the results obtained through the confusion matrix, the performance calculation is computed. According to the result computed, it is shown that the result of DNN classifier performance using the first dataset is quite good. This classifier achieved 99.40% of accuracy in classifying adult with ASD. Meanwhile, the sensitivity value obtained is 97.89% and specificity value of 100%. The model is able to produce balanced results between the three metrics.

To be able to see the overall performance, a graph of model accuracy and a graph of model loss is plotted using matplotlib library. The graphs are as shown in Fig. 2 below.

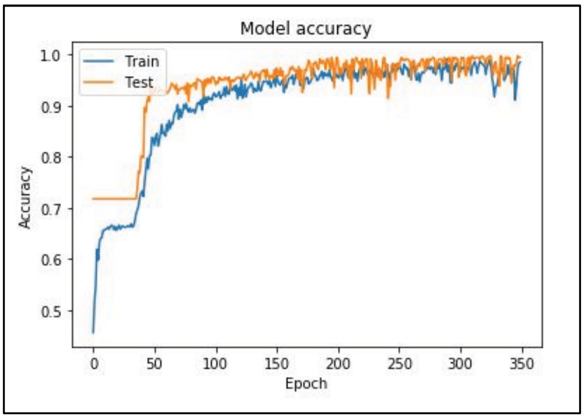



Fig. 2. Graph of train accuracy and test accuracy of proposed classifier (DNN) for Dataset 1of a figure caption.

Based on Fig. 2 above, the observed gap between the train and test accuracy indicates the amount of overfitting of the model. The smaller the gap between the two, the smaller the chance of overfitting. In this case, it can be seen that the gaps between the train and test accuracy decreased across the epochs. However, by the end of the last epoch, the gaps no longer exist. Thus, it can be concluded that the model does not overfit. However, the model accuracy graph also shows that there is a constant increase in accuracy between the training and testing set, starting from 200th epochs. The epochs should be reduced, but it may cause overfitting of the model. Therefore, epochs of 350 are set since it produces the best result of the model. In this study, the dropout parameter is set for each layer. By using the dropout hyperparameter, the chance of overfitting is minimized.

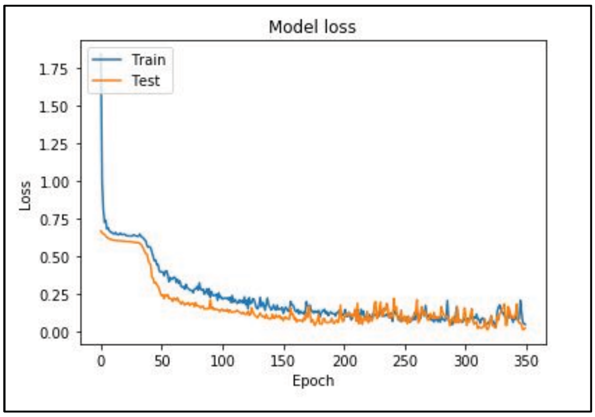



Fig. 3. Graph of train loss and test loss of proposed classifier (DNN) for Dataset 1

Other than plotting the model accuracy graph to analyse the model performance, the model loss graph can also give some insights into the model performance. The graph of loss over epochs shows the learning rate of the model. The lower the loss value, which is close to a zero value, the better the classification model. Based on Fig. 3 above, the loss value constantly decreases over epochs and shows a good learning rate by the model.

### *B. Performance of DNN Classifier using Dataset 2*

DNN model for the second dataset was run on 150 epochs. Just like dataset 1, the accuracy of the classification model is consistently increasing over the epochs and continuously scores in a range of 95% to 100% for the last 10 epochs. The loss results also show a constant decrease for the last 10 epochs, which indicates a good classification performance.

As mentioned previously, the model assessment is also computed using confusion matrix to further evaluate the classifier in terms of accuracy, sensitivity and specificity. Before the evaluation metrics are computed, the values of predicted against actual classes are executed. Table VII below shows the confusion matrix generated for DNN model using dataset 2.

TABLE VII. CONFUSION MATRIX OF DATASET 2

| | | **True Condition** | |
|---|---|---|---|
| | | Actual: Non-ASD | Actual: ASD |
| **Predicted Condition** | Predicted: Non-ASD | TN = 109 | FP = 3 |
| | Predicted: ASD | FN = 3 | TP = 38 |

Based on the results obtained through the confusion matrix, the performance calculation is computed. Table VIII below shows the results of accuracy, sensitivity and specificity using confusion matrix. The results are computed based on the equations that have been stated previously.

