OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

# Assessing the Importance of Eye Position, Pupil Position, and Point of Regard as Relevant Biomarkers for Autism Spectrum Disorder Detection Using Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) Models

Shriya Pawgi
shriyapawgi@gmail.com

**ABSTRACT**

More efficient and precise diagnoses, enabled by identifying important biomarkers, can significantly improve the standard of living of those with ASD, or Autism Spectrum Disorder. This study investigates the comparative performance of Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) models by utilizing different hyperparameter configuration methods across ten trials to determine the relevance of three primary factors that contribute to autism detection: eye position, pupil position, and point of regard. On a dataset of 59 participants, with 547 time-window rows, both architectures achieved accuracy of 88%. However, through subsequent feature ablation experiments, it was observed that the accuracy of the models remained consistent despite the removal of the pupil position and point of regard parameters for both the left and right eye, whereas the accuracy of the model decreased when the eye position variable was removed from the code, ultimately postulating that solely the eye position of the subject is pertinent for effective autism detection.

*Index Terms - Autism Spectrum Disorder (ASD), Biomarker, Deep Neural Network (DNN), Long Short-Term Memory (LSTM)*

## I. INTRODUCTION

Globally, approximately one in 127 people around the world have autism spectrum disorder [1], and those that receive late diagnoses are often plagued by secondary mental health issues as a result [2]. For the past two decades, the number of autism diagnoses has increased by around 300% [3]. Despite the increasing prevalence of autism affecting members of society, there still remain gaps in autism research due to the spread of misinformation and environmental disruptions in investigation [4]. Prior analysis conducted demonstrates that earlier diagnoses are associated with a better quality of life [5], and those diagnosed

later in life reported feelings of isolation and inadequacy, sentiments caused by lacking essential information about themselves and their identities [6]. Late diagnoses can lead to anxiety, depression, and can potentially result in suicidal thoughts, severely impacting education and career opportunities of individuals. This paper develops different machine learning techniques by incorporating two types of machine learning models, Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTMs) and explores the different biomarkers to investigate autism in order to prevent late diagnoses and the tragic outcomes associated with them. Identifiable biomarkers can also help introduce more non-invasive methods for autism detection. Earlier diagnoses, enabled by more efficient methods, will help to mitigate this issue by leading to early intervention therapies that improve individual standards of living [7] and more extensive autism research can help patch the divides in autism research and help raise communal awareness regarding autism spectrum disorder (ASD).

## II. RELATED WORKS

There are many diverse approaches taken to detect ASD, including using neurophysiological signals, neuroimaging, genetic testing, developmental screening questionnaires, and eye-tracking technology. Neurophysiological signals, specifically electroencephalography (EEG) can be utilised to illustrate neural network architecture through electrical impulses, as determined by the network theory and time series analysis findings. EEG can diagnose various brain disorders, and abnormalities found in the EEG are confirmed in ASD, but this comparison alone is insufficient for diagnosis unless more extensive study is done to evaluate specific stimuli that occur in individuals with ASD [8]. Furthermore, ASD is often associated with a brain that is connected atypically. Neuroimaging, such as Magnetic Resonance Imaging (MRIs) can detect volume abnormalities, and evidence of such abnormalities in both gray and white matter are correlated with atypical structural and function connectivity in the brain. However, such atypicalities need to be clarified and must be analysed within appropriate contexts before equating them with an autism diagnosis [9]. ASD has a strong genetic contribution, and genetic testing can help identify the cause in unique autism cases. Chromosomal microarray analysis to detect chromosomal abnormalities can be implemented to assess similarities in ASD diagnoses, but the target audience for the scope of such studies would be very small as this method can not be implemented to a broader audience [10]. Routine development screening in early childhood, particularly with both broadband and autism-specific instruments at specific age milestones can detect autism risk. Specifically, the Modified Checklist for Autism in Toddlers (M-CHAT-R) could be instrumental in detecting autism at an early stage, but would need to be administered by pediatricians, and not parents, to prevent missed red flags due to conflicts of interest. Furthermore, pediatricians must have the most effective screening tools, and the requirement for more sensitive means of detection can lead to an overreliance on clinical surveillance that can lead to unreliable diagnoses [11]. Eye-tracking technology can be used to effectively detect autism non-invasively by identifying key patterns in attention span and eye orientation that can be used as baseline comparisons, especially in younger individuals such as children and infants, as it does not require advanced motor responses or language [12]. The machine learning classification with eye-tracking

technology to accomplish a comprehensive hyper-parameter search using DNN and LSTM blocks. The particular objective of this research was to understand what features are more prominently significant to ASD for a more focused diagnosis. By having a constant baseline in the experiment, the accuracies of the various parameter configurations between the two models became more efficient to compare, and enabled the identification of a potential biomarker for ASD detection.

