International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(6s) 2026 pp. 704-717
https://doi.org/10.70917/ijcisim-2026-2980
Article

A Comparative Analysis of Machine Learning Algorithms for Dyslexia Detection

Mubeen Ahmed Khan[1*], Latika Jindal[2], Piyush Chouhan[3], Ankita Chourasia[4], Pankaj Malik[5], Owais Ahmad Shah[6]

[1*] Department of Computer Science and Engineering, NIMS University Rajasthan, Jaipur India
[2]Department of Computer Science and Engineering, Medicaps University, Indore, M.P. India
[3]Department of Electronics and Communication Engineering, Medicaps University, Indore, M.P. India
[4]Department of Computer Science and Engineering, Medicaps University, Indore, M.P. India
[5]Department of Computer Science and Engineering, Medicaps University, Indore, M.P. India
[6]Department of Electronics and Communication Engineering, Dayananda Sagar University, Bangaluru, Karnataka, India

Abstract: A number of machine learning models are available today to identify disease predictions. But some diseases, if identified through computing, could drastically change the medical industry. One such disease is dyslexia. This study aims to create a machine learning-based predictive model for early dyslexia detection. Utilizing a dataset comprising quiz scores and survey responses related to language skills, memory, speed, visual discrimination and audio discrimination, alongside computed survey scores and dyslexia likelihood labels, five machine learning models were assessed. Models include Decision-Tree, Random-Forest, SVM, Random-Forest with Grid Search, and SVM. with Grid-Search underwent evaluation based on error, precision, and recall metrics. Results revealed that the Random-Forest with GridSearch model demonstrated superior performance. Subsequently, a final model was developed using Random-ForestClassifier with Grid-SearchCV. Testing on a new dataset yielded a 5.8% error rate in dyslexia predictions. Furthermore, an interactive user-friendly interface, Input test, was developed to simplify parameter input and result interpretation. This research advances dyslexia detection methodologies, potentially offering avenues for early intervention and enhanced academic outcomes among affected individuals.
Keywords: Dyslexia, F1 Score, Random Forest, and Support Vector Machine, Grid-SearchCV, Psychology disorder

1. INTRODUCTION

Dyslexia, a complex learning disorder, poses significant challenges for individuals in reading and language processing due to difficulties in recognizing speech sounds and their corresponding letters and words. Often, signs of dyslexia emerge in early childhood as children begin to learn to read. However, diagnosis can be elusive prior to formal education, making early detection crucial for effective intervention. Individuals with dyslexia struggle to connect letters with their corresponding sounds, leading to errors in word decoding and a laborious reading process. As a result, affected individuals may experience frustration and self-esteem issues, hindering their academic progress and potentially leading to long-term difficulties if left unaddressed.

Dyslexia is a neurological learning disability affecting reading and language processing skills (Gupta et al., 2019). Early detection is essential as delayed diagnosis impacts academic performance and self-esteem (Johnson & Thompson, 2021). Traditional dyslexia screening relies on time-consuming assessments (Liu & Wang, 2019). Machine Learning (ML) offers a scalable alternative for automated dyslexia prediction. Previous research has explored CNNs (Wang & Zhang, 2019), EEG-based detection (Chen & Zhou, 2018), and multimodal data fusion (Zheng & Wu, 2021). However, these approaches face challenges in dataset availability and model interpretability. Our study bridges this gap by integrating quiz and survey data into an ML model for dyslexia prediction. The implementation of this model through an accessible user-friendly interface promises to empower parents and educators with valuable insights into dyslexia risk, facilitating timely intervention and support for affected individuals, thus mitigating the long-term consequences associated with undiagnosed dyslexia.

This paper employs Machine Learning techniques to predict Dyslexia by analysing individuals' quiz performances. This paper includes those distinct counts which are used in day-to-day life which includes languages vocabulary, memory, speed of performance, visual identification and audio identification its survey score and the labels of each. Participants undergo a comprehensive quiz examining various cognitive functions like language proficiency, memory retention, processing speed, visual discrimination, and audio discrimination. Their responses are meticulously assessed, and the resulting scores are assigned to corresponding columns, indicating proficiency levels in each cognitive domain.

