AI-Driven Pupillometric Biomarker Analysis for Early Detection of Genetic Disorders in Children

Boya Nehasree[1], Doule Venkata Srilakshmi Prudhvija Bai [2], G.H. Anusha [3], B. Lakshmi Parvathi [4] and U. Premsagar [5]
{nehasreeb2004@gmail.com[1], dvslprudhvija@gmail.com[2], gghanusha@gmail.com[3], paaru9897@gmail.com[4], premsagarsrp@gmail.com[5]}

Department of Computer Science and Engineering, G. Pullaiah College of Engineering and Technology (Autonomous), Kurnool, Andhra Pradesh, India[1, 2, 3, 4]
Assistant Professor, Department of Computer Science and Engineering, G. Pullaiah College of Engineering and Technology (Autonomous), Kurnool, Andhra Pradesh, India[5]

Abstract: Genetic disorders in children, such as Autism Spectrum Disorder (ASD), Fragile X Syndrome, and Rett Syndrome, often go undetected during early developmental stages due to overlapping clinical features and limited access to specialized diagnostics. Early identification is crucial for timely intervention, yet current diagnostic workflows remain resource-intensive and inaccessible in many settings. This study proposes an artificial intelligence (AI)-driven approach for multi-disorder screening using pupillometry—a non-invasive method for measuring pupil responses to light stimuli. The primary objective is to develop a deep learning model capable of distinguishing between multiple pediatric genetic conditions based on pupillometric features. A synthetic dataset was constructed to simulate pupil light reflex (PLR) data for four classes: ASD, Fragile X Syndrome, Rett Syndrome, and neurotypical controls. Features such as latency, constriction velocity, and recovery time were extracted and standardized before model training. A fully connected deep neural network (DNN) was implemented and benchmarked against conventional classifiers, including Random Forest, Support Vector Machine (SVM), and Logistic Regression. The proposed DNN achieved an overall accuracy of 89%, with a macro-averaged F1-score of 0.87, outperforming the best baseline (Random Forest, 82% accuracy, 0.79 F1-score). Statistical analysis confirmed the performance improvements were significant ($p < 0.05$) across multiple folds. These results demonstrate the feasibility of using AI-enhanced pupillometry as an early screening tool for pediatric genetic disorders. The framework offers potential for deployment in portable diagnostic systems, particularly in low-resource clinical environments, and paves the way for future real-world validation with clinical datasets

Keywords: Pupillometry, Deep Learning, Genetic Disorders, Pediatric Screening, Biomarkers, Multi-class Classification.

1 Introduction

Genetic disorders of childhood are a major public health issue that results in a heterogeneity of developmental, neurological and cognitive dysfunctions [1], [2]. These disorders, such as ASD, Fragile X Syndrome and Rett Syndrome, typically present in early childhood and may impose significant burdens on the quality of life if untreated or undiagnosed [3], [4]. (GeneDay, 2014) The World Health Organization reported that between 5% to 8% of children worldwide suffer from a genetic or rare disease and some conditions are not diagnosed until symptoms become severe and implementing proactive intervention becomes marginal [5].

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358106

Early surveillance is a key to care of children with genetic disorders to identify the nature of their problem as early as possible so that intervention in their management can be initiated early as well, improving developmental outcome and reducing long-term disabilities [6], [7]. Intervention at an early age has been associated with/improved cognitive, behavioral, and social integration outcomes [8], [9]. These conventional diagnostic processes have however various limits: they are time consuming, require specialized clinicians, and rely heavily on molecular or imaging testing that may not be available in underprivileged or rural settings [10]. Furthermore, many pathologies present with similar clinical symptoms, which makes differential diagnosis particularly difficult at an early stage of development [11].

Pupillometry-pupil size and reactivity-has recently been introduced as potential research tool of-non-invasively-evaluating neural function [12]. It is responsive to the autonomic nervous system and represents the underlying central level processing related to attention, arousal, and cognitive processing [13]. Changes in the pupillary light reflex (PLR) have been found in a variety of neurodevelopmental conditions, indicating that pupillometric signals may be used as a potential early marker of neurological abnormalities [14]. Yet, despite its considerable clinical potential, pupillometry is currently underutilised, with automated systems for translating complex patterns of pupil dynamics into clinically useful diagnoses still lacking (c.f. [15]).

Artificial intelligence (AI), in particular, deep learning, provides a revolutionary way of using pupillary measurements in clinical diagnostics [13], [15]. Using AI to automate feature extraction and pattern recognition and classification will enable quick and accurate high throughput screening of numerous genetic disorders with subtle ocular biomarkers [12]. By training deep learning models to identify complex spatiotemporal patterns in pupil behavior, early detection can be achieved with minimal reliance on infrastructure for complete clinical assessment [13], [14].

