

# NEURODEVELOPMENTAL PREDICTION USING SVC ALGORITHM AND DEEP LEARNING MODEL

**Hemaprabha<sup>1</sup>, Pooja H N<sup>2</sup>, Spoorthi P<sup>3</sup>**

Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India<sup>1-3</sup>

**Abstract:** Neurodevelopmental disorders (NDDs) usually develop during early childhood and affect a person's ability to think, feel, and interact with others. Common conditions associated with NDDs include ASD, ADHD, as well as other issues like intellectual disabilities, learning challenges, and cerebral palsy. This study introduces a machine learning method to classify different types of NDDs into multiple categories using both traditional and deep learning models. A dataset containing more than 6,000 entries was created, including information about gender, age, and clinical symptoms. The data was cleaned and prepared using label encoding and feature scaling methods. Two models were built and tested: a Support Vector Classifier (SVC) with a linear kernel and a DNN that uses a fully connected structure based on CNNs. The SVC model achieved an accuracy of 84%, while the deep learning model performed better, with an accuracy of 88%. This was verified using ROC curves and AUC analysis. These findings demonstrate that deep learning is effective in identifying complex patterns in data and may help enhance diagnostic processes for individuals with neurodevelopmental disorders.

**Keywords:** svc, linear svc, machine learning, mental health, neurodevelopmental, CNN.

## I. INTRODUCTION

The human brain is the control organ of the entire nervous system. The growth and functioning of the brain can be disturbed by neurodevelopmental disorders at times. These disorders can result in learning disorders and impair the intellect of humans. ASD and ADHD are forms of neurodevelopmental disorders that have become extremely common in children [1]. Neurodevelopmental disorder affects brain development and results in depression, social behaviour changes, cognitive abilities, and communication delays [2]. They affect the lifetime and create a lot of high medical costs for their families [3].

Though ASD symptoms are similar for both men and women, the studies show that severity is higher in men [4]. ML can identify patterns and establish relationships in data [5]. Weka, scikit-learn, MATLAB toolbox, and R are the software tools that can be used to implement various ML algorithms. Machine learning classifier algorithms can help in the early identification of such disorders. The accuracy of the predictions and diagnoses can be improved by bridging artificial intelligence, statistical modelling, and other data analysis technologies.[6].

Spatial and frequency abnormalities in the EEG can assist in the diagnosis of ADHD.CNN models utilising the EEG signal have also been explored [2].

The predictions from the model can support doctors in diagnosing conditions early. This paper has achieved accurate predictions by implementing a deep learning algorithm with an SVC.

## II. LITERATURE SURVEY

The neurodevelopmental disorders (NDDs), encompassing autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD or ADHD, intellectual disability ID, and specific learning disorder SLD), have such a large symptom profile that overlaps among them some and other conditions can easily arise. The earlier the complications are identified better. But traditional diagnosis is not entirely dependent on the behaviour of a person and clinical judgement, which may be time-consuming processes, creating subjective results. Two machine learning (ML) techniques, Support Vector Machines (SVM) and their classifier version, SVC, have been shown to provide promising effects on implant time reduction for the accuracy enhancement of NDD diagnosis and execution. Given that SVC is a

powerful model for high-dimensional data to distinctly identify decision boundaries, it proves advantageous in the case of complex biomedical and behavioural data.

Toki et al. [2] created a diagnostic model that uses ML on behavioural data gathered from a serious game to evaluate children's developmental traits. The authors considered many algorithms (including SVM) but ultimately chose logistic regression for the final predictive model. Indeed, the SVM approach was able to differentiate between children with NDDs and TD controls correctly in more than 85 % of test cases. This is the ability that differed most between groups when clustering was performed along dimensions characterising behaviour (e.g. phonological processes, verbal IQ, or cognitive flexibility).

Thabtah [7] discussed the application of SVM to ASD screening through behaviour-based datasets and proved that SVM performed brighter than some traditional classifiers, such as naïve Bayes and Decision tree, with 96.7% accuracy. Deshpande and Joseph [8] also used SVM on structural and functional MRI data in diagnosing ADHD, where the model was qualified to attain around 85% accuracy with showing SVM's strength in neuroimaging-based classification tasks.

