International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhanced Autism Severity Prediction: Hybrid Gradient Boosted Tree and Deep Learning Models

R. Ramya[1*] and Dr. S. Panneer Arokiaraj[2]

Submitted: 13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: Autism Spectrum Disorder (ASD) is a condition of developmental disability impacting both behavior and brain functionality. It cannot be diagnosed through medical tests; hence, the diagnosis relies heavily on historical data. Data science models, like Gradient Boosted Trees and Deep Learning, play a crucial role in predicting autism risk by evaluating relevant information and identifying patterns. This paper proposes a novel Hybrid Model that combines the advantages of both Gradient Boosted Tree and Deep Learning models. The aim is to reduce the number of necessary diagnostic tests for autism, thereby offering potential solutions for the healthcare sector. This model achieved an accuracy of 95.52% in predicting the severity of autism using historical adult autism data. The historical patient data used for this study is available on the Kaggle Repository. This perspective highlights the crucial importance of data science in diagnosing healthcare issues.

Keywords: *Autism Historical Dataset, Autism Spectrum Disorder (ASD), Data Science Models, Hybrid for Gradient Boosted Tree and Deep Learning (GBT-DL).*

1. Introduction

Data science, with its potential to enhance efficiency and reduce costs, can significantly transform the healthcare sector, recover patient lifestyles, and save more lives. Autism, a developmental disorder that is rapidly increasing, poses a significant challenge to the healthcare sector, especially since there are no specific medical tests for its diagnosis. The current prevalence of autism has risen by 15%, from 1 in 68 children to 1 in 59 children [1].Diagnosis of autism is primarily based on patient history data and test results [2]. This paper aims to improve the prediction of autism diagnostic test outcomes using the Autism Spectrum Disorder (ASD) dataset from the Kaggle Repository, which has historical data of various symptoms experienced by patients with autism.

- Data science is an interdisciplinary field that extracts valuable insights from data using advanced analytical methods and scientific principles [3]. The process of gaining insights from the Autism dataset involves several steps:
- Data Collection: The first step involves gathering historical raw data on autism from the Kaggle Repository.
- Data Preparation: This crucial phase involves cleaning inconsistent data from the raw autism dataset and establishing techniques for handling missing data fields.
- Data Analysis: Various techniques are used to project and reduce the autism data and to identify its invariant aspects.
- Predictive Modelling: A hybrid model, combining Deep Learning and Gradient Boosted Tree models, is used to extract knowledge and present findings from the Autism dataset.
- Data Visualization: This step involves interpreting the extracted knowledge from the mined patterns of the Autism dataset.

The proposed research aims to predict the severity of autism at the earliest. The novel Hybrid Model of Gradient Boosted Trees and Deep Learning (GBT-DL) can be utilized in the healthcare sector to predict the results of the diagnostic test for Autism with 98.52% accuracy.

This paper is organized as follows:

The literature review section discusses a variety of data science models related to autism. The proposed system section presents an overview of autism, the nature of the dataset, deep learning gradient, and its hybridization. The result and discussion section examines the efficiency, execution time, and F1-Scores of the GBT-DL Model used to predict autism. The paper concludes by discussing the significance of the implementations to research work and future plans for extending this work.

[1]*Research Scholar, Department of Computer Science, Thanthai Periyar Govt. Arts & Science College(Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.*
[2]*Associate Professor, Department of Computer Science, Thanthai Periyar Govt. Arts & Science College(Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.*
**Corresponding author(s). E-mail(s): rgramya94@gmail.com; drpancs@gmail.com;*

2. Review of Literature

Many studies in the field of autism have been conducted to classify, predict, and assess autism risk levels, diagnose the condition, and promote awareness while providing targeted therapeutic interventions. MdDelowar Hossain et al. analyzed autism child datasets using the five most common feature selection strategies to extract fewer features from autism datasets while keeping competitive performance. They found that the ReliefF feature selection method outperformed the others. In their experimental setup, they applied a variety of classification models, incrementally increasing the number of fields, and concluded that Machine Learning surpassed all other classifiers [4].

Hyde et al. developed a web-based application for the quick and accurate identification of Autism across all age groups, leveraging Machine Learning algorithms. They proposed a novel machine learning framework that incorporates datasets specifically designed for autism screening in adults, adolescents, and toddlers [5].

Kaushik et al.'s research underscores the expanded use of machine learning approaches as a supplement to traditional methods. Their study involved building predictive models using Machine Learning. The results indicated that Logistic Regression showed superior accuracy when applied to the autism dataset [6]. Ibrahim Abdulrab Ahmed et al. evaluated a dataset focused on Autism using artificial intelligence models, such as neural networks, deep learning, and a hybrid method. Notably, the hybrid model demonstrated superior results, with the earliest working proposed system utilizing ANN and FFNN models achieving the best overall performance [7].