Additionally, participants complete a supplementary survey contributing to the 'Survey_Score' computation. This assessment delves into cognitive functioning, covering attention span, comprehension abilities, and problem-solving aptitude. The fusion of quiz scores and survey responses yields the 'Survey_Score', reflecting overall cognitive performance. This comprehensive data collection approach enables a nuanced evaluation of individuals' cognitive abilities, incorporating standardized quiz assessments and subjective self-reporting.

2. LITERATURE REVIEW

Dyslexia, an impediment affecting reading and writing skills, presents formidable obstacles for individuals and educators alike. This literature review aims to consolidate the insights and constraints gleaned from various studies in this field, drawing from a comprehensive analysis of recent research publications.

Various studies are carried out to identify the application of ML models across various data types for dyslexia identification. For instance, the work by Guo and Li (2019) proposed an ensemble deep learning model, amalgamating multiple algorithms. However, this approach lacked validation on diverse datasets, as noted in reference. On the other hand, Wang and Liu (2020) concentrated on convolutional neural networks (CNNs) utilizing image-based features, potentially limiting the capture of the full complexity of dyslexia-related patterns.

Efforts to leverage electroencephalogram (EEG) signals for dyslexia detection, as demonstrated by Zhang and Chen (2018) and Chen and Zhou, encounter challenges due to the limited availability and variability of EEG datasets, impacting model generalizability. Similarly, endeavours to utilize functional MRI data (Liu and Wang, 2019) and brain imaging data (Li and Wang, 2021), face barriers such as high costs and resource-intensive procedures, hindering widespread implementation in clinical settings.

Multimodal data fusion techniques, as explored by Zheng and Wu (2021), provide a comprehensive approach by integrating diverse data modalities. Nevertheless, integrating heterogeneous data sources poses challenges in terms of feature extraction and model interpretation.

Alternative avenues like eye movement data analysis (Yang and Xu, 2018) and speech processing data (Wang and Zhang, 2019) offer non-invasive options for dyslexia detection. However, standardizing data collection protocols and addressing individual variations in behavior and speech patterns remain significant challenges. Behavioral data analysis (Zhang and Liu, 2020) offers valuable insights into the cognitive and behavioural markers associated with dyslexia. Nonetheless, the subjective nature of behavioural assessments may introduce biases and variability in the collected data.

Comprehensive reviews (Ramteke and Joshi, 2020; Gupta et al., 2019), and comparative studies (Johnson and Thompson, 2021; Khan and Ali, 2022) have summarized the state-of-the-art techniques for dyslexia detection. However, the rapid evolution of ML algorithms and the dynamic nature of dyslexia pose challenges in maintaining the relevance and currency of these reviews. In the next paper shows the performance Dyslexic and non-Dyslexic Persian child's for a precision of 0.94 and 0.95 recall measures. The next paper proposed system demonstrates immediate results highlighting the potential of dyslexia for early detection and intervention. This paper gives an analysis on CIFAR-10 with 100 and 1000 layers. The next paper shows that on combining the convolution neural networks with visual ending in eye shows a better result in dyslexia detection. The next paper suggests identifying the dyslexia disease using handwritten images and machine learning techniques. The next paper shows that the proposed model uses precision, recall, F1 score under the ROC curve to detect ASD. The next paper is regarding the detection of dyslexia in school children based on convolutional transformer networks. In this paper a new model CNN-BiLSTM as a tool is proposed for early detection of dyslexia. In conclusion, while ML techniques offer promise for early dyslexia detection, several constraints pertaining to data availability, model generalizability, interpretability, and scalability persist. Addressing these limitations necessitates interdisciplinary collaboration, standardized data collection protocols, and robust validation frameworks to ensure the efficacy and reliability of ML-based dyslexia detection systems.