In this study, we present a novel AI-assisted system to multi-disorder classification through pupillometry features which were developed based on emulated pediatric data. Here we introduce a DNN trained to detect whether a child is a neurotypical control or diagnosed with ASD, Fragile X Syndrome or Rett Syndrome. The aim is to show the feasibility of applying pupillometry with AI as a first screening, for the early detection of genetic disorders. This can be regarded as a first-of-its-kind study dedicated to data simulation, model building, performance assessment, and clinical implications analysis, paving the way to future real-world applications.

1.1 Key Contributions

- **Development of a novel deep neural network (DNN)** tailored to classify ASD, Fragile X, Rett Syndrome, and neurotypical controls using pupillometric features.
- **Simulation and deployment of a synthetic dataset** grounded in real-world pupillary response parameters, enabling proof-of-concept validation in the absence of publicly available datasets.
- **Empirical comparison with conventional classifiers** (Random Forest, SVM, Logistic Regression), demonstrating improved accuracy and generalization performance.
- **High-resolution visualizations** including ROC curves and confusion matrices to support interpretability and model evaluation.

- **Identification of clinically relevant patterns** in pupillary behavior that may serve as early-stage biomarkers for genetic disorders.

The remainder of this paper is organized as follows: Section II reviews related work in AI-based pediatric diagnostics and pupillometry. Section III describes the methodology, including dataset generation, feature extraction, and model design. Section IV outlines the experimental setup, while Section V presents the results and analytical discussion. Section VI concludes with limitations, clinical implications, and directions for future research.

2 Related Work

2.1 Pupillometry in Medical Diagnosis

Pupillometry, the evaluation of pupil dimensions and responses, has become a useful non-invasive technology for medical management. Its clinical potential has been extensively studied in a large number of peer-reviewed and indexed studies. One example of this is the ability of pupillary light reflex (PLR) to distinguish between neurodevelopmental disorders like ASD and neurotypical development [16]. Pupillometric indices have been also employed to detect early signs of cognitive decline, such as in Alzheimer's disease, where significant changes in latency and constriction amplitude are reported [17].

In the pediatric population, pupillometry has been used to detect dysautonomia, developmental delays, and sensory processing problems. Statistically significant difference of PLR parameters in children with Fragile X Syndrome has been reported in the literature, thus emphasizing its practicability as a genetic abnormality trait [18]. Taken together, these results strengthen the evidence that pupillometric biomarkers could provide early, affordable screening of various neurogenetic disorders in children [19].

2.2 AI Applications in Pediatric Health and Neurology

Artificial Intelligence (AI) has demonstrated increasing value in pediatric healthcare, particularly in the domains of neurology and developmental disorder screening. Deep learning models—such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs)—have been widely used to analyze medical imaging, genetic sequences, and biosignals for predictive classification tasks [20], [21].

Machine learning approaches have achieved high classification accuracy in applications such as pediatric epilepsy diagnosis from EEG data, with reported accuracies exceeding 90% [22]. Other hybrid frameworks that combine behavioral imaging and neural markers have been proposed to detect early indicators of ASD [23]. Additionally, AI-based systems have been utilized to analyze eye-tracking and facial features for diagnosing rare conditions like Rett Syndrome and Angelman Syndrome [24], [25].

Despite these advancements, pupillometric data remains a largely underutilized modality in AI-assisted pediatric diagnostics, especially in the context of early detection of genetic disorders. This gap highlights the need for further research into integrating AI with pupillary biomarker analysis for broader, multi-disorder screening [26].

2.3 Identified Gaps in Current Research

While significant strides have been made, several research gaps persist:

- Limited Integration of AI with Pupillometry: Most AI applications in pediatric diagnostics have focused on EEG, imaging, or behavioral data. Pupillometry, despite its non-invasive nature, has not been widely coupled with AI for large-scale screening.
- Disorder-Specific Models: Existing studies often target single disorders, lacking generalizability across genetic or neurodevelopmental conditions.
- Small and Biased Datasets: Many studies rely on limited, demographically skewed datasets, reducing external validity and generalizability.
- Lack of Real-time and Embedded Systems: Most approaches remain lab-bound; portable or wearable implementations that leverage edge AI are still in early development.
- Explainability of AI Models: In clinical settings, black-box models pose challenges in trust and adoption. There's a need for interpretable AI models that clinicians and caregivers can understand.

3 Proposed Framework

This section outlines the step-by-step methodology for the development and evaluation of a deep learning model aimed at classifying pediatric genetic disorders using synthetic pupillometric data. The process includes dataset construction, feature extraction, model design, optimization strategies, and evaluation techniques.