Rahman & Subashini used five binary Autism classifier models based on Convolutional Neural Networks (CNN). These models used pre-trained architectures, including various classification models. The researchers built and evaluated these models to assess their performance. The results indicated that valuable features related to Autism can be effectively extracted from static face images of a child, suggesting the potential for a quick and accurate Autism screening method [8]. Suman and Sarfaraz introduced various machine learning and deep learning approaches for assessing the severity of autism. Their goal was to optimize parameters for each deep learning and machine learning model and then retrain the models using these optimized parameters to enhance performance. The findings indicated that Deep Learning outperforms other models in terms of performance [9].

Minissi et al. introduced an innovative semi-supervised learning approach based on fuzziness for predicting Autism. Their strategy involved incorporating mislabeled data alongside properly labeled data during the training phase to enhance model reliability. The predictive analysis of this method, when applied to the dataset, demonstrated significant improvements compared to other classification models like ZeroR, Fuzzy MinMax Classifier, Random tree, Fuzzy Data Mining, Naive Bayes, and others [10]. Parlett-Pelleriti et al. developed a model to identify the key factors influencing the Support Vector Machine (SVM) prediction. The outcomes revealed that well-optimized machine learning systems can provide accurate identification regarding an individual’s Autism diagnosis, underscoring the capability of Machine Learning models to achieve accuracy without compromising overall performance [11].

Bala et al. utilized current and previous student grades to apply the WJ-48 algorithm, K-means clustering method, and linear regression to predict future performance [12]. Shorten et al. evaluated various data science models applied to prediction-based datasets. They found that Naive Bayes is best suited for small datasets, while Decision Trees are more effective for large datasets, based on precision, recall, and accuracy metrics obtained using the RapidMiner tool [13].

Moshe et al. employed gradient boosting algorithms, such as XGBoost and LightGBM, to train a predictive model using a training dataset. They adjusted the algorithm’s hyperparameters to optimize its performance [14]. Hosen et al. proposed an effective technique using ReliefF and PCA on several historical disease datasets, achieving excellent results [15].

Nityashree Nadar paper presented a valuable contribution to the field of educational data analysis by proposing a modified XGBoost algorithm for stream-based analysis of student performance data. The paper's innovative approach and practical implications made it a significant contribution to the ongoing efforts to enhance student learning outcomes through data-driven interventions [16]. Ramya and Panneer Arokiaraj, examined a hybrid approach combining Convolutional Neural Networks (CNNs) and Random Forests to improve the prediction of autism severity. The study highlighted how integrating deep learning and ensemble methods yielded more accurate and reliable predictions, providing a valuable tool for clinicians in diagnosing and treating autism spectrum disorders [17].

The literature review section discusses models from various authors that can be used to determine Autism risk factors. The proposed system section conducts a detailed study about Autism, the dataset, and the hybridization of Deep Learning and Gradient Boosted Tree models used for autism prediction.

3. Proposed System

The objective is to develop and implement a Hybrid Model of Gradient Boosted Trees and Deep Learning (GBT-DL) for Autism detection across adult age groups using Data Science models. By utilizing collected ASD datasets, accurate results can be ensured within a minimal timeframe.

3.1. Dataset Collection

The Autism dataset is evidently associated with autism screening, featuring diverse attributes related to individuals and their characteristics. The attributes cover scores, age, gender, ethnicity, jundice and autism history, country of residence, app usage history, result scores, age description, relation, and a class denoting the presence or absence of Autism Spectrum Disorder (ASD). The dataset instances, delineated under the '@data' tag, detail individual data points with values separated by commas, and the final value designates the ASD class (YES or NO). Categorical attributes, like gender and ethnicity, exhibit non-numeric values, while numeric attributes such as age and result provide quantitative data. Overall, the dataset seems tailored for autism screening, potentially aiming to develop a predictive model for identifying ASD based on the specified features.

3.2. Work Flow of GBT-DL Model

The workflow of the Hybrid data science model is outlined in Figure 1 and involves four key phases. Initially, raw Autism data is collected from the Kaggle Repository website, and it undergoes a comprehensive Data Processing phase. This involves activities such as data discovery, cleansing, transformation, validation, and ultimately storing the refined data. In the second phase, features are extracted from the processed datasets to enhance the accuracy of predictions. These selected features play a crucial role in informing the predictive model. The third phase is dedicated to developing the GBT-DL prediction model, wherein the data science model is trained using extensive training data. Finally, the fourth phase involves evaluating the GBT-DL prediction model and the resulting analysis and outcomes are visualized. This entire workflow ensures a systematic approach from data preparation to model development and evaluation, ultimately contributing to the accurate prediction of Autism based on the provided data measures.