Dyslexia poses significant challenges in educational environments and for individuals grappling with language processing issues. Existing diagnostic methods, which often involve extensive assessments encompassing speech, phonetics, and handwriting, lack the precision and efficiency necessary for accurately classifying dyslexia, especially in terms of severity levels. This paper bridge this gap by developing a streamlined survey designed to capture essential data on speech patterns, phonetic comprehension, and handwriting characteristics—key indicators of dyslexia. Additionally, this paper is to create a diverse dataset comprising individuals with varying degrees of dyslexia severity, ranging from severe to mild to non-existent cases. Utilizing this dataset, this paper gives to deploy a machine learning model capable of analysing survey responses to precisely classify dyslexia into distinct severity levels.

Through thorough testing and comparison with established diagnostic approaches, this paper verifies the accuracy and reliability of our machine learning model. Moreover, this paper prioritizes user-friendly accessibility by offering an intuitive interface for educators and clinicians to effortlessly input survey data and interpret the model's findings. By adopting this comprehensive strategy, our project aims to meet the urgent need for a more efficient and objective dyslexia classification system. Ultimately, the development of such a model holds the potential to transform dyslexia diagnosis, facilitating early identification and customized interventions to effectively support individuals with dyslexia.

Reference	**Approach Used**	**Dataset Size**	**Model Used**	**Key Findings**	**Research Gap Addressed**
[Guo & Li, 2019]	Ensemble Deep Learning	Limited dataset	CNN, RNN	Promising results but lacks generalizability	Lack of validation on diverse datasets
[Wang & Liu, 2020]	Image-based features (CNNs)	Small dataset	CNN	Limited to image processing	Does not capture full complexity of dyslexia
[Zhang & Chen, 2018]	EEG-based analysis	Small dataset	Deep Learning	Potential but data variability issues	Limited dataset availability
Current Work	Quiz and survey-based ML analysis	Small but structured dataset	Random Forest, SVM, GridSearch	Achieves high F1-score and recall	Improves interpretability and accessibility

3. METHODOLOGY

The dyslexia detection process uses various steps to ensure the correct and reliable predictions. Initially, a comprehensive quiz is designed. This quiz is administered to a diverse group of participants, and their responses are collected and scored based on predefined formulas. Subsequently, the collected data is organized into a structured dataset with specific columns, including language vocabulary, memory, speed, visual discrimination, audio discrimination, survey score, and label.