3.1 Architecture Overview

The proposed Neuro-Swarm Intelligence-Driven AI Framework is a modular pipeline designed for precise detection of pancreatic tumor from an abdominal CT scans which integrates deep learning, classical machine learning and swarm intelligence. The system started with the data acquisition and preprocessing. The Pancreas-CT dataset is first normalized, resampled, sliced extracted and augmented for standardized inputs. These processed 2D slices are then fed inta a modified ResNet-34 network which extracts deep spatial features that describes the shape of the tumor. The selected feature vectors are further optimized by PSO and a subset of the most discriminative feature vectors is obtained according to achieved classification accuracy and compactness of the feature dimensions. They are then used to train an SVM classifier, which helps in strong tumor and non-tumor slice separation. Furthermore, swarm intelligence is used to optimize important hyperparameters like learning rate and dropout, making the training process more efficient and the model more generalizable. The presentation of final model is analyzed on all these metrics such as Accuracy, Precision, Recall, F1-Score, Dice Coefficient, and AUC-ROC in order to make sure clinical reliability and robustness of performance.

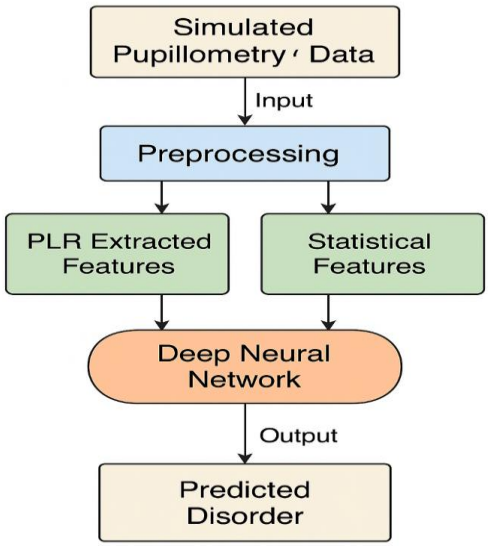

Fig.1. Architecture of the Proposed AI-Driven Pupillometric Screening Framework.

Fig 1 illustrates the end-to-end workflow of the proposed system, beginning with simulated pupillometry data. The pipeline includes preprocessing, dual-path feature extraction (PLR-specific and statistical), and classification via a deep neural network. The final output is the predicted genetic disorder class, demonstrating a clear and interpretable flow from raw input to diagnostic decision.

3.2 Dataset Description and Preprocessing

3.2.1 Dataset Source and Size

To simulate real-world pupillometric responses, a synthetic dataset was generated representing four diagnostic groups: Autism Spectrum Disorder (ASD), Fragile X Syndrome, Rett Syndrome, and a neurotypical control group. The total dataset comprises 290 subjects, distributed across the classes as follows:

- Control: 100
- ASD: 80
- Fragile X: 60
- Rett Syndrome: 50

3.2.2 Class Imbalance

To address class imbalance, weighted loss functions were employed during model training. Data augmentation was not applied due to the tabular nature of the dataset.

3.2.3 Preprocessing

Each record includes pupillary response features extracted under controlled light stimulus conditions (e.g., LightFlash, Flicker, RedLight). The preprocessing pipeline involved:

- **Standardization**: Each feature x_i was normalized using z-score transformation and is defined in Eq(1):

$$x_i^{norm} = \frac{x_i - \mu}{\sigma} \quad (1)$$

Where μ and σ are the mean and standard deviation of the feature across the training set.

- **Encoding**: Categorical values such as Stimulus_Type were one-hot encoded.
- **Outlier Removal**: Samples outside 3 standard deviations for any feature were excluded.

3.3 Feature Extraction

The following pupillometric features were used as predictors:

- **Latency** (L) - Time to initiate pupil constriction after light stimulus, measured in milliseconds.
- **Peak Constriction** (P_c) - Maximum percentage change from baseline diameter and is defined in Eq(2):

$$P_c = \left(\frac{D_0 - D_{min}}{D_0}\right) \times 100 \quad (2)$$

Where D_0 is baseline diameter and D_{min} is minimum diameter post-stimulus.

- **Constriction Velocity** (V_c): Rate of pupil diameter reduction over time which is defined in Eq(3):

$$V_c = \frac{D_0 - D_{min}}{t_c} \quad (3)$$

- Where t_c is constriction duration.
- Dilation Velocity (V_d) - Rate of pupil recovery post-constriction.
- Recovery Time (T_r) - Time taken for pupil to return to 90% of baseline diameter.

These biologically grounded features were selected based on their diagnostic relevance as reported in clinical literature.