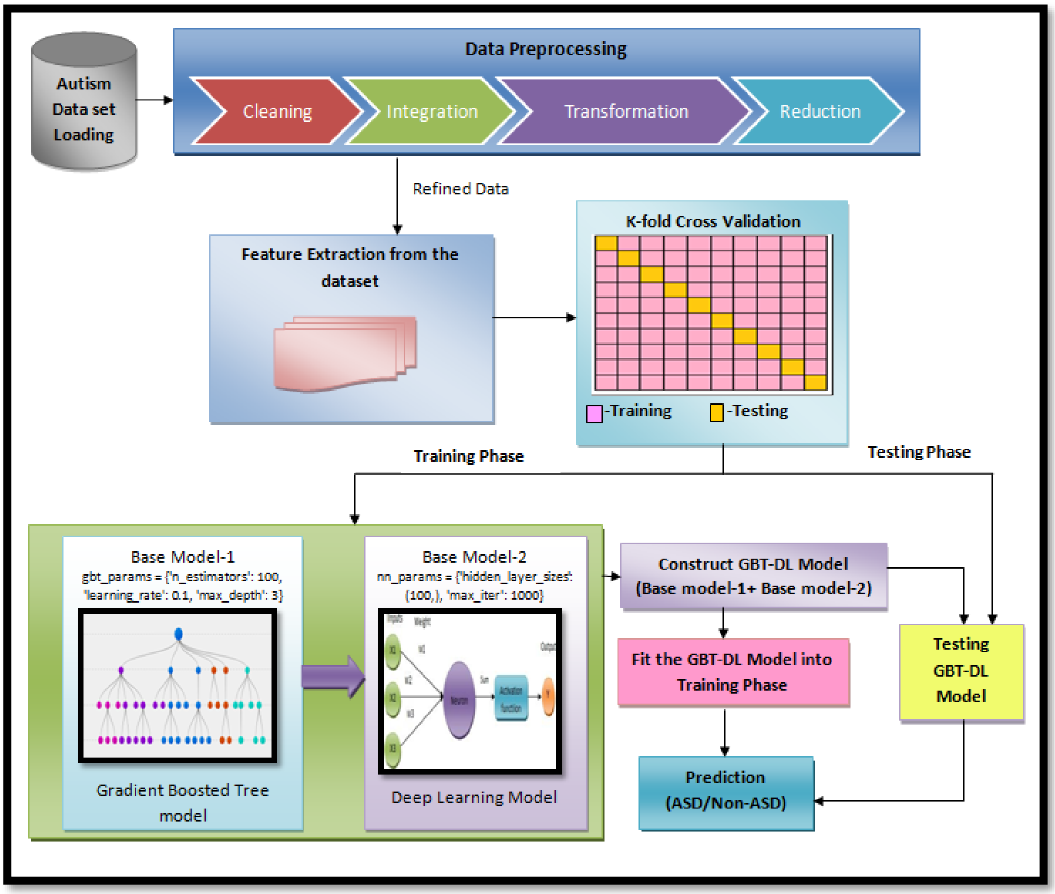

Fig 1: Workflow Architecture of the GBT-DL Model

The research consists of four phases:

- The first phase is data preparation: data collection, discovery, cleansing, transformation, validation, and storing the data.

- The second phase is extracting features from the datasets.
- The third phase consists of developing the GBT-DL prediction model.
- Finally, the GBT-DL prediction model is evaluated, and the results and analysis are visualized.

3.3. Data Preparation

The data preparation phase is a crucial step within the data science workflow, involving the conversion of raw data into a well-organized, structured format suitable for analysis and model construction. The following sections will outline several typical steps integral to the data preparation process.

The implementation phases are discussed in subsequent sections:

3.3.1. Discover and Access

The first step of the operation is to gather, discover, and access the ASD dataset in CSV file format. Preprocessing is the most crucial step of implementation because inconsistent historical Autism dataset should not be used directly. The Preprocessing dataset will be noiseless and clear data.

3.3.2. Cleanse and Transform Data

This phase entails refining and converting the initial data into a structure conducive to subsequent analysis and model construction. This encompasses addressing missing values, outliers, and discrepancies within the dataset. Data preparation for analysis entails properly formatting the data. This includes adjusting data types, standardizing units of measurement, and scaling variables to ensure they fall within a comparable range. The attribute of the autism dataset are normalized to ensure they have similar scales. One such normalization method is z-score normalization, where field values are preprocessed based on the standardization of the dataset.

$$\mathbf{Dscaled} = \frac{\mathbf{D-mean}}{\mathbf{Sd}} \quad (1)$$

3.3.3. Validation and Enrich Data

This Validation operator performs thorough data validation checks to ensure the data is accurate, complete and consistent. This involves checking for any anomalies or unexpected patterns in the data. Finally store the ASD dataset. The Validation operator performs comprehensive data validation checks to ensure the data is accurate, complete, and consistent. This involves checking for any anomalies or unexpected patterns in the data. The final step is to store the Autism Spectrum Disorder (ASD) dataset.