## III. MATERIALS AND METHODS

### *A. Dataset Description*

The dataset [13] centers on eye-tracking information to determine autism. Each participant in the dataset was given a unique identifying number, and there were a total of 59 participants, and **547 time-window rows**. Eye-tracking data from 21 females and 38 males was recorded, and participant age ranged from 2.7 years to 12.9 years, with the mean age being 7.88 years. Of these participants, 49% were diagnosed with ASD, while 51% were identified as typically developing (TD). This reiterates that the participant sample of this dataset demonstrates a reasonably balanced distribution of gender, age, and diagnostic status.

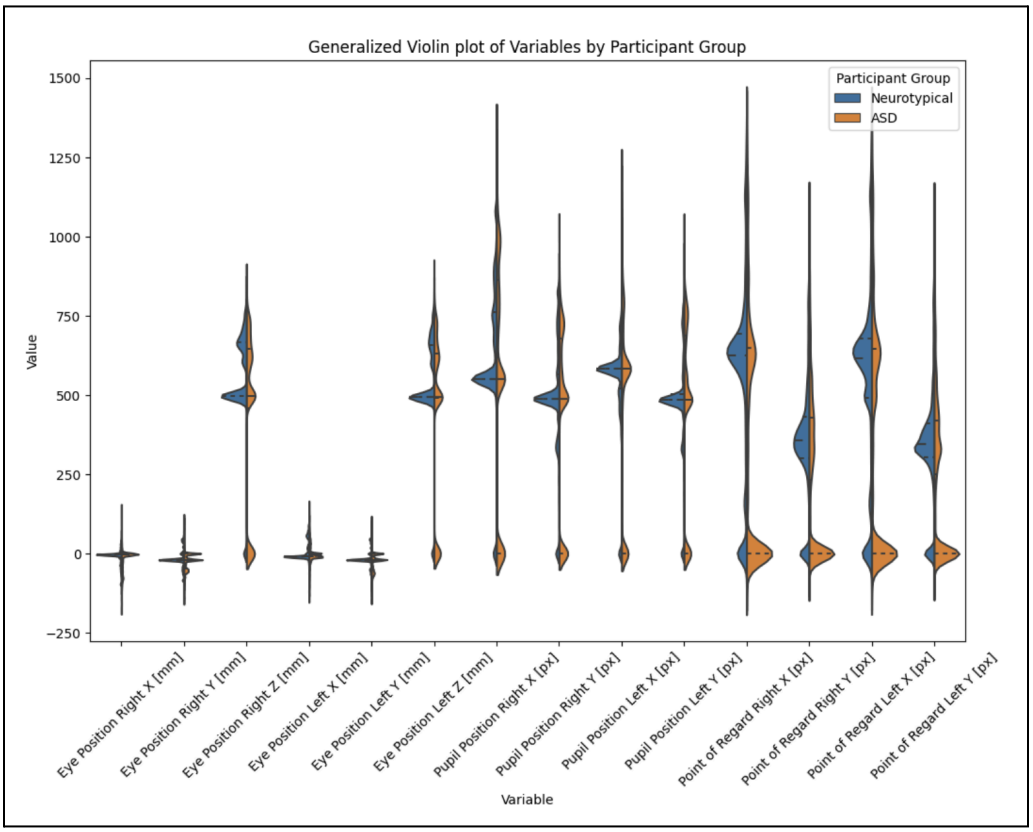


Figure 1 (fig 1); A violin plot of variables displaying two types of participant groups: neurotypical and ASD.