TABLE VIII. RESULTS OF DNN CLASSIFIER FOR DATASET 2

| **Metrics** | **Accuracy (%)** | **Sensitivity (%)** | **Specificity (%)** |
|---|---|---|---|
| **Result** | 96.08 | 92.68 | 97.32 |

According to the result computed, it is shown that the result of DNN classifier performance is quite good. The classifier achieved 96.08% of accuracy in classifying adult with ASD. Meanwhile, the sensitivity value obtained is 92.68% and specificity value of 97.32%. This shows that this model has quite a low true positive rate (TPR) in classifying actual cases of ASD, but higher ability in classifying actual cases of non-ASD.

Again, in order to be able to see the overall performance, a graph of model accuracy and model loss on the training and validation datasets over training epochs is plotted using matplotlib library. From the graph, the model can be evaluated whether it suffers overfitting or not. The graphs are as shown in the Fig. 4 below.

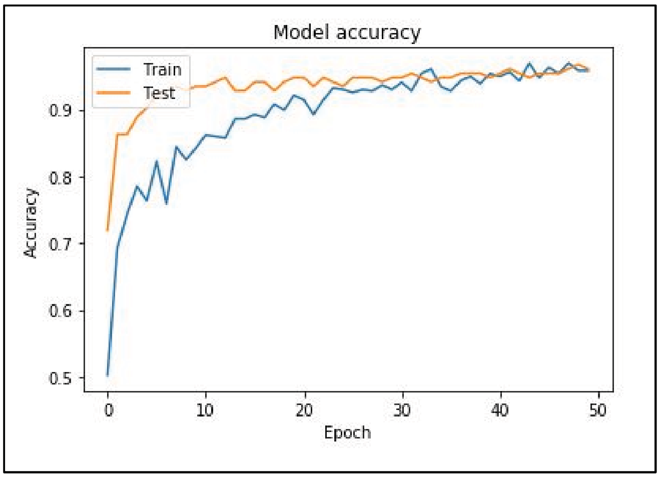



Fig. 4. Graph of train loss and test loss of proposed classifier (DNN) for Dataset 2

Based on Fig. 4 above, the gaps between the train and test accuracy decreased across the epochs. By the end of the last epoch, the gaps no longer exist. Thus, it can be concluded that the model does not overfit. The graph also depicts that the number of epochs set for this model using the second dataset is enough for the training process to take place since there is no constant accuracy increment is found. The dropout parameter is also set for each layer in the model created for this dataset to minimise chances of model overfitting.

Next, model loss graph is plotted to give a better insight of the model performance. As mentioned in the previous subchapter, the graph of loss over epochs indicates the learning rate of the model. The lower the loss value, which is close to a zero value, the better the classification model. Based on Fig. 5, the loss value constantly decreases over epochs and are closer to zero, which shows a good learning rate by the model. However, there is a little gap at the end of the epoch, but it is very close to each other, which would be fine.

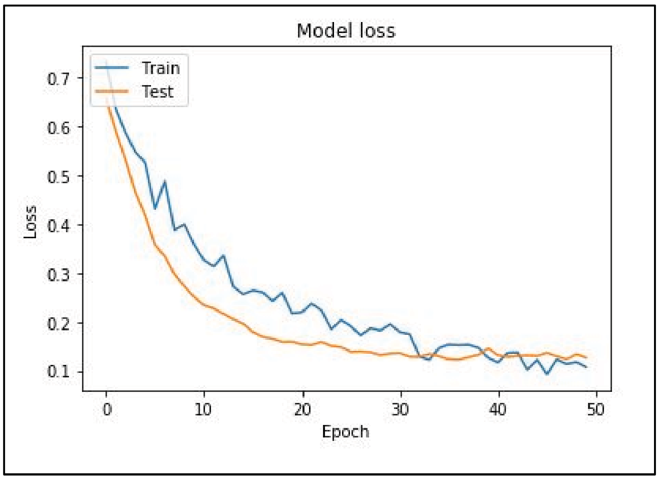



Fig. 5. Graph of train loss and test loss of proposed classifier (DNN) for Dataset 2

As a conclusion for the assessment of DNN model for both datasets, the overall performance of the DNN model is summarized in the form of the bar chart. The bar chart in Fig. 6 below depicts the accuracy of DNN model on both dataset 1 and dataset 2.