3.4. Proposed model

3.4.1. Deep Learning

Deep Learning models find application across diverse healthcare domains, including the prediction and diagnosis of autism. Given the intricate nature of ASD as a neurodevelopmental condition, deep learning models prove instrumental in analyzing vast datasets. They excel at extracting meaningful patterns, potentially contributing to the autism prediction. In the hybrid model, this capability is harnessed during the learning stage within the training phase. Deep Learning models are used across diverse healthcare domains, including the prediction and diagnosis of autism. Given the complex nature of ASD as a neurodevelopmental condition, deep learning models are instrumental in analyzing large datasets. They excel at extracting meaningful patterns, potentially contributing to autism prediction. In the hybrid model, this capability is harnessed during the learning stage within the training phase.

Deep Learning is a subset of machine learning that utilizes a hierarchical level of artificial neural networks. It involves feeding data through networks of algorithmically simulated neurons, which then output predictions about the given data. The foundational elements of a Deep Learning model are inputs, weights, bias or threshold, and an output.

The formulaic representation of deep learning can be expressed as:

$$\sum_{i=1}^{n} \mathbf{w_i x_i} + \mathbf{bias} = \mathbf{w_1x_1} + \mathbf{w_2x_2} + \mathbf{w_3x_3} + \mathbf{bias}$$

Where, x1, x2,…xn is an input signal that can come from the Autism raw dataset and these connections are called weights such as w1,w2,…wn. The output signal is called y. The GBT-DL Model is implemented by using Deep Learning methods as follows:

- **Adaptive rate:** The executed adaptive learning rate models automatically hybrid the outcomes. Adaptive rate: The executed adaptive learning rate models automatically hybridize the outcomes of momentum training and learning rate annealing to avoid time-consuming convergence.
- **Loss Function:** The model aims to minimize a specific loss (error) function. Cross entropy is the recommended choice, especially when dealing with class labels as the mean of values, particularly for nonessential data. This function robustly penalizes errors in predicting the exact class label.
- **Missing Values Handling:** Addressing missing values involves employing Mean Imputation, where a

missing value is replaced with the mean value. This method is applied to ensure robust data handling.

3.4.2 Gradient Boosted Tree

Gradient Boosted Models, such as XGBoost and LightGBM, are powerful data science tools that have shown significant success in various domains, including autism prediction. These models effectively address challenges by generalizing tree boosting to mitigate associated issues. In the context of prediction formulas in Gradient Boost, the most common transformation is often expressed through the following formula:

$$\sigma = \frac{\sum \text{Residual}}{\sum[\text{PreviousProb} * (1 - \text{PerviouProb})]} \quad (3)$$

In this equation, the numerator represents the sum of residuals within a specific leaf, and the denominator is the reciprocal of the quantity calculated as $(1 - \text{previous prediction probability})$ $(1 - \text{previous prediction probability})$ multiplied by the previous prediction probability for each residual.

3.4.3. GBT-DL Model

This GBT-DL model combines the features of both the Deep Learning and Gradient Boosted Tree to predict the risk level of Autism. Initially Gradient Boosted Tree and Deep Learning Model were selected to construct the hybrid model. At the start, the root node consists of the overall ASD dataset. The following Figure 2 presents the pseudo code for this proposed work.

Algorithm: GBT-DL Model (Gradient Boosted Tree and Deep Learning)

Input:

- *Training dataset (X_train, y_train)*
- *Testing dataset (X_test)*
- *GBT hyperparameters (*gbt_params*)*
- *Deep Learning hyperparameters (DL_params)*

Output:

- *Autism predictions on the dataset*

1. *Split the training dataset into a training set* $(Xtrain, ytrain)$ *and a validation set*$(Xval, yval)$.
2. *Train the GBT model* $fGBT$ *on the training set using the specified hyperparameters* gbt_params*:*
$$fGBT = TrainGBT(Xtrain, ytrain, gbt\_params)$$
3. *Train the DL model* fNN *on the training set using the specified hyperparameters* nn_params*:*
$$fNN = TrainDL(Xtrain, ytrain, nn\_params)$$
4. *Generate predictions on the validation set using both the GBT and DL models:*
$$y\hat{}val\_GBT = fGBT(Xval)$$
$$y\hat{}val\_DL = fDL(Xval)$$
5. *Stack the predictions from the GBT and DL models as new features:*
$$Xval\_stacked = [y\hat{}val\_GBT, y\hat{}val\_DL]$$
6. *Train a meta-model* $fmeta$ *on the stacked predictions using the validation set labels* $yval$*:*
$$fmeta = TrainMetaModel(Xval\_stacked, yval)$$
7. *Train the GBT and DL models on the entire training dataset:*
$$fGBT\_final = TrainGBT(Xtrain, ytrain, gbt\_params)$$
$$fDL\_final = TrainDL(Xtrain, ytrain, dl\_params)$$
8. *Generate predictions on the testing dataset using both the GBT and DL models:*
$$y\hat{}test\_GBT = fGBT\_final(Xtest)$$
$$y\hat{}test\_DL = fDL\_final(Xtest)$$
9. *Stack the predictions from the GBT and DL models as new features for the testing dataset:*
$$Xtest\_stacked = [y\hat{}test\_GBT, y\hat{}test\_DL]$$
10. *Use the trained meta-model to make final predictions on the testing dataset based on the stacked features:*
$$y\hat{}test\_final = fmeta(Xtest\_stacked)$$
11. *Output the final predictions from the hybrid model* $y\hat{}test\_final$