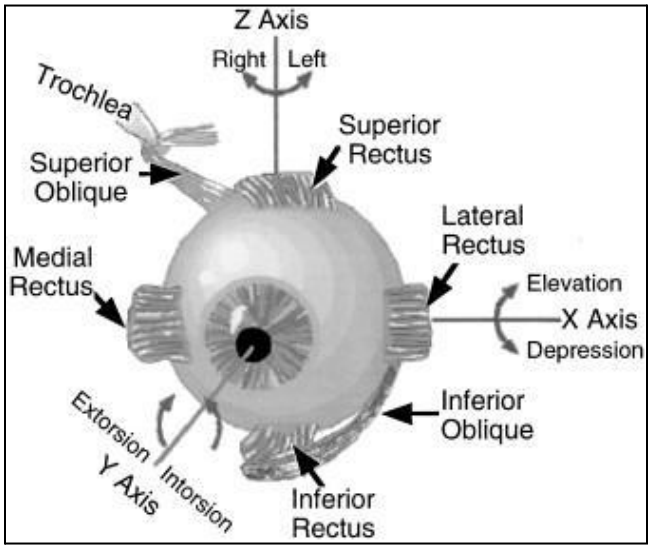



Figure 2 (fig 2); Labeled diagram of the eye with muscles and the three axes of rotation to describe horizontal, vertical, and torsional movements (X, Y, and Z respectively). From *Encyclopedia of the Human Brain*, "Eye Movements" by Charles J. Bruce, Harriet R. Friedman, 2002. https://www.sciencedirect.com/topics/immunology-and-microbiology/eye-position

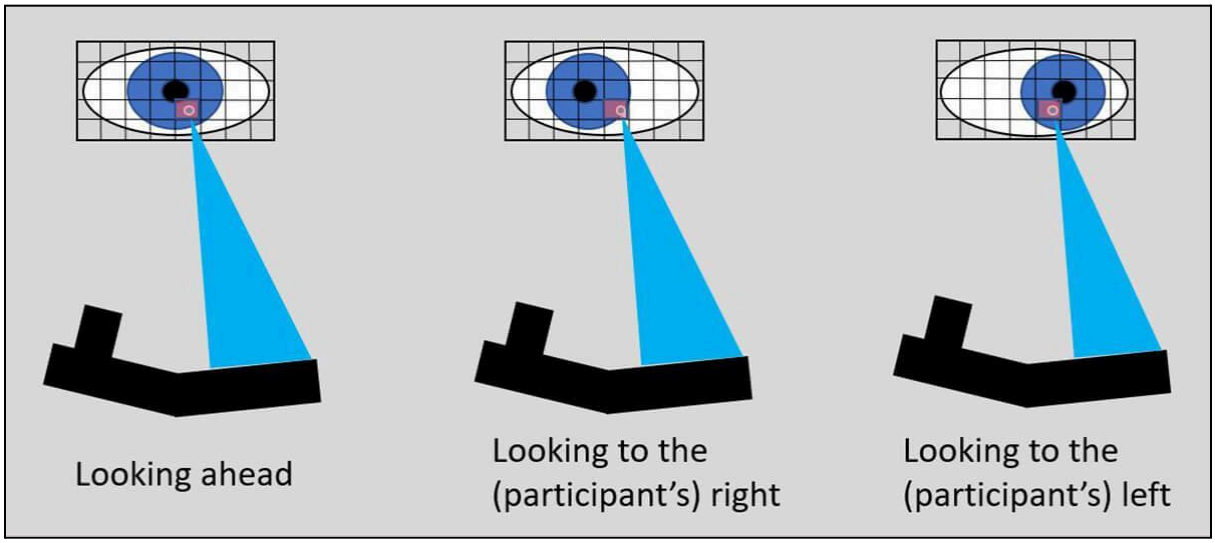


Figure 3 (fig 3); Three different diagrams of the human eye to outline the difference in pupil movements when looking ahead, looking right, and looking left. From *"Fast, Accurate, Reliable Eye Tracking"* by SR Research, EyeLink, 2025. https://www.sr-research.com/about-eye-tracking/

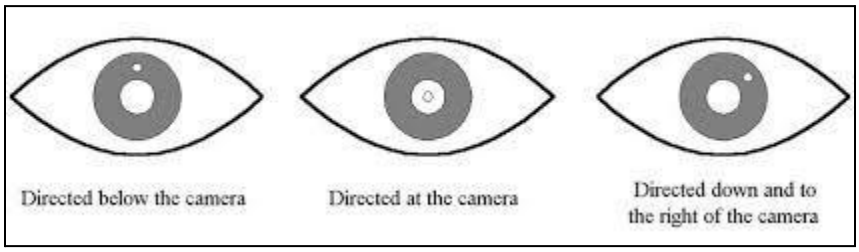



Figure 4 (fig 4); Three different diagrams of the human eye to outline the difference in various point of regard positions in reference to the camera. From Research Gate by Linden J. Ball, 2013. https://www.researchgate.net/figure/Using-the-corneal-reflection-pupil-centre-method-to-determine-point-of-regard-when-the_fig1_322406088