Alam et al. [9] compared several ML methods to predict neurodevelopmental disorders early on, including SVC, k-nearest neighbors (k-NN), random forests, and DNN. The evaluation's findings showed that SVC is particularly suitable for smaller or unbalanced datasets and performs well in terms of both precision and recall.

SVM has also been utilized to analyse the EEG signal in studies on the identification of ASD and ADHD. Using nonlinear SVMs for SVM-extracted features of EEG, Ghiassian et al. [10] revealed that they could detect attention and hyperactivity deficits with high sensitivity. Its possible utility in multimodal diagnostic approaches was also proved through research that combined SVM with speech features and eye-tracking parameters.

Despite its advantages, Support Vector Classification (SVC) has some significant drawbacks. Clinical applications require slow kernel and regularisation parameter tuning, and the algorithm's decision function is not interpretably defined. However, when paired with deep learning networks, ensemble techniques, and dimensionality reduction, it is a good choice for early diagnostic systems.

In conclusion, it has been proven in earlier studies that Support Vector Classification (SVC) is a reliable predictor of neurodevelopmental disorders. Thus, the reliable performance of this model across a variety of implementation contexts (e.g. biomedical signal processing and/or behavioural assessment) supports its status as an important component in predictive modelling paradigms aimed at early/accurate prediction/timely detection of NDDs.

In this paper, The Neurodevelopmental disorders are difficult to diagnose using traditional methods, so predictive models are useful in enhancing accuracy. this study utilised both ML and DL on clinical and behavioural data. The SVC gave 84% accuracy, providing a good baseline. The initial deep learning model (DL-V1) performed better at 86% accuracy, showing it could find more detailed patterns. With optimisation, the deep learning model (DL-V2) had 88% accuracy, highlighting its strength in managing complex data. Compared to SVC, deep learning provides more accuracy and earlier diagnosis of neurodevelopmental disorders.

### III. METHODOLOGY

#### 3.1 Data Collection:

A specially designed synthetic dataset has been employed for this research. The dataset consists of greater than 6,000 samples mimicking single cases of different Neurodevelopmental Disorders [NDDs], such as ASD, ADHD, ID, and LD. Every record has both demographic and clinical features related to NDD diagnosis.

There are many key features in the neurodevelopment data, like age, gender, family history of genetic disorder, speech delay, motor skill problem, attention span, IQ level, ability to control emotions, learning difficulty, neurological disorder, and type of disorder, which collectively preserve demographic, intellectual, behavioural, and clinical features informative of neurodevelopmental examination.