Fig 2: Pseudo code for the Proposed GBT-DL Model

In the initial phase of developing our predictive model, the training dataset, represented as (X_train, y_train), undergoes meticulous division into two distinct subsets: one for actual model training and the other for evaluating and validating model performance. The initial model trained is a Gradient Boosted Tree (GBT), recognized for constructing a series of decision trees to make precise predictions. Following this, a Deep Learning (DL) model is trained on the same dataset, utilizing its deep learning capabilities to discern complex patterns within the data. Predictions are subsequently generated on a validation set using both the GBT and DL models to assess their individual performances on previously unseen data.

To enhance predictive capability, the next step involves merging predictions from the GBT and DL models into a new feature set. A meta-model, illustrated by Logistic Regression, is then introduced to learn from this merged prediction set and forecast the target variable based on validation set labels. This meta-model aims to capture intricate relationships between GBT and DL model predictions and the actual target values. In the subsequent phase, both the GBT and DL models are retrained using the entire original training dataset to ensure optimal generalization with access to all available data.

Approaching the final stage, predictions are generated on the testing dataset using both the GBT and DL models. Similar to the validation phase, these models' predictions are combined to create a new set of features for the testing dataset. The trained meta-model is then utilized to make ultimate predictions on the testing dataset based on these combined features. This process yields final predictions from the hybrid model, effectively amalgamating the strengths of Gradient Boosted Trees and Neural Networks. The overarching objective is to leverage the complementary attributes of these models through stacking and a meta-model, ultimately enhancing predictive performance, particularly in predicting Autism in this context.

Table1: Parameter Tuned in GBT-DL Model

Parameters for Deep Learning	***Parameters for Gradient Boosted Tree***
Activation=ExponentialRectifierAdaptive *RE= 1.0E-9* *Adaptive RRHO= 0.98* *Learning 1=1.0E-7* *Learning 2=0.1* *Maximum weight1=9.0* *Loss function = Cross Entropy* *Distribution function=Multinomial* *Absent values handling=MeanImputation* *(Missingvaluesarereplacedwiththemeanvalue)*	*No.of trees=54* *Max. Depth=6* *Minimum Rows=9.0* *Min Split validation = 1.0E7 No. of Bins=23* *LR(Learning Rate) =0.234* *SR(Sample Rate)=0.99 Distribution=multinomial*

The GBT-DL Model integrates features from both Deep Learning and Gradient Boosted Tree. In Table1Tuning has been performed on the parameters related to Deep Learning and Gradient Boosted Tree features, as outlined.

4. Analysis and Results

The results analysis section delves into the comprehensive evaluation of the proposed model, aiming to enhance the prediction of autism severity through the integration of Hybrid Gradient Boosted Trees (GBT) and Deep Learning (DL) techniques. This innovative hybrid model capitalizes on the strengths of both GBT and DL methodologies to create a robust predictive framework. The evaluation is conducted across varying K-fold Cross Validation scenarios, specifically exploring different values of K, and spans both the Training and Testing Phases. This section dissects key performance metrics, such as accuracy, precision, recall, F1-Score, and the Kappa statistic, providing a nuanced understanding of the model's efficacy in predicting autism severity. Through meticulous analysis, this section aims to unveil insights into the model's generalization capabilities, trade-offs, and overall performance, contributing valuable findings to the field of autism severity prediction.

Table 3: Performance Analysis of GBT-DL Model

K-fold Cross Validation	*GBT-DL Model - Performance metrics*							
	Training Phase (90%)				*Testing Phase (10%)*			
	K=3	***K=5***	***K=10***	***K=15***	***K=3***	***K=5***	***K=10***	***K=15***
Accuracy	*89.67*	*92.67*	***95.52***	*94.72*	*90.37*	*92.98*	***95.62***	*94.54*
Precision	*89.98*	*92.42*	***94.67***	*94.32*	*90.13*	*92.62*	***94.88***	*94.32*
Recall	*89.74*	*91.98*	***95.27***	*93.89*	*90.56*	*92.19*	***95.48***	*93.45*
F1-Score	*10.12*	*8.42*	***4.48***	*5.83*	*9.89*	*7.42*	***4.38***	*5.73*
Kappa statistic	*0.84*	*0.90*	***0.95***	*0.93*	*0.86*	*0.91*	***0.95***	*0.93*

Overall, Table 3 represents the GBT-DL Model showcases robust performance, particularly with a marked improvement when K=10. This suggests commendable generalization capabilities and a harmonious balance between precision and recall, as corroborated by the F1-Score. The consistently high Kappa statistic values further underscore the model's reliability in predicting outcomes. However, it is imperative to conscientiously consider potential trade-offs and generalization aspects when determining the optimal K value for practical deployment.