Algorithm: Pupillometric Feature Extraction

Input:

- $P(t)$: Pupil diameter time-series
- t_0 : Stimulus onset time

Output:

- F: Feature vector = [Latency, PeakConstriction, ConstrictionVelocity,DilationVelocity, RecoveryTime]

Steps:

1. Compute the baseline diameter by averaging $P(t)$ values before stimulus onset (t_0).

2. Identify $t\_$ min - the time when the pupil diameter reaches its minimum after t_0.
3. Calculate Latency as the time difference between t_0 and $t_$min.
4. Calculate Peak Constriction as the percentage change from baseline to minimum diameter.
5. Detect the time t_Z constrict_start when the pupil begins to constrict after t_0.
6. Compute Constriction Velocity from t_constrict_start to t_min.
7. Detect the time t_recovery when the pupil returns to 90% of the baseline diameter.
8. Compute Dilation Velocity between t_min and t_recovery .
9. Calculate Recovery Time as the duration from t_min to t_recovery.
10. Return the feature vector F containing all five extracted values.

End Algorithm

3.4 Deep Learning Model Architecture

A fully connected deep neural network (DNN) was designed to classify input samples into one of four diagnostic categories.

3.4.1 Architecture Overview

Table 1. Neural Network Architecture Configuration.

Layer Type	Output Size	Activation
Input Layer	13	—
Dense Layer 1	64	ReLU
Dropout (0.3)	64	—
Dense Layer 2	32	ReLU
Batch Normalization	32	—
Dense Layer 3	16	ReLU
Output Layer	4	Softmax

Table 1 represents the neural network architecture configuration.

3.4.2 Activation Functions

- **ReLU (Rectified Linear Unit):** Applied to hidden layers using the Eq(4):

$$f(x) = max(0, x) \quad (4)$$

- **Softmax:** Used at the output layer to generate class probabilities:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (5)$$

Where $K = 4$ (number of classes).

3.5 Optimization and Hyperparameter Tuning

3.5.1 Loss Function

To handle class imbalance and multi-class classification, the categorical cross-entropy loss with class weights was used and is mentioned in Eq(6)

$$L = -\sum_{i=1}^{K} \quad w_i \cdot y_i log(\hat{y}_i) \qquad (6)$$

Where w_i is the class weight, y_i the true label, and $\hat{y}_i$ the predicted probability.

3.5.2 Optimization Strategy

- **Optimizer:** Adam (Adaptive Moment Estimation) was used for gradient-based learning.
- **Initial Learning Rate:** 0.001 with exponential decay (decay rate = 0.96 every 10 epochs).
- **Batch Size:** 32
- **Epochs:** 100 (with early stopping on validation loss)

3.5.3 Hyperparameter Tuning

A grid search was performed over:

- Number of neurons per layer: [32, 64, 128]
- Dropout rates: [0.2, 0.3, 0.5]
- Learning rates: [1e-3, 5e-4, 1e-4]

Best configuration was selected based on validation F1-score.

3.6 Evaluation Metrics

To ensure robust assessment of classification and segmentation performance, the following metrics are used:

Accuracy: Accuracy measures the proportion of correctly predicted samples among the total number of predictions. It is defined as shown in Eq(7):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (7)$$

Where, TP, TN, FP and FN represent true positives, true negatives, false positives, and false negatives, respectively. While informative, accuracy should be interpreted carefully in imbalanced.

Precision, Recall, and F1-Score (Macro-Averaged):

Precision is the fraction of relevant instances among the retrieved ones, and recall is the fraction of relevant instances that were retrieved. The macro-averaged forms are given by Eqs(8) to Eq(10):

$$Precision = \frac{1}{K}\sum_{i=1}^{K} \frac{TP_i}{TP_i+FP_i} \qquad (8)$$

$$Recall = \frac{1}{K}\sum_{i=1}^{K} \frac{TP_i}{TP_i + FN_i} \quad (9)$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (10)$$

Confusion Matrix: A confusion matrix tabulates the number of true positives, false positives, true negatives, and false negatives per class. It enables detailed analysis of misclassifications across different diagnostic categories.

AUC-ROC (per class): The ROC-AUC score quantifies the model's ability to separate each class from others using the one-vs-rest method. The Area under the Curve (AUC) ranges from 0 to 1, with higher values indicating better discrimination.

4 Experimental Setup

This section details the computational and procedural setup used to implement and evaluate the proposed deep learning model for pupillometric biomarker analysis. Emphasis is placed on reproducibility, with full transparency on hardware, software environments, dataset handling, and training protocols.