As shown in figure 1, the violin plot illustrates the distribution of several eye-tracking variables. As interpreted from the legend, orange represents ASD participants, and blue represents neurotypical participants. Figures 2, 3, and 4 depict the three primary variables being compared in the violin plot of figure one: eye position, pupil position, and point of regard, respectively. On the x-axis of the violin plot, the first six variables are associated with the physical position of the left and right eyes, and each eye position is described by an X-Y-Z axis. The next six variables are associated with the location of the left and right pupils on a screen, and the plots for the X-and-Y-positions of both pupils appear very similar to each other. However, there is a noticeable difference in the Z-positions of both pupils, as the blue half spreads out vertically, indicating a wider distribution as opposed to its orange counterpart, which is more concentrated and narrow at the lower values. A more noticeable difference can be identified from the point of regard parameters. The X, or horizontal gaze, positions display a broader distribution of blue with greater variation than the orange half, which is more concentrated. Similarly, there is a taller and wider distribution for the blue half, whereas the orange lines are more focused on specific heights. These disparities in the graphs elucidate that within this dataset, individuals with ASD, represented by the color orange, have more restricted and consistent gaze patterns across the measured eye-tracking variables as compared to neurotypical participants.

### *B. Algorithms*

The research utilized the data from figure 1 for training and validating different neural network architectures and AI models, with a particular focus on two families of AI architectures, DNNs and LSTM models. The objective of the research was to determine the capacity of the fundamental neural network, represented by DNNS, and to use LSTMs to determine the specialized version of processing input data, as DNNs and LSTMs are from two different families that process data differently. DNNs are a resource-intensive extension of artificial neural networks, and contain multiple hidden, fully-connected

layers, which enable identification of more sophisticated data patterns. DNNs consist of an input layer, multiple hidden layers, and an output layer, and they use feature hierarchy to learn about the data, meaning that each layer learns features from the previous layer, and these features increase accordingly in complexity; the output of the previous layer becomes the base for the features on which the upper layer trains on [14]. A result is produced from the first layer using data received and an activation function calculation, and is transmitted to the next layer of neurons. How data impacts the next layer's result is determined by the weight of the connections between the corresponding layers [15]. Weights represent the strength of the connections, and are numerical values that determine how much influence the output of one neuron has on another. DNNs are trained iteratively. Data flow, or forward propagation, is the hierarchical learning process previously referenced. A DNN calculates loss by performing comparisons between admin-provided targets and its own predictions, quantifying the difference using an objective of loss function in order to measure the accuracy of the network's predictions. Backpropagation enables errors to be corrected. The network minimized errors by adjusting neuron connection, and a backward flow of information exemplifies learning efficiency. Overfitting occurs when the model memorizes data and does not recognize patterns, and can be identified if the model performs poorly on new data, but well on the training data. This issue can be mitigated by reducing model complexity and data augmentation. Underfitting occurs when the model is not specific enough to capture data structure, and this can be prevented by adding more features and increasing training data [16].