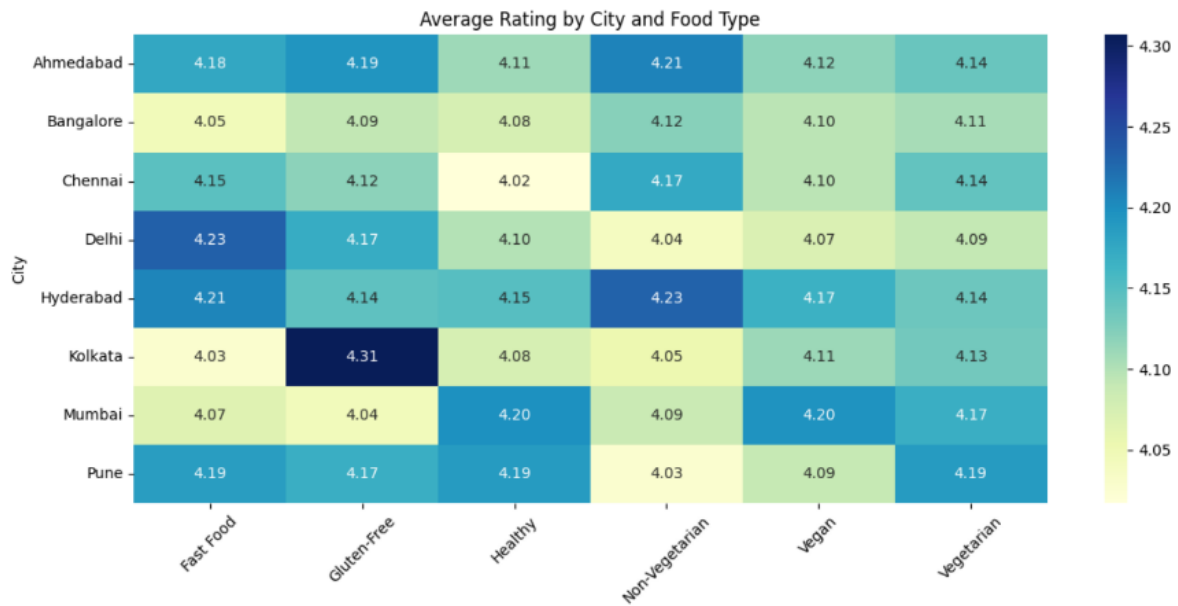


Figure1. Heap Map of Dataset

### 3.2 Data Preprocessing:

**Selection of Associated Variables from Dataset:** Important features were identified from the initial dataset based on previous clinical studies and knowledge in the domain. These include demographic variables (for example, age, sex), cognitive test measures (for example, IQ test score, attention span), and behaviour tests (for example, emotional control test, measure of social interaction). Irrelevant and redundant variables were not employed to reduce noise and dimensionality.

**Exclusion of Missing and Extreme Data:** Missing or null value records in the important variables were either imputed with proper statistical methods (e.g., mean/median imputation) or excluded on exceeding acceptable limits. Furthermore, outliers and Extreme observations that could distort or have a detrimental effect on model training were detected through Z-score and interquartile range (IQR) techniques and treated accordingly.

**Setting Dependent Variables and Splitting Datasets into 16 Regions:** The dependent target variable (disorder\_type) was set as per clinically established neurodevelopmental disorders, i.e., ASD, ADHD, ID, and LD, and control participants. For region-wise analysis (if geographic or institutional regions are present in the dataset), the dataset was divided into 16 different subsets to facilitate region-wise comparison of performance and tests of generalisation.

**Feature Selection and Grouping:** To make certain that only the most effective predictors were kept, feature selection integrated correlation analysis with recursive feature elimination (RFE). The chosen features were subsequently ranked into three wide categories: demographic features (gender, age, family history), cognitive and behavioural indicators (IQ score, attention span, social interaction), and clinical or medical characteristics (motor problems, speech delay, genetic predisposition). This classification enhanced the interpretability of the result and gave more informative insights into feature importance in the ML models

**Evaluation of ML Classification Algorithm:** After preprocessing, several classification methods were employed, including DNN to detect non-linear relationships, SVC to separate linearly, and Optional ensemble or decision tree methods to attain interpretability. Metrics like overall accuracy, confusion matrices, precision, recall, and F1-score were used to assess the models' performance after they were trained on the pre-processed data.

### 3.3 Support Vector Classifier (SVC):

A Support Vector Machine with a linear kernel was used as a traditional ML baseline. The model was used because it can handle multi-class problems that are nearly linearly separable.