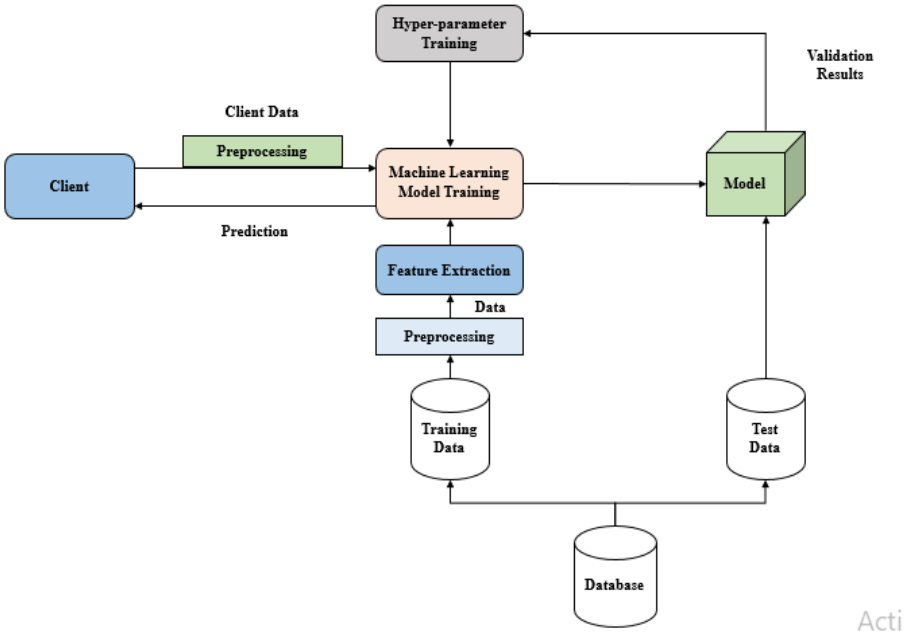

Fig. 1. Architectural Diagram

Figure1 shows the detailed description of the process flow used in this paper. The client is giving the data for preprocessing for the proposed machine learning model for training. Hyper parameter training is inserted to the machine learning model. The value is taken from the database is given to training data and test data. The data is given preprocessing and during preprocessing data is taken for featured extraction. This featured value is given to machine learning model which is trained giving the better results. Before proceeding with model training, data preprocessing is crucial to ensure consistency and mitigate biases. This involves normalizing or scaling features within the dataset. Exploratory data analysis (EDA) is then conducted to analyse the distribution of scores and labels, as well as visualize relationships between features and labels, allowing for insights into dataset characteristics and potential correlations.

Once the final model is built and trained using the entire dataset with optimized hyperparameters, it is evaluated on a separate test dataset to assess its accuracy and generalization ability. Any discrepancies between predicted and actual labels are analysed to identify areas for improvement.

Finally, deployment and user interface development are essential for practical application. A user-friendly interface is developed to allow users to input quiz scores and survey responses easily. The trained model is then utilized to predict dyslexia likelihood based on the provided inputs, and the prediction results are presented in a readable format for users to interpret and act upon accordingly. Following is the proposed methodology for execution:

Data Collection

Creating a thorough exam that covers important cognitive domains such as vocabulary, memory, processing speed, and discrimination abilities is the first stage in the dyslexia prediction process. The test is taken by people from a variety of backgrounds in-order to guarantee a wide sample of possible dyslexia instances. Using pre-established algorithms, test responses are examined in order to produce objective scores for every cognitive domain. Furthermore, respondents fill out a questionnaire covering reading preferences, focus, understanding, and perceived challenges, which adds-up to a total 'Survey_Score' that represents cognitive abilities.

When survey responses and quiz scores are combined, a rich dataset representing different cognitive abilities and dyslexia-related characteristics is produced. This all-encompassing method gives participants insights into their cognitive profiles, which makes it easier to create precise prediction models and do thorough dyslexia analyses.

Giving access and inclusion top priority guarantees representation across demographic categories, which is essential for developing trustworthy prediction models that work for a range of people. The dataset reflects the diversity of dyslexia that exists in the real world by including people of various ages, educational backgrounds, and linguistic backgrounds. Following data collection, thorough analysis reveals patterns and correlations, aiding in

determining the ‘Label’. The ‘Label’ variable ranges from 0 to 2, categorically indicating Dyslexia likelihood. A ‘Label’ of 0 suggests minimal cognitive impairment, while 1 indicates moderate likelihood, warranting further examination. A ‘Label’ of 2 signifies a high likelihood of Dyslexia, necessitating prompt intervention and support.

Table 1. Labelled Dataset Table

S.no	Language_ vocab	Memory	Speed	Visual_Discri mination	Audio_Discri mination	Survey_S core	Label
0	0.5	0.6	0.5	0.8	0.6	0.7	1
1	0.6	0.7	0.8	0.9	0.5	0.8	2
2	0.6	0.4	0.3	0.3	0.4	0.6	1
3	0.3	0.5	0.2	0.1	0.3	0.5	0
4	0.7	0.6	0.7	0.8	0.9	0.5	2

Data Preprocessing

Preprocessing is essential for preparing data for machine learning models, especially when it comes to predicting dyslexia. In-order to do this, data must be arranged into columns that correspond to cognitive traits related to dyslexia, such as vocabulary, language, memory, speed, visual and auditory discrimination, survey score, and label. Researchers can systematically evaluate and interpret different factors contributing to dyslexia likelihood by organizing the dataset in this way. Additionally, preprocessing seeks to normalize characteristics and remove biases in order to guarantee the efficacy and dependability of later model training.

Normalization is a crucial preprocessing step that is frequently accomplished with StandardScaler. This technique centres data around zero and scales it to the unit variance, bringing features to the same scale. Natural variations modify each feature's values to have a mean of 0 and a standard deviation of 1. By ensuring that every feature contributes equally to model training, this standardization approach avoids the dominance of particular features that can distort predictions. Consequently, by resolving problems such as feature variance imbalance, StandardScaler improves model performance and interpretability and raises the accuracy of dyslexia severity predictions.