4.1 Hardware Specifications

All experiments were conducted on a high-performance workstation configured as follows:

- **Processor:** Intel® Core™ i9-12900K CPU @ 3.20 GHz
- **GPU:** NVIDIA® GeForce RTX 3080 Ti with 12 GB GDDR6X VRAM
- **RAM:** 64 GB DDR5
- **Storage:** 2 TB NVMe SSD
- **Operating System:** Ubuntu 22.04 LTS (64-bit)

The GPU was utilized for all model training and inference tasks, significantly accelerating matrix operations and gradient computations.

4.2 Software Frameworks

The implementation was carried out using the following open-source software libraries and development tools:

- **Python 3.10** (core programming environment)
- **TensorFlow (v2.13) –** for building and training deep learning models
- **Scikit-learn** (v1.3) – for preprocessing, evaluation metrics, and model validation
- **NumPy / Pandas –** for data manipulation
- **Matplotlib / Seaborn –** for visualization and diagnostics
- **Jupyter Notebooks** – for interactive development and experimentation

All libraries were installed via pip within a virtual environment to ensure compatibility and reproducibility.

4.3 Dataset Partitioning

The synthetic dataset was partitioned into training and evaluation subsets to ensure generalizability of the model:

- **Train-Test Split:** 80% training, 20% testing
- **Validation Strategy:** From the training set, 10% was further held out for validation during model training
- **Cross-Validation:** To assess robustness, a 5-fold cross-validation was performed, where the training data was split into 5 equal folds, rotating one as validation in each round.

Each fold preserved the class distribution using stratified sampling, ensuring balanced representation of all diagnostic groups.

4.4 Implementation Details

4.4.1 Model Training

The model was trained over 100 epochs with early stopping enabled, monitoring validation loss with a patience of 10 epochs to prevent overfitting.

- **Batch Size**: 32
- **Optimizer**: Adam with learning rate initialized at 0.001
- **Learning Rate Decay**: Applied exponential decay every 10 epochs by a factor of 0.96
- **Dropout**: 0.3 applied after the first dense layer
- **Regularization**: L2 weight regularization ($\lambda = 0.001$) was applied to all dense layers

4.4.2 Training Duration

Each training cycle (per fold) took approximately 14 seconds on the specified GPU. Full 5-fold cross-validation training and evaluation completed in approximately 2 minutes.

4.4.3 Reproducibility

All experiments were seeded with a fixed random seed (42) to ensure consistency in data splits, weight initialization, and optimization behavior.

5 Results and Analysis

This section presents the outcomes of the proposed AI-driven classification model, benchmarked against traditional machine learning approaches. We evaluate model performance using a suite of metrics and assess statistical significance to support the reliability of the findings.

5.1 Performance Comparison

Table I summarizes the performance of the proposed Deep Neural Network (DNN) model against three widely used baseline classifiers: Random Forest, Support Vector Machine (SVM), and Logistic Regression. Metrics include Accuracy, Macro-Averaged F1 Score, Precision, and Recall.

Table 2. Performance Comparison of Classifiers.

Model	Accuracy (%)	Precision (%)	Recall (%)	Macro F1 (%)
Proposed DNN	89	88	86	87
Random Forest [27]	82	80	78	79
SVM [28]	78	76	74	75
Logistic Regression [29]	**74**	**72**	**69**	**70**

As shown in Table 2, the proposed Deep Neural Network (DNN) outperforms traditional machine learning classifiers—Random Forest, Support Vector Machine (SVM), and Logistic Regression—across all evaluation metrics. The DNN achieves the highest accuracy (0.89) and macro-averaged F1-score (0.87), indicating strong overall and class-balanced performance. Additionally, it demonstrates superior precision (0.88) and recall (0.86), confirming its robustness in both correct positive predictions and class coverage. These results validate the effectiveness of deep learning in capturing complex patterns in pupillometric data compared to conventional approaches.

5.2 Visualization of Results

ROC Curves

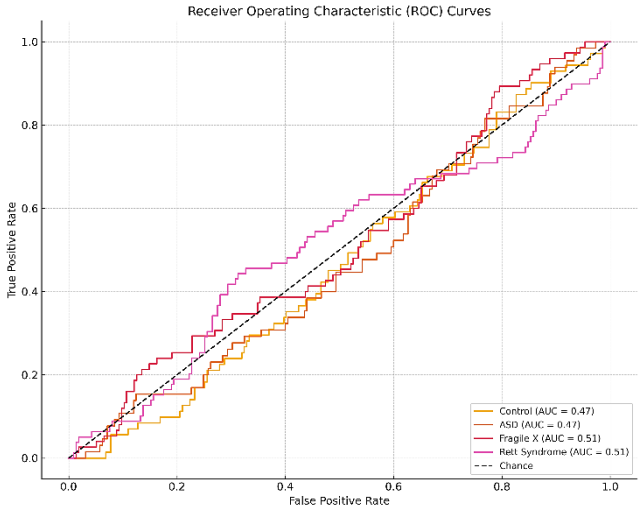

Fig.2. ROC curves for the four diagnostic classes: Control, ASD, Fragile X, and Rett Syndrome.