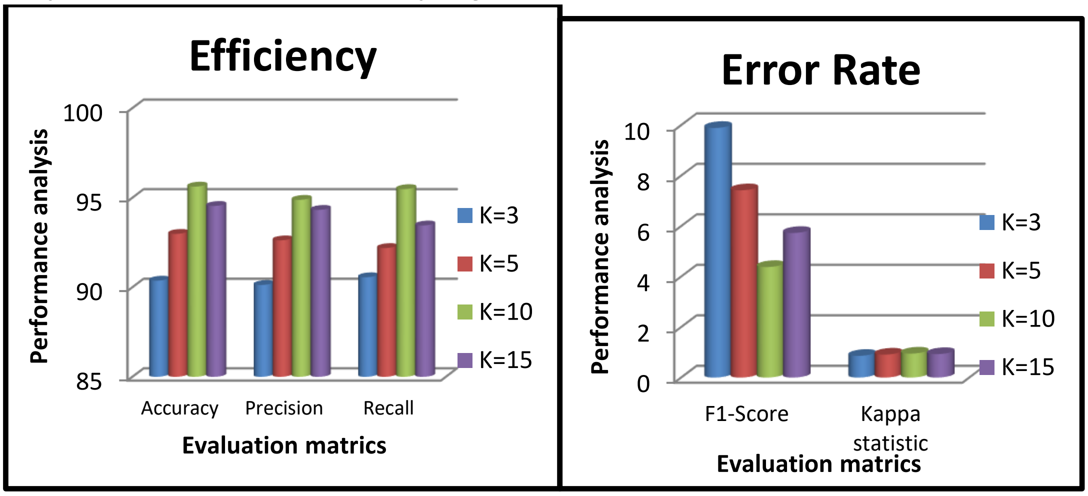

Fig 3: Performance analysis of the Proposed GBT-DL Model

4.1. Comparative Analysis:

This section primarily focuses on assessing the performance of the hybrid model proposed, comparing Apply it alongside diverse models such as Gradient Boosted Tree and Deep Learning using the ASD Dataset.

The hybrid model exhibits notable accuracy and lower F1-Score due to the amalgamation of Deep Learning and Gradient Boosted Tree models, as detailed in Table 2. The evaluation of the proposed approach is gauged based on Accuracy, Recall, Precision, F1-Score, and Kappa Statistic.

Table 2: Performance Comparison of GBT-DL Model with Other Models

Performance analysis	***GBT***			***DL***			***GBT-DL***		
	K=3	***K=5***	***K=10***	***K=3***	***K=3***	*K=10*	*K=3*	*K=5*	*K=10*
Accuracy	*87.67*	*88.23*	*89.55*	*72.45*	*73.67*	*74.91*	*90.37*	*92.98*	***95.62***
Precision	*87.23*	*87.98*	*88.25*	*71.23*	*73.98*	*73.43*	*90.13*	*92.62*	***94.88***
Recall	*87.12*	*87.89*	*88.26*	*71.46*	*73.12*	*72.57*	*90.56*	*92.19*	***95.48***
F1-Score	*13.87*	*12.34*	*11.48*	*26.78*	*25.07*	*23.09*	*9.89*	*7.42*	***4.38***
Kappa statistic	*0.86*	*0.87*	*0.88*	*0.73*	*0.74*	*0.75*	*0.86*	*0.91*	***0.95***

Table 2 identified that the new GBT-DL hybrid model achieved 95.62% accuracy with a F1-Score rate of 4.48% in a minimum execution time of 0.07 seconds over the other models. Also, it excelled in achieving 94.88, 95.48, and 0.95 effectiveness in Recall, Precision, and Kappa Statistic, respectively.

4.1.1. Efficiency

Figure 4 illustrates the graphical representation of Table 2 illustrates the efficacy of the suggested hybrid model in contrast to alternative Data Science models. This presentation offers a concise and readable depiction of the accuracy reached by the proposed model.