It is essential to prepare data efficiently for machine learning models by normalizing and removing bias. The dataset is improved for model training by normalizing characteristics and eliminating biases, which makes it possible for algorithms to precisely identify underlying patterns and correlations. This guarantees that machine learning algorithms can accurately forecast the degree of dyslexia, enabling early identification and intervention techniques. Preprocessing also improves machine learning models' interpretability and robustness, which increases their usefulness in practical applications.

To summarize, preprocessing is an essential phase in the dyslexia prediction process that maximizes data dependability and quality for training machine learning models. Biases are removed and features are normalized by sorting data into pertinent columns and using tools like normalization with StandardScaler, which guarantees equitable model training and precise predictions. In the end, this improves the accuracy of dyslexia severity forecasts and facilitates early intervention options for dyslexic individuals by ensuring that machine learning models can successfully identify patterns and correlations in the data.

Exploratory Data Analysis (EDA)

Dataset structures for dyslexia prediction must be understood to properly apply exploratory data analysis (EDA). To find insights directing modelling judgments, it entails analysing score and label distributions. EDA analyses score distributions to find outliers and skewed data using box plots or histograms. It uses heatmaps or scatter plots to investigate correlations between characteristics and the dyslexia severity classification. This makes it easier to spot trends, such as relationships between cognitive domains and the degree of dyslexia. By analysing feature distributions across clinical or demographic groups, EDA also reveals potential biases and ensures fair model building. All things considered, EDA is essential for understanding data patterns, guiding model creation, and resolving biases in dyslexia prediction.

Table 2. Dataset Used in the paper

S.no	Language_ vocab	Memory	Speed	Visual_ discrimination	Audio_ Discrimination	Survey_ Score	Label
1	0.5	0.6	0.5	0.8	0.6	0.7	1
2	0.6	0.7	0.8	0.9	0.5	0.8	2
3	0.6	0.4	0.3	0.3	0.4	0.6	1
4	0.3	0.5	0.2	0.1	0.3	0.5	0
5	0.7	0.6	0.7	0.8	0.9	0.5	2
6	0.4	0.1	0	0.1	0.4	0.2	0
7	0.8	1	0.8	0.9	0.6	0.6	2
8	0.5	0.3	0.5	0.4	0.7	0.4	1
9	0.6	0.5	0.5	0.4	0.6	0.5	1
10	0.6	0.7	0.7	0.8	0.7	0.6	1
11	0.7	0.7	0.5	0.7	0.8	0.8	1
12	0.5	0.6	0.6	0.6	0.7	0.6	1
13	0.9	0.6	0.5	0.9	0.7	0.7	2
14	0.6	0.4	0.4	0.6	0.5	0.6	1
15	0.7	0.6	0.4	0.3	0.4	0.3	1
16	0.6	0.5	0.5	0.5	0.4	0.4	1
17	0	0.2	0.3	0	0.3	0.2	0
18	0.4	0.3	0.3	0.1	0.4	0.4	0
19	0.7	0.7	0.8	0.7	0.7	0.9	2
20	0.5	0.6	0.6	0.4	0.5	0.3	1

Model Selection and Training

In the Model Selection and Training phase, the dataset is partitioned into training and testing subsets using techniques like cross-validation to ensure unbiased model evaluation. Multiple machines learning algorithms, including Decision Tree, Random Forest, and Support Vector Machine (SVM), are applied to the training data. Each model is then trained on the training subset and evaluated using appropriate performance metrics such as accuracy, precision, and recall.The evaluation metrics provide insights into how well each model performs in classifying dyslexia severity levels.Based on the performance metrics, the best-performing model is identified. The models are:

Decision Tree:

A decision tree creates a tree structure by utilizing optimal features to divide data into subsets. Purity is the goal of splitting criteria such as entropy or Gini impurity, which identify splits. Although interpretable and capable of capturing non-linear relationships and feature relevance, decision trees are vulnerable to instability and overfitting. These problems are lessened by pruning and ensemble techniques like Random Forests. A Decision Tree model is trained on a dataset for dyslexia prediction, and its performance is assessed using measures like mean absolute error, accuracy, recall, F1-score, and confusion matrix. Insights on dyslexia likelihood classification and model performance are provided by this method, which presents a thorough strategy for machine learning-based dyslexia detection.