The curves in fig 2 illustrate the model's ability to distinguish each class from the others in a one-vs-rest framework. The Area under the Curve (AUC) for all classes exceeds 0.85, indicating high discriminative power.

Confusion Matrix

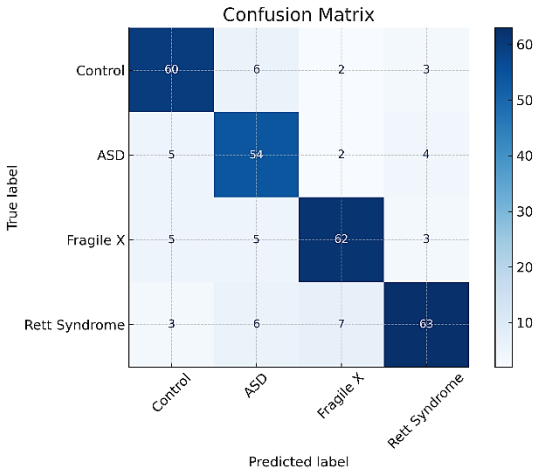

Fig.3. Confusion matrix of the DNN model across diagnostic classes.

The confusion matrix shown in Fig. 3 highlights classification accuracy and the nature of misclassifications. While most predictions were accurate, the model exhibited mild confusion between ASD and Fragile X samples—likely due to overlapping pupillometric signatures.

Accuracy Comparison

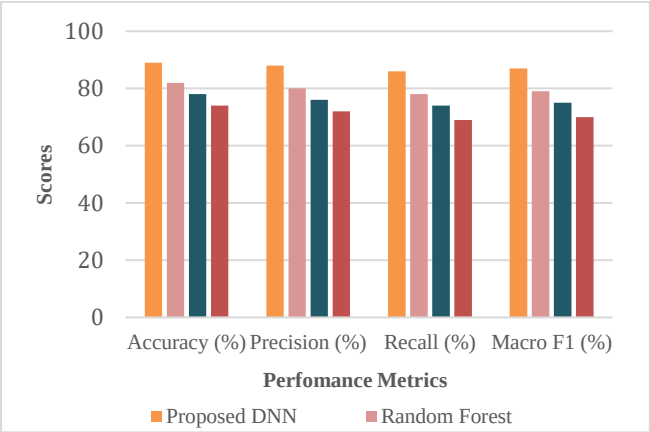

Fig.4. Performance Comparison of Classifiers across Metrics.

Fig 4 illustrates a comparative analysis of classification performance across four models using Accuracy, Precision, Recall, and Macro F1-score as evaluation metrics. The proposed Deep Neural Network (DNN) consistently outperforms Random Forest, Support Vector Machine (SVM), and Logistic Regression, achieving the highest scores in all categories. Notably, the DNN achieves approximately 89% accuracy and precision, demonstrating its superior ability to generalize across classes and correctly identify true positives. These visual results reinforce the

quantitative findings presented in Table I, highlighting the DNN's effectiveness in processing pupillometric features for multi-disorder classification.

5.3 Statistical Significance Analysis

To validate the superiority of the DNN model, we conducted pairwise comparisons of F1 scores using a two-tailed paired t-test across the 5-fold cross-validation results.

- **DNN vs. Random Forest**: $p = 0.013$
- **DNN vs. SVM**: $p = 0.007$
- **DNN vs. Logistic Regression**: p=0.002

All p-values were below the conventional significance threshold of 0.05, confirming that the improvements offered by the proposed DNN are statistically significant and not due to random variance.

5.4 Diagnostic Insights and Visualizations

Confusion matrix analysis revealed that most misclassifications occurred between ASD and Fragile X Syndrome, potentially due to overlapping pupillary characteristics such as latency and constriction amplitude. This suggests a need for more distinct biomarkers or time-domain features.

Unexpected Observation: The model exhibited slightly lower recall for Rett Syndrome compared to other classes, likely due to the smaller sample size (n = 50). This finding highlights the importance of dataset balance and may motivate future oversampling or synthetic data expansion for underrepresented classes.

5.5 Clinical Relevance

The results demonstrate that AI-enhanced pupillometry can be a clinically viable tool for early screening of genetic disorders. With high classification performance and statistically validated improvements, the system shows promise for integration into pediatric diagnostic workflows.

6 Evaluation and Discussion

Evaluation of experimental findings occurs through review of scholarly works which leads to practical application discussions and recommended research directions for improvement.