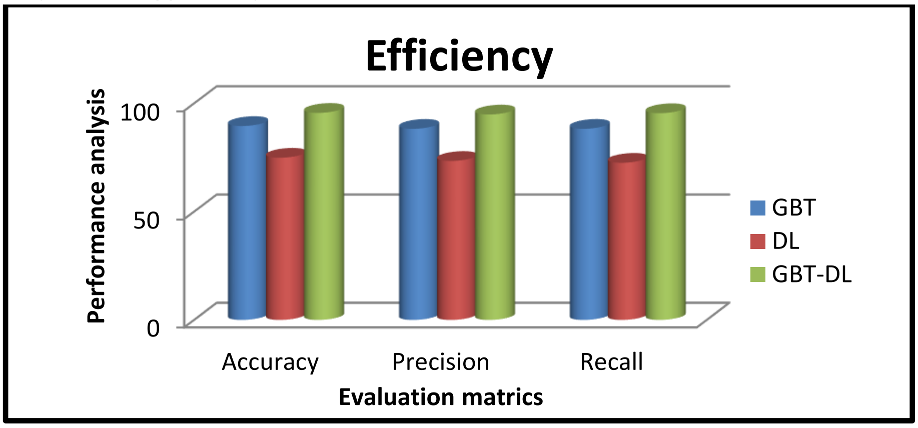

Figure 4: Performance Comparison of GBT-DL Model Based on Accuracy, Recall and Precision with Other Models.

4.1.2. Error Rate

The F1-Score of the GBT-DL Model depends on the number of samples incorrectly classified and the kappa statistic measures the random accuracy. Figure 6 defines F1-Score and kappa statistics of the proposed hybrid model with other Data Science models.

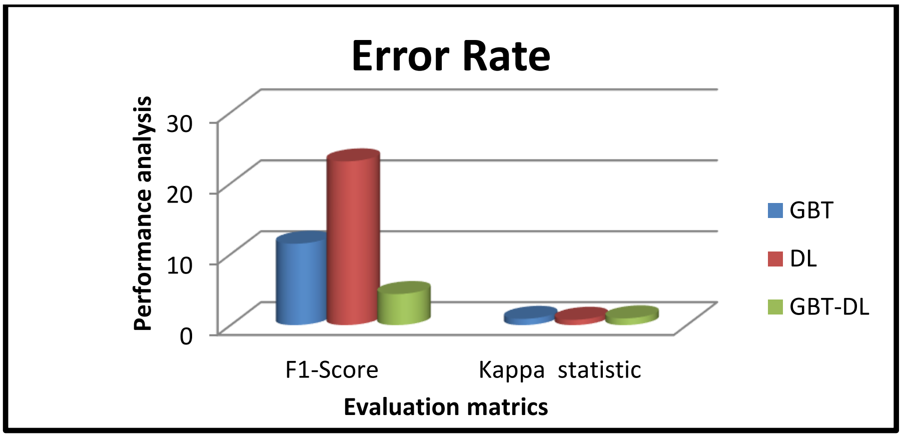

Fig 6: F1-Score and Kappa Statistic Comparison of the GBT-DL Model with Other Data Science Models

The above figure 5 and figure 6 provides a comparative analysis of the performance metrics for different prediction models in the context of enhancing autism severity prediction. Three models are considered: Gradient Boosted Trees (GBT), Deep Learning (DL), and a hybrid model combining both (GBT-DL). The evaluation is done with varying values of K, representing different scenarios or experimental conditions. The performance metrics include Accuracy, Precision, Recall, F1-Score, and Kappa statistic. For each model and K value, the corresponding values of these metrics are presented. The hybrid GBT-DL model generally outperforms the individual GBT and DL models across all K values in terms of accuracy, precision, recall, F1-Score, and Kappa statistic. For instance, at K=10, the hybrid model achieves the highest accuracy (95.62%), precision (94.88%), recall (95.48%), F1-Score (4.38%), and Kappa statistic (0.95) compared to GBT and DL. The table provides a comprehensive overview of how the hybrid model combines the strengths of GBT and DL to enhance the prediction of autism severity.

5. Conclusion

In conclusion, this research has demonstrated the significant potential of data science, particularly the Hybrid Model of Gradient Boosted Trees and Deep Learning (GBT-DL), in predicting Autism Spectrum Disorder (ASD). The GBT-DL model, which combines the strengths of both Deep Learning and Gradient Boosted Tree models, has shown promising results in terms of accuracy and efficiency. The model achieved an accuracy of 95.52% in predicting the severity of autism using historical adult autism data, outperforming other existing models such as Gradient Boosted Tree and Deep

Learning. This underscores the potential of machine learning and artificial intelligence in healthcare diagnostics, particularly in conditions like ASD where diagnosis is primarily based on patient history data and test results. However, the research also highlighted the limitation of the model, which is its applicability to a adult age group of autism prediction. Future work will aim to expand the model to accommodate a different age group of autism prediction, thereby enhancing its predictive power and utility in the healthcare sector. This research contributes to the growing body of literature on the application of data science in healthcare and provides a foundation for future studies aiming to leverage machine learning and artificial intelligence in diagnosing and predicting various health conditions. The findings of this study underscore the transformative potential of data science in healthcare, paving the way for more accurate, efficient, and timely diagnosis and treatment of conditions like Autism Spectrum Disorder.