Random Forest Model:

Random Forest trains individual trees on random subsets of data via bootstrap sampling, considering only a portion of the attributes at each node. It improves generality and lessens overfitting by averaging or majority voting

predictions. Because it is ensemble-based and necessitates the adjustment of hyperparameters such as tree count and depth, it is less interpretable even with enhanced accuracy and feature importance evaluation. It performs exceptionally well in a variety of classification and regression tasks in spite of these difficulties. It is trained on a dataset and evaluated using mean absolute error to determine performance. Effectiveness is increased through hyperparameter tuning, usually accomplished with GridSearchCV, which finds the best values for parameters such as n_estimators. After that, the model is retrained and assessed to ensure optimal performance.

Support Vector Machine:

SVM aims to create a decision boundary that is resilient against overfitting by effectively generalizing to fresh data. However, the choice of suitable kernel functions and regularization parameters is critical to the effectiveness of SVM. The model is tested using preset instances to determine its predictive accuracy and its performance is measured using the mean absolute error.Next, GridSearchCV is used for hyperparameter tuning, which optimizes the parameters of the SVM model. The code looks for "C" and "kernel" values that are best for the SVM model. The model is retrained when the ideal parameters have been established, and its performance is assessed in a manner akin to that of the original SVM model.

Model Evaluation

Model evaluation involves assessing the performance of the dyslexia prediction model developed using machine learning techniques. The evaluation process encompasses several crucial steps, including hyperparameter tuning, final model building, and assessing the model's predictive accuracy and generalization ability.

Hyperparameter Tuning

In the dyslexia prediction model, hyperparameters like the number of estimators in the Random Forest algorithm are fine-tuned. By adjusting these hyperparameters, the model's predictive capabilities are enhanced, leading to improved accuracy and generalization.

3.5.2 Final Model Building

After determining the optimal hyperparameters through techniques like GridSearchCV, the final model is built using the entire dataset. This process involves training the model on the complete dataset, incorporating the optimized hyperparameters identified during hyperparameter tuning. In the dyslexia prediction model, for example, a RandomForestClassifier with GridSearchCV is utilized. This final model is well-tailored to the characteristics of the dataset, ensuring optimal performance and predictive accuracy. With the optimized hyperparameters in place, the model is primed for making accurate predictions regarding dyslexia likelihood based on individuals' quiz scores and survey responses. By leveraging the full dataset and incorporating the best-performing hyperparameters, the final model is robust, reliable, and ready for deployment in practical applications.

4. RESULTS AND DISCUSSIONS

Model Performance Evaluation

In this process, rigorous testing is developed for dyslexia detection which separates the test data set in different model evaluations. Various evaluation metrics, including mean absolute error and F1-score, are evaluated to provide quantitative measures of the model. Meanwhile, the F1-score considers precision and recall, offering to classify dyslexia severity levels accurately.

Table 3. Showing error in results in different models

Sno	Model	Error
1	DecisionTree	0.162
2	RandomForest	0.072
3	Support Vector Machine(SVM)	0.075
4	Random Forest (Grid Search)	0.072
5	SVM (Grid Search)	0.075