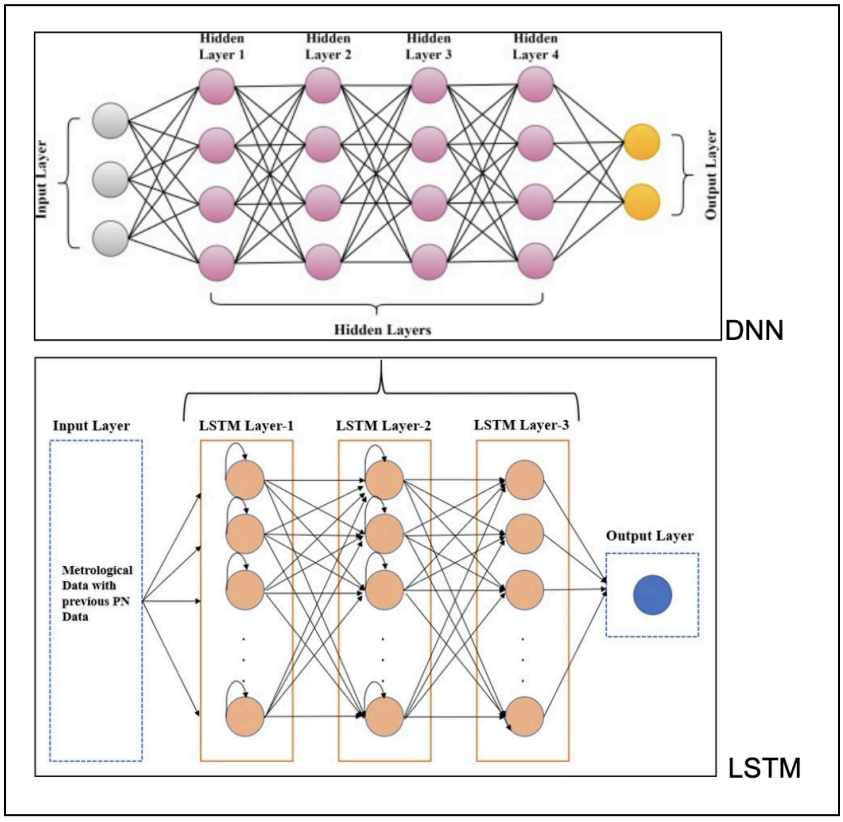


Figure 5 (fig 5); Detailed diagrams describing the main components and data flow in two different types of machine learning models: DNN and LSTMs. Box 1 is from Dutch Harbor Properties by ai dnn, 2026. https://dutchharborproperties.com/?o=80190161041850. Box 2 is from *"Complete Guide to Learn LSTM Models: Types, Applications, and When to Use Which Model"*, Stacademic, Medium, by Poornam Sai G, 2024. https://blog.stackademic.com/complete-guide-to-learn-lstm-models-types-applications-and-when-to-use-which-model-f9b779f31714.

Long Short-Term Memory (LSTM) models are advanced Recurrent Neural Networks (RNNS) with memory, and are referenced when working with sequential data [14]. A LSTM consists of a chain structure that contains memory cells, four neural networks, and three main gates. Information flows through an input gate, which decides which new information is necessary to add. The forget gate determines which data to discard from the previous cell state. The cell state is updated based on the decisions of these two gates, and the output gate determines the output information as a hidden state, which is passed to the next time step. The three gates manipulate memory by selectively retaining and discarding data to enable long-term learning dependencies without gradients that can vanish or explode. Sequential processing enables the memory cell to maintain context across time steps, and can be used for generating network output [17]. Backpropagation Through Time (BPTT) is a method used to train LSTMs. BPTT accounts for the temporal dependency caused by multi-step process sequencing by backpropagating errors across time steps and through network layers. Training incorporates the chain rule to calculate the weight contribution to the general error margin, and product chains are created with each factor representing the ratio of change [18].

## IV. EXPERIMENTAL RESULTS

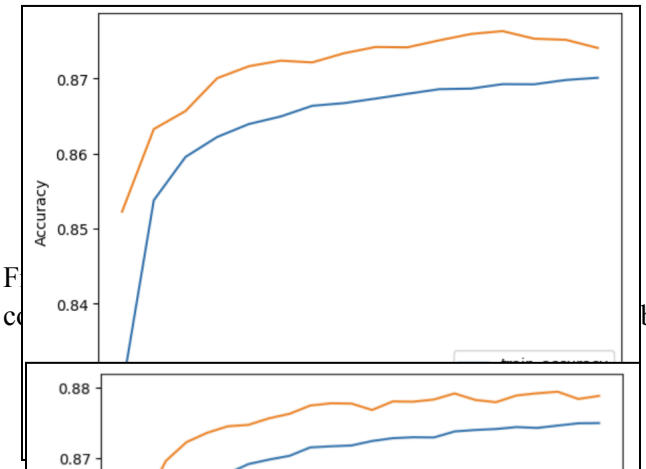
Analysis of the dataset and experimentation was divided into two stages, consisting of multiple trials. In both stages, a DNN and LSTM model were used.

### *A. Stage I*

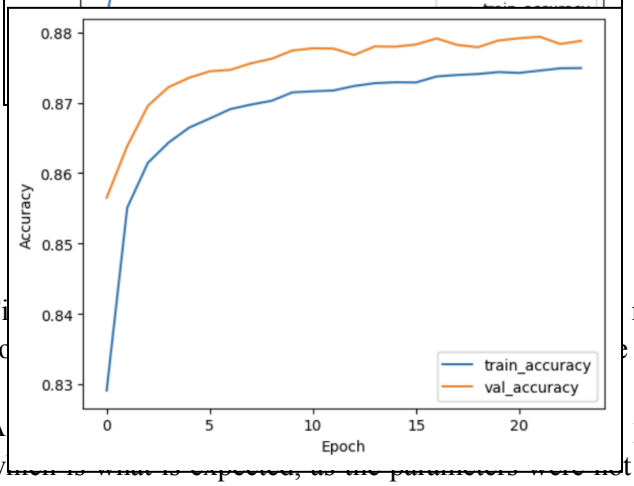
In stage 1, five hyperparameter sets were implemented for each type of model. Both models had a learning rate of 0.001 and were compiled using the Adam optimizer. Training was conducted to a maximum of 50 epochs, and the batch size was 32. To restore the best weights and prevent overfitting, early stopping was implemented with a patience of 3. The 547 data samples were divided into training and validation, with a respective 80% to 20% split with a fixed random seed. Approximately, this yields 438 training samples and 109 validation samples. For consistency, the same split was used across all trials. The split performed at the row level, and not the participant level. Class imbalance was addressed with the application of RandomOverSampler to the training prior to model fitting. To derive a comparative performative analysis of the LSTM and DNN models, it was imperative to determine a baseline, or standard of comparison.