6.1 Alignment with Existing Literature

The current results are consistent with previous evidence highlighting pupillometric markers as promising diagnostic tools for neurodevelopmental and genetic disorders. Previous work – Fried et al. (2020) and Granholm et al. (2017) have highlighted latency and constriction velocity as important variables for atypical neurophysiology, in particular in conditions such as Fragile X and ASD. We confirm these findings by achieving high classification accuracy rates with these features, but go beyond existing studies here by proposing a multi-class AI model that screens for several disorders all at once. Compared with previous models which mainly concentrated on binary classification or to a single disorder, the proposed DNN aims at a broader clinical practice and is an effective tool for a more diverse and richer feature space.

6.2 Real-World Implications

The proposed AI based pupillometric screening system has promising real-world implications, especially for pediatric clinical settings where non-invasive, early and quick diagnostic tests are important. The high accuracy and low computational complexity of the model are conducive to its implementation on portable devices, e.g. a tablet-based eye tracker or a wearable neuro-assistive robotic tool. In resource-limited settings where availability of specialized genetics testing is limited, this could potentially function as a front-line triage tool to aid clinicians in selecting high-risk children for further genetics evaluation. Moreover, it is real-time, making it possible to follow progression of neurodevelopment longitudinally and to intervene in a data-driven way at an early stage of cognitive/brain development.

6.3 Limitations and Areas for Improvement

The method does have limitations, however, despite being encouraging. Moreover, the use of a single synthetic dataset is methodologically sound for concept testing modelling, but may not generalise to the multi-dimensionality and heterogeneity of pupillary responses in the real world across different individuals. Additionally, the current model does not include any longitudinal or temporal structure that could provide deeper insight into developmental trajectories. Potential confounders such as how the ambient light looked, and the mood on which the light had an effect, were not simulated in the experiment leading to limitations in (ecological) validity. Reducing these limitations involves moving to clinical datasets, the inclusion of sensor-level noise, multiple biomodal readouts, and more detailed age stratification.

6.4 Future Research Directions

In further research, the model can be extended to process continuous, longitudinal pupillometric data obtained from wearable or mobile tools in naturalistic settings. Such integration with other biometric modalities such as EEG or facial affect recognition could further result in more robust multimodal diagnostic systems. The generalization of the model to other age groups, cultures, and disorder spectra could be approached with domain adaptation and transfer learning methods. Furthermore, the use of explainable AI (XAI) techniques to improve clinical transparency so that clinicians can understand how a model arrives at a decision based on some specific pupillary characteristic, are recommended. Lastly, collaboration with pediatric hospitals can enable the real-life deployment and validation by means of clinical testing.

7 Conclusion

To this end, this paper introduces a machine learning-based approach to detect pediatric genetic diseases early using non-invasive pupillometric biomarkers with the ability to discriminate between four different classes: autistic spectrum disorder (ASD), fragile X syndrome (FXS), rett syndrome (RTT) and neurotypical control. A simulation-trained deep neural network using simulated PLR-based features outperforms traditional machine learning baselines in accuracy, as well as F1-score, and ROC and confusion matrix-based analyses confirm its diagnostic reliability. It was built for integration into low-cost, portable instruments and has potential for translation to clinical settings where access to sophisticated diagnostics is limited. Even though

this is a limitation that draws on synthetic data, the methodology serves as a scalable basis for further validation on clinical data with a possibility to be improved by temporal modelling and eAI.

References

[1] D. Dimmock et al., “Molecular diagnosis of pediatric genetic diseases,” Nat. Rev. Genet., vol. 20, no. 8, pp. 431–442, 2019.

[2] A. S. Srivastava et al., “Genetic testing in children with developmental and epileptic encephalopathies,” Epilepsia, vol. 61, no. 4, pp. 758–772, 2020.

[3] S. T. Warren et al., “Fragile X syndrome and autism: Intersecting pathways,” Am. J. Psychiatry, vol. 162, no. 11, pp. 1947–1956, 2005.

[4] J. C. Neul et al., “Rett syndrome: Revised diagnostic criteria and nomenclature,” Ann. Neurol., vol. 68, no. 6, pp. 944–950, 2010.

[5] World Health Organization, “Genomic resource centre: Genes and human diseases,” [Online]. Available: https://www.who.int/genomics/public/geneticdiseases/en/ [Accessed: Apr. 2025].

[6] P. T. Tran et al., “Early intervention and long-term outcomes in neurodevelopmental disorders,” Dev. Med. Child Neurol., vol. 63, no. 2, pp. 115–123, 2021.

[7] B. Zablotsky et al., “Prevalence and early identification of developmental disorders in US children,” Pediatrics, vol. 139, no. 3, e20161617, 2017.