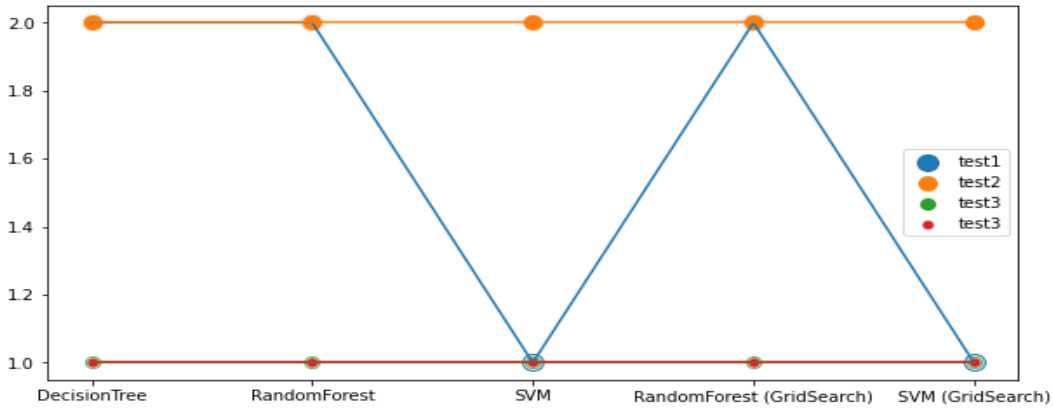

Fig.2 Scatterplot for comparison of the results of different models

From the analysis in Fig. 2 it is observed that the error is maximum in SVM and SVM with GridSearch model, and the error is minimum in the DecisionTree Model. All the formulas are given below:

Formulas for Precision, Recall, F1- Score

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1\ Score = = 2 * \frac{Precision-Recall}{Precision+Recall}$$

To visualize the model's prediction to the actual labels, a confusion matrix is generated. The job of this matrix is to enable a detailed examination of the model's performance across different dyslexia likelihood categories, highlighting any discrepancies or misclassifications[31]. By analysing the confusion matrix, areas for improvement in the model can be identified, guiding further refinements or adjustments.

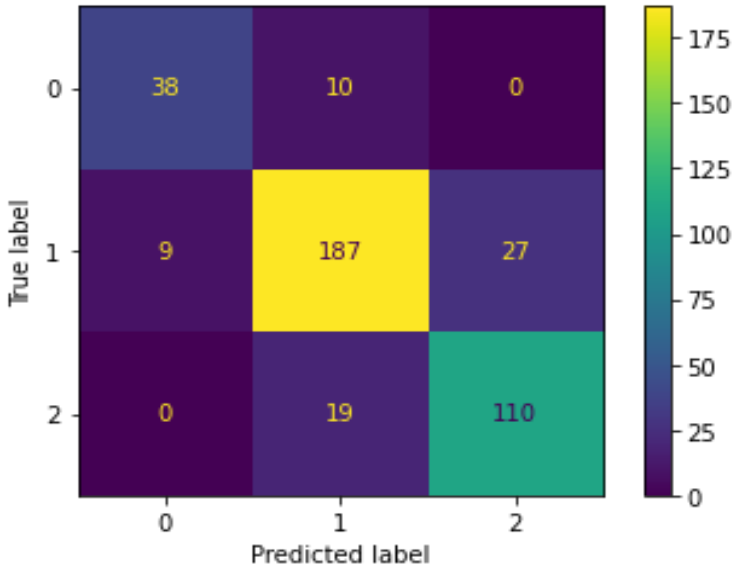

Fig. 3. Confusion Matrix for Decision Tree Model

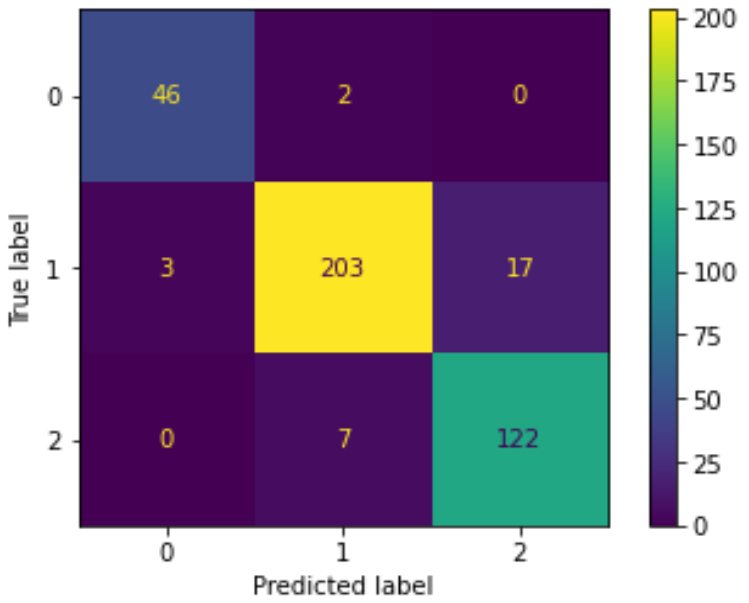

Fig. 4. Confusion Matrix for Random Forest Model

For a DecisionTree Classifier: Precision = 0.826, Recall = 0.828, F1-score = 0.826