| Model | Number of Layers | Neurons of Layer 1 | Neurons of Layer 2 | Neurons of Layer 3 | Dropout Rate | Accuracy | AUC | Time per Step |
|---|---|---|---|---|---|---|---|---|
| DNN | 3 | 128 | 64 | 2 | 0.2 | 0.88 | 0.0512 | 1 ms/step |
| LSTM | 3 | 128 | 64 | 2 | 0.88 | 0.88 | 0.0480 | 2 ms/step |

Figure 6 (fig 6); Compiled table of the variable configurations for the DNN and LSTM models for Stage 1. Note that the dense and drop-out variables are consistent for both models.




F results of the DNN Model in Stage 1, after the c ble.




Fi results of the LSTM Model in Stage 1, after the c table.

A per figure 8, the accuracy was also 0.88, or 88%, w t changed. However, as shown through figure 6, the LSTM model required 2 ms per step, while the DNN model processed each step in 1 ms, meaning that the DNN was twice as fast per step. The DNN model also had an AUC of 0.95, while the AUC of the LSTM model was 0.95. As the accuracies of both models were the same, the DNN model is at par with the LSTM model, due mostly due to the eye-tracking features incorporating temporal patterns that reduce the sequential processing advantage of the LSTM model.

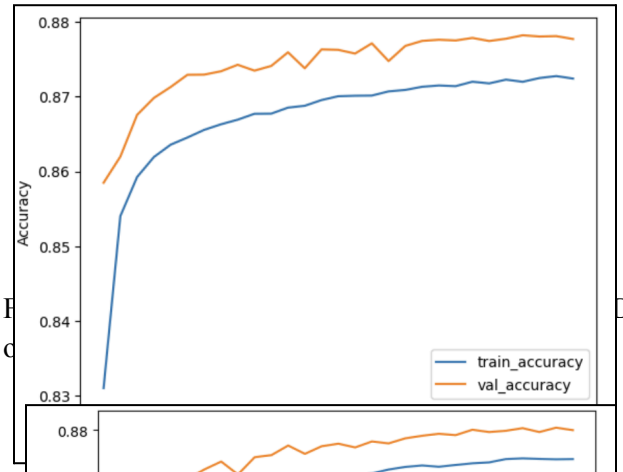
After performing five experiments for each model by altering the Dense numerical quantity, it was determined that for the DNN Model, architectures of a range of 64 to 128 units achieved the consistent

accuracy of 88%, whereas for the LSTM Model, architectures of a range of 128 to 256 with a Dense of 64 output layers achieved a consistent accuracy of 88%. This corroborates that below the maximum threshold of 32 units, neither model is adequate for ASD detection.

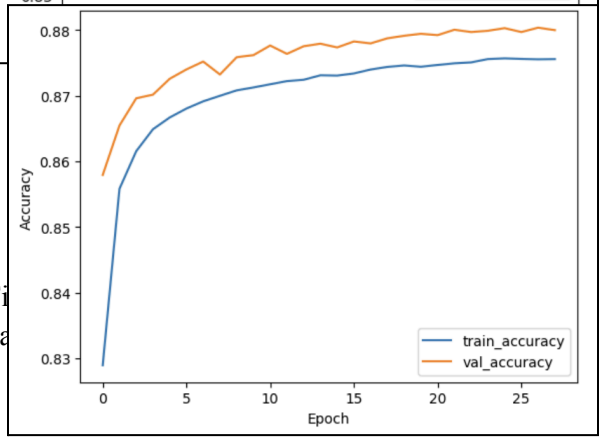
*B. Stage II*

Stage 2 of experimentation detailed hyperparameter configurations via the elimination and isolation of eye position, pupil position, and point of regard variables. This was accomplished by removing the appropriate variable from every segment of code that detailed the columns of interest. A thorough analysis of the features was then conducted to determine the relevance of the variables.