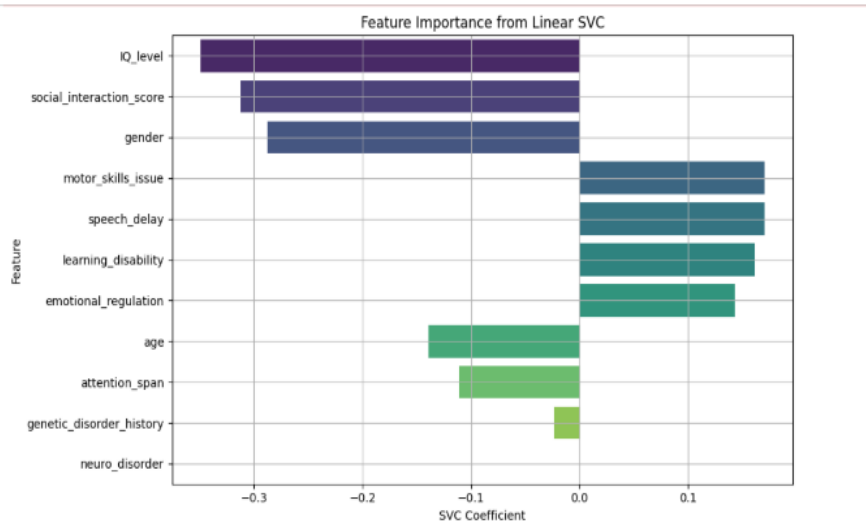


Figure2. Feature Importance

The dataset contains 6,000 records with information on individuals’ age, gender, and medical or developmental history. The dataset contains information on genetic disorder history, speech delays, motor skill difficulties, attention span, social interaction, IQ levels, emotional regulation, and learning challenges. It also notes the presence of any neurodevelopmental disorder and the corresponding disorder type. Most of the data is numeric, while gender and disorder type are categorical. Overall, this dataset offers a detailed view of developmental, cognitive, and social characteristics useful for analysis.

**3.4 Deep Learning Model (DNN)**

CNNs are specialized deep learning architectures for structured grid-like data, e.g., images. CNNs, unlike traditional neural networks, apply local spatial patterns in the form of convolutional layers that act on small receptive fields. The model can thus progressively extract features of different complexity levels. Such models perform well in object detection, image classification, and medical imaging because they learn hierarchical features automatically.

**Deep Learning V1 Module (Version 1):**

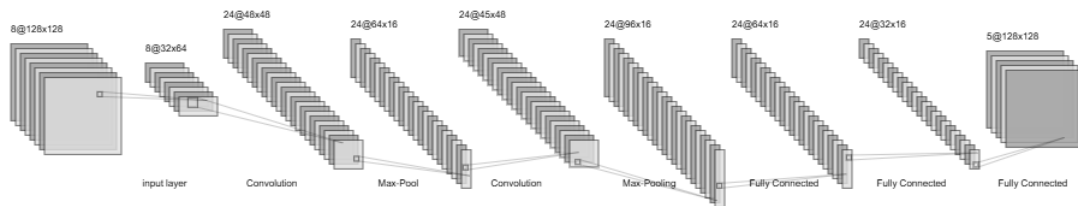


Figure 3. Convolutional Neural Network (1)

**Architecture:**

- Input Layer: 128 neurons, ReLU Activation
- Convolutional & Dense Layers:
- Conv Layer: 32 filters, 3×3 kernel, ReLU
- MaxPooling (2×2)
- Conv Layer: 64 filters, 3×3 kernel, ReLU
- MaxPooling (2×2)
- Dense (96 neurons), ReLU
- Dropout (0.2)
- Dense (64 neurons), ReLU
- Dense (32 neurons), ReLU
- Dropout (0.3)
- Output Layer:
- Dense (5 neurons for 5 classes: ASD, ADHD, ID, LD, Control)
- Softmax activation

**Training Configuration**

Optimiser: Adam

Loss Function: Categorical Cross-Entropy

Batch Size: 32

Epochs: 150 (with early stopping)

Validation Split: 10%

Callbacks:

-EarlyStopping (patience = 15, monitors validation accuracy)

-ModelCheckpoint (saves best model as best\_model\_v1.h5)

Evaluation

Metric: Test set accuracy

Deep Learning Accuracy (V1 CNN): 86%

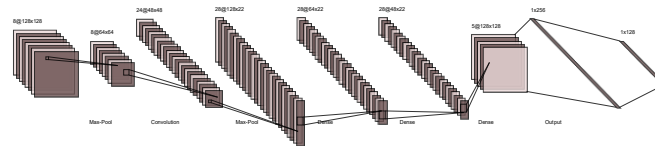
**Deep Learning V2 Module (Version 2):**

Figure 4. Convolutional Neural Network (2)

**Architecture:**

A CNN was implemented to discover intricate nonlinear patterns in the data and surpass conventional ML classifiers.