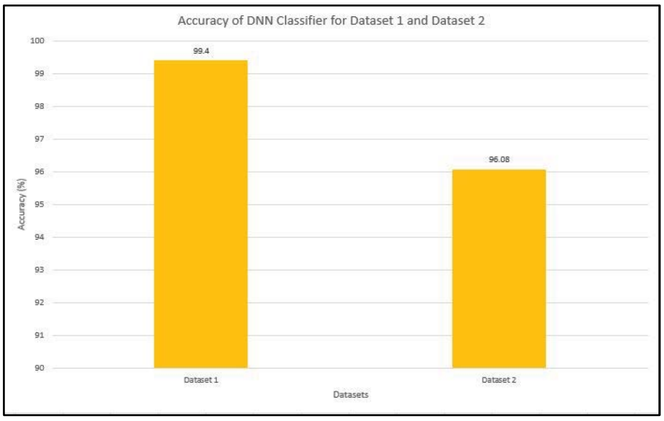



Fig. 6. Performance of DNN classifier for Dataset 1 and Dataset 2

Based on the bar chart above, it can be concluded that DNN classifier worked well for both datasets. Dataset 1, which consist of larger samples compared to that of Dataset 2, has achieved an accuracy of 99.40%.

Meanwhile, Dataset 2 also achieved a good classification accuracy of 96.08%. DNN model performance is better using Dataset 1 may be due to its larger samples, which supplies more training dataset for the model to learn. The more the training samples, the better the model learns to recognise the patterns in the data.

### *C. Discussion*

In order to validate the performance of the proposed classifier, the results obtained from the implementation of DNN model for both datasets were compared with previous studies by another researcher. In this study, the performance of DNN model is compared with a machine learning classifier, namely Support Vector Machine (SVM). The kernel used for the model is the Radial Basis Function (RBF), as performed by [7] in his research work. Fig. 7 shows the comparison of performance between DNN classifier and SVM classifier.

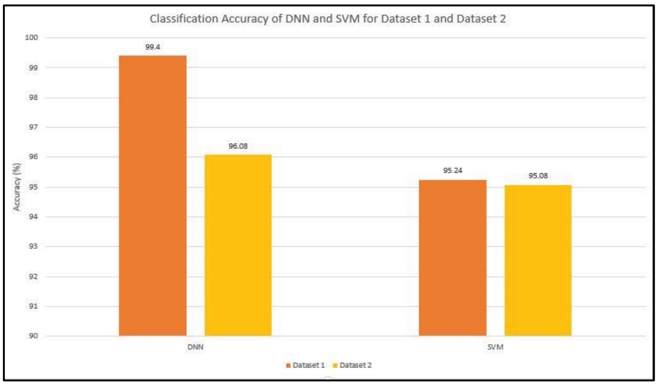



Fig. 7. Comparison of accuracy between DNN and SVM on Dataset 1 and Dataset 2

According to the illustrations in Fig. 7, it can be seen that DNN classifier has clearly outperformed SVM in terms of accuracy. DNN model using Dataset 1 performs better than by using Dataset 2. As mentioned before, the outcome is may be due to the large samples of Dataset 1 compared to Dataset 2. This finding corresponds to the fact that DNN is a suitable classifier for larger datasets. This is because the larger the samples of a dataset, the more training data the dataset can provide to the model to learn and recognise the pattern of the data. However, the samples in Dataset 1 is still not that sufficient enough to claim that the DNN model is optimum. In order to build a more accurate and robust model, a larger dataset is needed.

Finally, to further evaluate the model performance, the improvement percentage of DNN model is calculated. The improvement calculation is done by comparing the difference in results of DNN model and SVM model run in this study. Table IX below shows the improvement percentage of accuracy between the two models. It is shown that DNN model has improved the accuracy in classifying ASD diagnosis by 4.16% on the first dataset and improved 1.00% on Dataset 2.

Accuracy of DNN for dataset 1 (99.4%) and dataset 2 (96.08%) are different because of the different characteristics of the dataset and the different number of samples.