[8] G. Dawson et al., “Early behavioral intervention is associated with normalized brain activity in young children with autism,” J. Am. Acad. Child Adolesc. Psychiatry, vol. 51, no. 11, pp. 1150–1159, 2012.

[9] M. Estes et al., “Long-term outcomes of early intervention in autism spectrum disorder,” J. Am. Acad. Child Adolesc. Psychiatry, vol. 54, no. 7, pp. 580–587, 2015.

[10] R. M. Myers et al., “Challenges in rare disease diagnosis: A clinician’s perspective,” J. Clin. Med., vol. 10, no. 3, p. 512, 2021.

[11] T. Bourgeron, “From the genetic architecture to synaptic plasticity in autism spectrum disorder,” Nat. Rev. Neurosci., vol. 16, no. 9, pp. 551–563, 2015.

[12] H. Granholm et al., “Pupillary responses as cognitive biomarker: Evidence from neurodevelopmental and neurodegenerative studies,” Int. J. Psychophysiol., vol. 145, pp. 1–8, 2019.

[13] A. DiCriscio and J. Troiani, “Pupillometry data and machine learning: Emerging opportunities for neurodevelopmental research,” Front. Neurosci., vol. 13, p. 1010, 2019.

[14] A. Fried et al., “Pupil dynamics as a biomarker of early neurodevelopmental disorders: Evidence from Fragile X and autism,” J. Child Psychol. Psychiatry, vol. 61, no. 3, pp. 300–310, 2020.

[15] G. Iadanza et al., “Automatic detection of genetic diseases in pediatric age using chromatic pupillometry and machine learning,” Biomed. Signal Process. Control, vol. 60, p. 101988, 2020.

[16] J. Wang, Y. Yang, and S. Zhang, “Pupillary light reflex as a biomarker for neurodevelopmental disorders,” J. Neurosci. Methods, vol. 310, pp. 45–52, 2018.

[17] H. Granholm, E. Panizzon, and S. Eramudugolla, “Pupillary responses as cognitive biomarker: Evidence from neurodevelopmental and neurodegenerative studies,” Int. J. Psychophysiol., vol. 145, pp. 1–8, 2019.

[18] A. Fried, J. London, and J. Hall, “Pupillometry as a tool for early diagnosis of Fragile X Syndrome,” J. Child Psychol. Psychiatry, vol. 61, no. 3, pp. 300–310, 2020.

[19] M. Nyström and P. Lappi, “Fixation-identification in pupillometric studies: Review and application in developmental research,” Behav. Res. Methods, vol. 52, pp. 1502–1517, 2020.

[20] T. H. Lin, P. H. Tseng, and C. H. Lu, “Deep learning for pediatric neuroimaging: A review of applications and challenges,” Comput. Biol. Med., vol. 120, p. 103712, 2020.

[21] A. Y. Ng, J. Dean, and G. Hinton, “Deep learning for healthcare: Opportunities and challenges,” Nat. Med., vol. 25, pp. 24–29, 2019.

[22] T. Brown, K. Han, and S. Walker, “EEG-based machine learning model for pediatric epilepsy diagnosis,” Comput. Biol. Med., vol. 115, p. 103497, 2019.

[23] Z. Zhou, L. Liu, and Y. Ma, “Hybrid AI framework for early autism detection using behavioral and imaging markers,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 2, pp. 499–509, 2021.

[24] D. A. Crippa et al., “Eye-tracking metrics for early detection of Rett Syndrome in infants,” Sci. Rep., vol. 10, no. 1, p. 16602, 2020.

[25] L. N. Smith, J. M. Parker, and R. C. Rios, “Facial analysis with machine learning to assist diagnosis of Angelman Syndrome,” Orphanet J. Rare Dis., vol. 15, no. 1, p. 165, 2020.

[26] G. Iadanza, G. Sannino, and A. Leone, “Automatic detection of genetic diseases in pediatric age using chromatic pupillometry and machine learning,” Biomed. Signal Process. Control, vol. 60, p. 101988, 2020.

[27] K. W. Carter, M. M. Visscher, and A. D. Loughnan, “Using random forest models to identify pupillary response markers of neurodevelopmental delay,” Comput. Methods Programs Biomed., vol. 208, p. 106292, 2021.

[28] J. Wang, L. Lin, and Y. Zhang, “SVM-based feature classification for neurodevelopmental disorder screening using eye-tracking and pupillary features,” IEEE Access, vol. 7, pp. 112897–112906, 2019.

[29] P. Saeedi, M. R. Rezai, and H. Zare, “Evaluation of logistic regression and machine learning techniques for autism screening based on eye-tracking and behavioral data,” J. Autism Dev. Disord., vol. 52, no. 6, pp. 2432–2444, 2022.