Input Layer: 128 neurons, ReLU activation

Hidden Layers:

Dense (96), ReLU

Dense (64), ReLU

Dense (48), ReLU

Dense (32), ReLU

Dropout Layers: Following the first and third layers (Dropout (0.2) and Dropout (0.3))

Output Layer: Softmax activation for multiclass classification (5 classes)

**Training Configuration**

Optimiser: Adam

Loss Function: Categorical Cross-Entropy

Batch Size: 32

Epochs: 150 (with early stopping)

Validation Split: 10%

Callbacks:

- EarlyStopping: Patient of 15, validates against validation accuracy

- ModelCheckpoint: Saves the model with the best validation accuracy as 'best\_model.h5'

Evaluation Metric: Test set accuracy.

Deep Learning Accuracy:88%

**IV. RESULT**

An ML predictive model is proposed in this study to predict neurodevelopmental disorders (NDDs), including learning disabilities (LD), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and intellectual disability (ID). A Support Vector Classifier (SVC) and a deep learning-based network (DNN) were two of the predictive models that were created and assessed using a carefully selected dataset that contains both demographic and clinical parameters.

With a high accuracy of about 84%, the SVC model with the linear kernel demonstrated impressive performance, demonstrating its proficiency with structured, linearly separable data. With a higher accuracy of 88%, the DL model,

which consists of a number of dense and dropout layers, performed brighter than the SVC model, demonstrating its ability to model intricate nonlinear relationships between features.

The results demonstrate that DL models can yield better predictive performance when trained on enough data and fine-tuned, even though traditional ML techniques like SVC provide a reliable benchmark that is simple to interpret. This study illustrates the possibilities of data-driven models to support precise and early NDD screening. These models can support clinical judgment, especially in settings where access to comprehensive psychiatric evaluation is restricted.

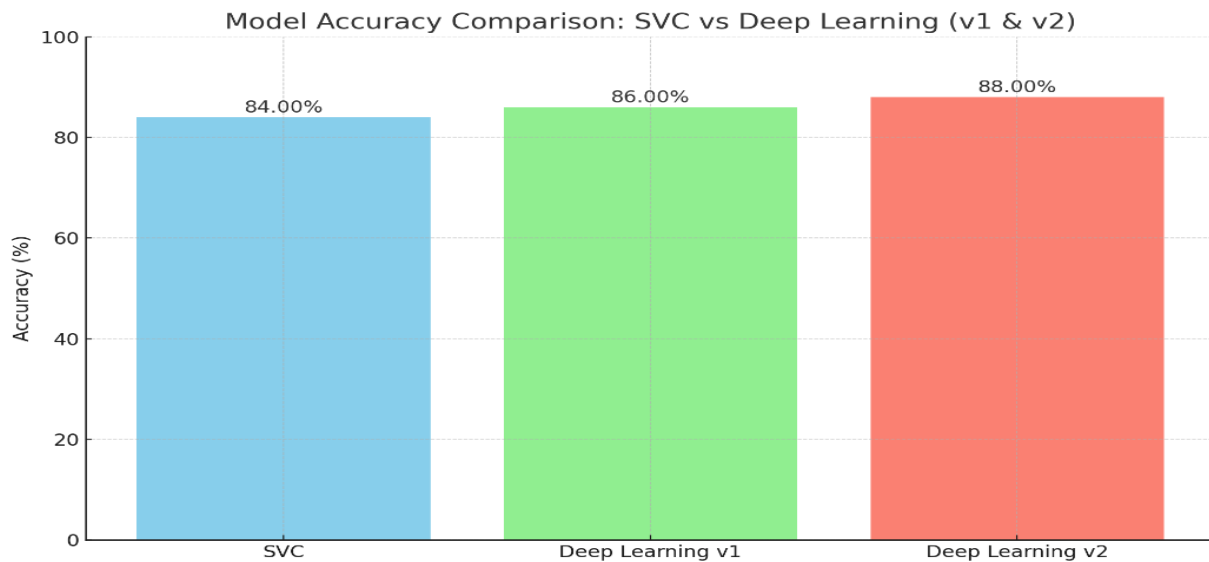


Figure 5: Model Accuracy Comparison SVC v/s Deep Learning

This bar chart indicates how accurate three models are: SVC, Deep Learning v1, and Deep Learning v2. Accuracy refers to how many correct predictions the model has. The results indicate an improvement that is consistent as we progress from SVC to Deep Learning.