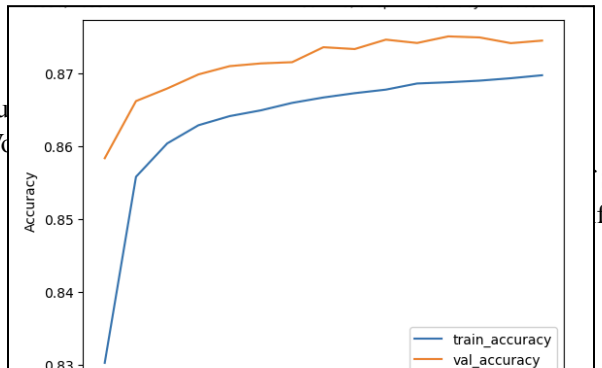
It was imperative to examine the independence of the features before further experimentation, and this was tested by keeping the dense numerical values consistent for each trial of this stage. These values were the same as Stage 1, as tabulated in figure 6.




F DNN model after removing the pupil position variable
c

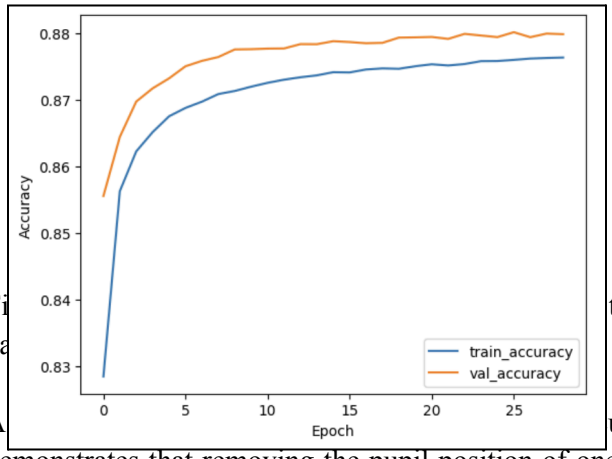



Fi the LSTM model after removing the pupil position
va




Ju Student Scholarship
Vo fordjss.org

Figure 11 (fig 11); Accuracy vs. Epoch graph for the DNN model after removing the pupil position variable of only the right eye.




Fi the LSTM model after removing the pupil position va

A uracy remained a constant 88% for each trial. This demonstrates that removing the pupil position of one eye versus the other did not make a difference, and this result conveys that the features are interdependent, meaning that despite removing a sub-variable of pupil position, the data eliminated was incorporated from the corresponding sub-variable. This is a reflection of redundancy, as the result is consistent with high correlation between the left and right eye sub-variables. As each sub-variable carries redundant information, when one is removed, the model's predicate capacity is unaffected. The left and right eye measurements proxy the same underlying gaze behavior as there is no reconstruction of missing data by the model. To confirm this interpretation, similar experiments were done for point of regard and eye position, and further comparison demonstrated the same result.

After it was deduced that the features were mutualistic, both sub-variables were removed, meaning that for each trial, both the left and right counterparts of a variable were removed. After removing both sub-sets for the eye position parameter, overall accuracy dropped to 86%. Across the five DNN architecture configuration tests in this stage, the validation accuracy ranged from 85% to 88% reflecting a natural variance across runs of 2% to 3%. This drop of 2% is at the boundary of the variance range and can be interpreted as preliminary evidence. After removing both sub-sets for the pupil position parameter, overall accuracy remained the consistent 88%. After removing both sub-sets for the point of regard parameter, overall accuracy remained the consistent 88%, implying that the eye position parameter was the only variable relevant for ASD detection in this dataset, as removal of only the eye position variable led to a drop in accuracy.

To test this inference, each variable was then isolated, and its accuracies were compared. Isolating the eye position parameter, Model 1 had an accuracy of 85%, and Model 2 had an accuracy of 86%. Isolating the pupil position parameter, both models had an accuracy of 84%. Isolating the point of regard parameter, both models had an accuracy of 73%. This demonstrates that models with only the eye position variable would have an accuracy very close to that of the original setup of the models, given that there was only an approximate 2% difference.