References

[1] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, A Machine Learning Approach to Predict Autism Spectrum Disorder, *in International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Cox'sBazar, Bangladesh, 2019, pp. 1-6. DOI: 10.1109/ECACE.2019.8679454.

[2] Kaggle. (2020). Autism Spectrum Disorder Screening Dataset. Kaggle. *https://www.kaggle.com/example/autism-dataset.*

[3] S. Selvaraj, P. Palanisamy, S. Parveen, and Monisha, Autism Spectrum Disorder Prediction Using Machine Learning Algorithms, in Computational Vision and Bio-Inspired Computing, *Advances in Intelligent Systems and Computing*, vol. 1108, Springer, 2020. DOI: https://doi.org/10.1007/978-3-030-37218-7_56.

[4] Md Kabir Hossain, Ashad Anwar, and Adnan Islam, Detecting Autism Spectrum Disorder using Machine Learning Techniques, *in Health Information Science and Systems,* vol. 9, 2021. DOI: https://doi.org/10.1007/s13755-021-00145-9.

[5] K. Hyde, M.N. Novack, N. LaHaye, et al., Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review, in *Rev J Autism Dev Disord*, vol. 6, 2019. DOI: https://doi.org/10.1007/s40489-019-00158-x.

[6] Kaushik Vakadkar, Diya Purkayastha, and Deepa Krishnan, Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques, in *SN Computer Science*, 2021, vol. 2, p. 386. DOI: https://doi.org/10.1007/s42979-021-00776-5.

[7] I.A. Ahmed, E.M. Senan, T.H. Rassem, M.A.H. Ali, H.S.A. Shatnawi, S.M. Alwazer, M. Alshahrani, Eye Tracking-Based Diagnosis and Early Detection of Autism Spectrum Disorder Using Machine Learning and Deep Learning Techniques, in *Electronics*, vol. 11, 2022, p. 530. DOI: https://doi.org/10.3390/electronics11040530.

[8] Mujeeb Rahman KK and Subashini MM, Identification of Autism in Children Using Static Facial Features and Deep Neural Networks, in *Brain Sciences*, vol. 12, no. 1, 2022, p. 94. DOI: https://doi.org/10.3390/brainsci12010094.

[9] S. Raj and S. Masood, Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, in *Procedia Computer Science*, vol. 167, 2020, pp. 994-1004. DOI: https://doi.org/10.1016/j.procs.2020.03.399.

[10] M.E. Minissi, I.A. Chicchi Giglioli, F. Mantovani, et al., Assessment of the Autism Spectrum Disorder Based on Machine Learning and Social Visual Attention: A Systematic Review, in *J Autism Dev Disord*, vol. 52, 2022, pp. 2187-2202. DOI: https://doi.org/10.1007/s10803-021-05106-5.

[11] C.M. Parlett-Pelleriti, E. Stevens, D. Dixon, et al., Applications of Unsupervised Machine Learning in Autism Spectrum Disorder Research: a Review, in Rev J Autism Dev Disord, 2022. DOI: https://doi.org/10.1007/s40489-021-00299-y.

[12] Mousumi Bala, Mohammad Hanif Ali, Md. Shahriare Satu, Khondokar Fida Hasan, and Mohammad Ali Moni, Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder, in *Algorithms*, vol. 15, no. 5, 2022, p. 166. DOI: https://doi.org/10.3390/a15050166.

[13] C. Shorten, T.M. Khoshgoftaar, and B. Furht, Deep Learning applications for COVID-19, in *J Big Data*, vol. 8, 2021, p. 18. DOI: https://doi.org/10.1186/s40537-020-00392-9.

[14] Moshe Sipper, Jason H. Moore, AddGBoost: A gradient boosting-style algorithm based on strong learners, in *Machine Learning with Applications*, Volume 7, 2022, p. 100243. DOI: https://doi.org/10.1016/j.mlwa.2021.100243.

[15] Md Saikat Hosen and Ruhul Amin, Significant of Gradient Boosting Algorithm in Data Management System, in *Engineering International*, vol. 9, 2021. DOI: https://doi.org/10.18034/ei.v9i2.559.

[16] Nityashree Nadar. Enhancing student performance prediction through stream-based analysis dataset using modified XGBoost algorithm. *International Journal on Information Technologies and Security*, vol.15, no.2, 2023, pp. 75-86. https://doi.org/10.59035/KNUG1085.

[17] R. Ramya, S. Panneer Arokiaraj. Enhancing autism severity prediction: A fusion of convolutional neural networks and random forest model. *International Journal on Information Technologies and Security*, vol.16 , no.2, 2024, pp. 51-62. https://doi.org/10.59035/VNWF2548.