With an accuracy of 84%, the SVC model provides a solid baseline against which to compare. Deep Learning v1 performed better, showing that it could recognise more complex patterns in the data with an accuracy of 86%. Even a 2% improvement can make a big difference in sectors like healthcare, where accuracy is essential.

With an accuracy of 88%, Deep Learning v2 produced the best result, demonstrating that greater predictions are produced by improved models. Overall, in general, the graph proves that deep learning models are stronger and more reliable compared to more traditional methods, such as SVC.

## V. CONCLUSION

In this study, NDDs were predicted using ML and DL techniques on a carefully selected dataset of more than 6,000 records. With an accuracy of 84%, the SVC provided a reliable conventional baseline; however, the CNN-inspired DNN outperformed it with an accuracy of 88%. The models' discrimination power and dependability were further evaluated using the ROC curve and AUC metrics, particularly for the DNN. The results show how DL can help with clinical decision-making and early detection of NDDs. Future studies employing bigger datasets and hybrid modelling techniques can improve the predictive models' accuracy and transferability in actual healthcare settings.

## VI. FUTURE WORK

This study could be made even better by using a huge dataset that includes more brain imaging and genetic information, giving a fuller picture of each participant. Adding real-time behaviour monitoring could make the assessments more accurate and up to date. A simple software tool or mobile app could help researchers and doctors use these results more easily in real-life situations. Also, using advanced methods like combining SVM with CNN or other ML techniques could make the model increased accurate and reliable.

**REFERENCES**

- [1]. Song, C., Jiang, Z. Q., Liu, D., & Wu, L. L. (2022). Application and research progress of Machine learning in the diagnosis and treatment of neurodevelopmental disorders in children. *Frontiers in Psychiatry*, 13, 960672. <https://doi.org/10.3389/fpsy.2022.960672>
- [2]. Toki, E. I., Tsoulos, I. G., Santamato, V., & Pange, J. (2024). Machine learning for predicting neurodevelopmental disorders in children. *Applied Sciences*, 14(2), 837.
- [3]. Alam, S., Raja, P., & Gulzar, Y. (2022). Investigation of machine learning methods for early prediction of neurodevelopmental disorders in children. *Wireless communications and mobile computing*, 2022(1), 5766386.
- [4]. <https://www.verywellmind.com/neurodevelopmental-disorders-definition-symptoms-traits-causes-treatment-522131>
- [5]. [Malik, M. A., Faraone, S. V., Michoel, T., & Haavik, J. (2023). Use of big data and machine learning Algorithms to extract possible treatment targets in neurodevelopmental disorders. *Pharmacology & Therapeutics*, 250, 108530.
- [6]. Rahman, M. M., Usman, O. L., Muniyandi, R. C., Sahran, S., Mohamed, S., & Razak, R.A. (2020). A review of machine learning methods of feature selection and classification for autism spectrum disorder. *Brain sciences*, 10(12), 949.
- [7]. Thabtah F. Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfilment. *PLoS One*.2017;12(5): e0178170.
- [8]. Deshpande G, Joseph L. Classification of ADHD using SVM on neuroimaging data. *Front Neurosci*. 2019 13:189.
- [9]. Alam S, Raja P, Gulzar Y. Investigation of machine learning methods for early prediction of neurodevelopmental disorders in children. *Wireless Commun Mob Comput*. 2022; 2022:5766386.
- [10]. Ghiassian S, Greiner R, Jin P, Zhang Y. A supervised machine learning algorithm to predict functional outcomes in ASD based on EEG features. *J Autism Dev Disord*. 2016;46(1):341–52.