## V. DISCUSSION AND CONCLUSION

The dataset was composed of 59 participants, which is a small sample for training deep learning models, causing findings to not generalise larger or demographically diverse populations. The validation split was performed at the row level, meaning that time-window samples from the same individual may appear in the training and validation partitions, which risks inflating the accuracy. Variables such as gender distribution and participant age may also have independently influenced the patterns in the data, and they may be unrelated to ASD. It is important to note that model performance on a dataset does not automatically establish a clinical biomarker. The limitations described must be addressed by testing larger, more diverse data sets with multiple random seeds and participant-independent splits in order to determine the prospect of a potential biomarker. The research involved the implementation of LSTM and DNN models and ran experiments on an eye-tracking autism data set by altering the hyperparameter configurations. Based on the performances, a comparative analysis of the features associated with the disorder was derived. The main features compared were eye position (left and right), point of regard (left and right), and pupil position (left and right). By comparing the accuracies of both models after isolating each variable, it was determined that eye position is the most predictive factor of ASD within this dataset, positing that this variable has the potential to be considered as a significant biomarker candidate for future deep learning works. The relevance of this biomarker can lead to more efficient research in the future, as the other variables can be ignored in newer models, resulting in faster analysis of eye-tracking data and enabling non-invasive autism detections of greater accuracy.

## VI. REFERENCES

[1] World Health Organization, "Autism spectrum disorders," WHO Fact Sheets, Nov. 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders
[2] K. Tolvaišaitė, K. Butkutė-Šliuožienė, and A. Kizienė, "Late Autism Spectrum Disorder (ASD) diagnosis: A new perspective on life," *Eur. Psychiatry*, vol. 68, no. S1, p. S808, Aug. 2025, doi: 10.1192/j.eurpsy.2025.1644.
[3] C. Ladd-Acosta, "Is there an autism epidemic?," Johns Hopkins Bloomberg School of Public Health, May 2025. [Online]. Available: https://publichealth.jhu.edu/2025/is-there-an-autism-epidemic

[4] Autism Science Foundation, "2025 Autism Research Year in Review," Dec. 2025. [Online]. Available: https://autismsciencefoundation.org/2025-year-in-review/
[5] Autism Research Institute, "Age and diagnosis," Autism.org, 2024. [Online]. Available: https://autism.org/age-and-diagnosis/
[6] Autism Parenting Magazine, "Late autism diagnosis," 2024. [Online]. Available: https://www.autismparentingmagazine.com/late-autism-diagnosis/
[7] The TreeTop, "Negative effects of an autism diagnosis later in life," ABA Therapy Resources, 2024. [Online]. Available: https://www.thetreetop.com/aba-therapy/negative-effects-of-an-autism-diagnosis-later-in-life
[8] O. Gurau, W. J. Bosl, and C. R. Newton, "Systematic review on EEG analysis to diagnose and treat autism by evaluating functional connectivity and spectral power," *Front. Psychiatry*, vol. 14, Art. no. 1090142, Feb. 2023.
[9] G. A. Stefanatos and I. S. Baron, "Review of neuroimaging in autism spectrum disorders: What have we learned and where we go from here," *Mol. Autism*, vol. 2, no. 4, Apr. 2011.
[10] Y. Shen, K. A. Dies, I. A. Holm, et al., "Clinical genetic testing for patients with autism spectrum disorders," *Pediatrics*, vol. 125, no. 4, pp. e727–e735, Apr. 2010.
[11] K. Downey, E. M. Maye, H. J. Engel, et al., "Can screening with the Ages and Stages Questionnaire detect autism?," *J. Dev. Behav. Pediatr.*, vol. 36, no. 7, pp. 536–543, Sep. 2015.
[12] T. Falck-Ytter, S. Bölte, and G. Gredebäck, "Eye tracking in early autism research," *J. Neurodev. Disord.*, vol. 5, no. 28, Oct. 2013.
[13] W. Bin and W. Bin, "Eye tracking autism code," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/code/wasim2019/eye-tracking-autism-code
[14] ScienceDirect, "Deep neural network," in *Topics in Computer Science*, Elsevier, 2024. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/deep-neural-network
[15] DataScientest, "Deep neural network: What is it and how is it working?," 2024. [Online]. Available: https://datascientest.com/en/deep-neural-network-what-is-it-and-how-is-it-working
[16] Zemim, "Deep neural network (DNN) explained," Medium, 2024. [Online]. Available: https://medium.com/@zemim/deep-neural-network-dnn-explained-0f7311a0e869
[17] GeeksforGeeks, "Deep learning: Introduction to Long Short-Term Memory (LSTM)," 2024. [Online]. Available: https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/
[18] O. Calzone, "An intuitive explanation of LSTM," Medium, 2024. [Online]. Available: https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c