TABLE IX. IMPROVEMENT PERCENTAGE OF DNN MODEL

| **Datasets** | **Accuracy of Classifiers (%)** | | |
|---|---|---|---|
| | **DNN** | **SVM** | **Improvement** |
| **Dataset 1** | 99.4 | 95.24 | 4.16 |
| **Dataset 2** | 96.08 | 95.08 | 1.00 |

## IV. CONCLUSION

The main goal of this study is to implement a supervised deep-learning algorithm using DNN with ASD screening dataset, with the hope to produce a model that can accurately classify ASD cases and non-ASD cases.

Before the model is developed, the datasets underwent a few pre-processing steps. One of them is eliminating missing values. The first dataset contained a lot of missing values in the variable named "why taken the screening" that causes the data to have 1094 missing values. Therefore, since this variable has no significance towards the classification of ASD, it is dropped so that the data can be well-utilised. Meanwhile, for the second dataset, 95 rows of data were dropped due to missing values, and that has reduced the number of instances in the dataset quite a lot. Data pre-processing was challenging since both datasets mostly consist of categorical variables and only two numerical variables. However, by using tools such as the scikit-learn, data pre-processing is done smoothly.

The datasets used in this study was published in December 2017. Therefore, there are only a small number of work that has been done using these datasets. With that, this proposed model is hoped to be able to become a benchmark for the future work related to these datasets.

Overall, the proposed DNN model is well-developed and has produced a good classification result in classifying ASD with the given attributes of the patients' behaviour and also medical information. However, it is very advisable to use a larger dataset in order to obtain a better performance of DNN model in order to claim that the model is accurate and robust enough to classify the data without the risk of overfitting. In this study, however, the results of the proposed model are good enough in classifying adults with ASD and have proven to be significant.

## ACKNOWLEDGMENT

The authors would like to express gratitude to the reviewers and editors for helpful suggestions and Universiti Teknologi Malaysia for sponsoring this research by the GUP Research Grant (Grant Number: Q.J130000.2528.19H17) and FRGS Research Grant (Grant Number: Q.J130000.7851.5F110). The research is also sponsored by the School of Computing, Faculty of Engineering, UTM.

## REFERENCES

[1] H. Li, C. A. Erickson, K. C. Dominick, L. J. Lu, X. Guo, and A. A. Minai, "Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method," Front. Neurosci., vol. 11, no. August, pp. 1–19, 2017.

[2] B. Van Den Bekerom, "Using Machine Learning for Detection of Autism Spectrum Disorder," Twente Student Conf. IT, 2017.

[3] R. R. Redfield et al., "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years-Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014 Surveillance Summaries Centers for Disease Control and Prevention MMWR Editorial and Production Staf," MMWR Surveill Summ., vol. 67, no. 6, p. 2, 2018.

[4] M. Dr Ananya Mandal, "Autism Causes," News Med. Life Sci., pp. 1–5, 2012.

[5] Centers for Disease Control and Prevention, "Screening and Diagnosis of Autism Spectrum Disorder for Healthcare Providers | CDC," 2018. [Online]. Available: https://www.cdc.gov/ncbddd/autism/hcp-screening.html. [Accessed: 30-Apr-2019].

[6] B. Mcnamara, C. Lora, D. Yang, F. Flores, and P. Daly, "Machine Learning Classi cation of Adults with Autism Spectrum Disorder," pp. 1–29, 2018.

[7] A. Demirhan, "Performance of Machine Learning Methods in Determining the Autism Spectrum Disorder Cases," Mugla J. Sci. Technol., pp. 79–84, 2018.

[8] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Informatics Heal. Soc. Care*, vol. 00, no. 00, pp. 1–20, 2018.

[9] E. Frank, M.A. Hall, and I.H. Witten. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[10] Gulli, Antonio, and Sujit Pal. *Deep Learning with Keras*. Packt Publishing, 2017.

[11] D. Stathakis, "How many hidden layers and nodes?," *Int. J. Remote Sens.*, vol. 30, no. 8, pp. 2133–2147, Apr. 2009.

[12] Aswathi B.L. "Sensitivity, Specificity, Accuracy and the Relationship between Them." *Life'n Science*, 12 Mar. 2009, [Online]. Available: www.lifenscience.com/bioinformatics/sensitivity-specificity-accuracy-and. [Accessed: 30-Apr-20].