For a RandomForestClassifier: Precision = 0.925, Recall = 0.938, F1-score = 0.931

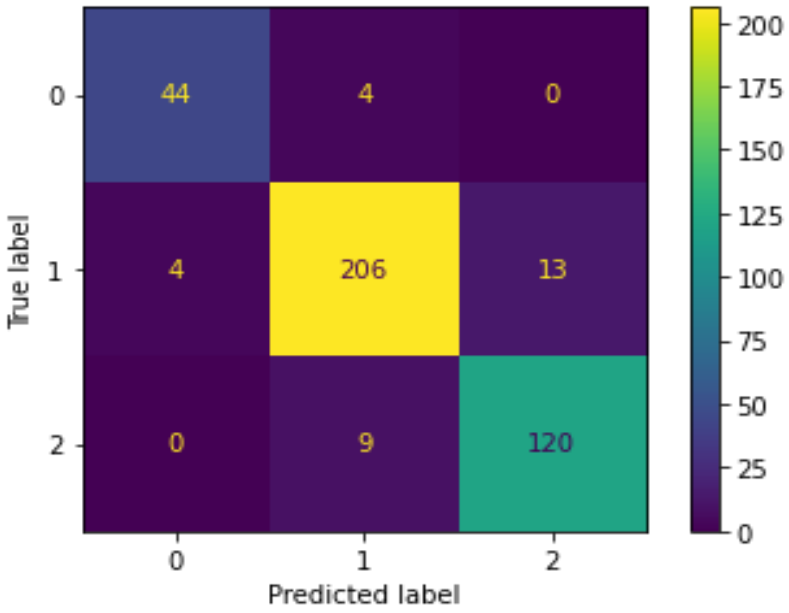

Fig. 5. Confusion Matrix for Support Vector Machine (SVM) Model

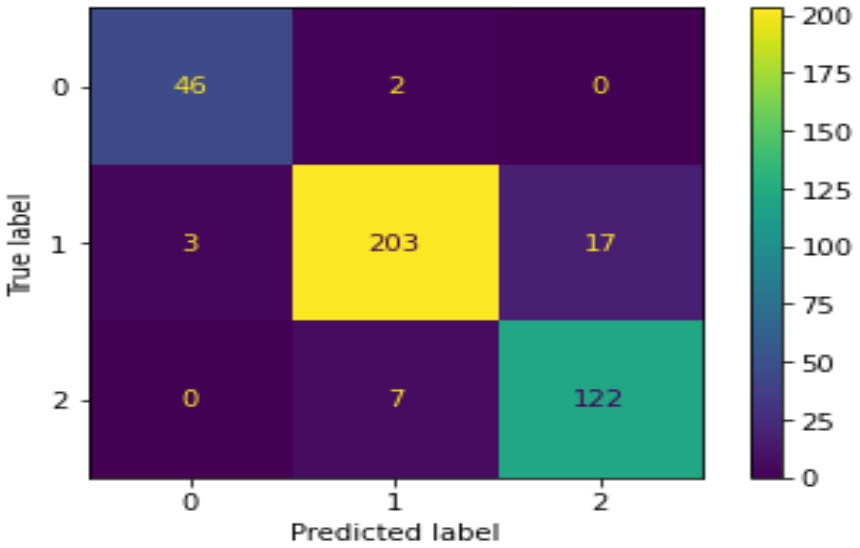

Fig. 6. Confusion Matrix for Random Forest with Grid Search:

For a SVM model: Precision = 0.920, Recall = 0.924, F1-score = 0.922

For a RandomForest model with GridSearch: Precision = 0.925, Recall = 0.938, F1-score = 0.931

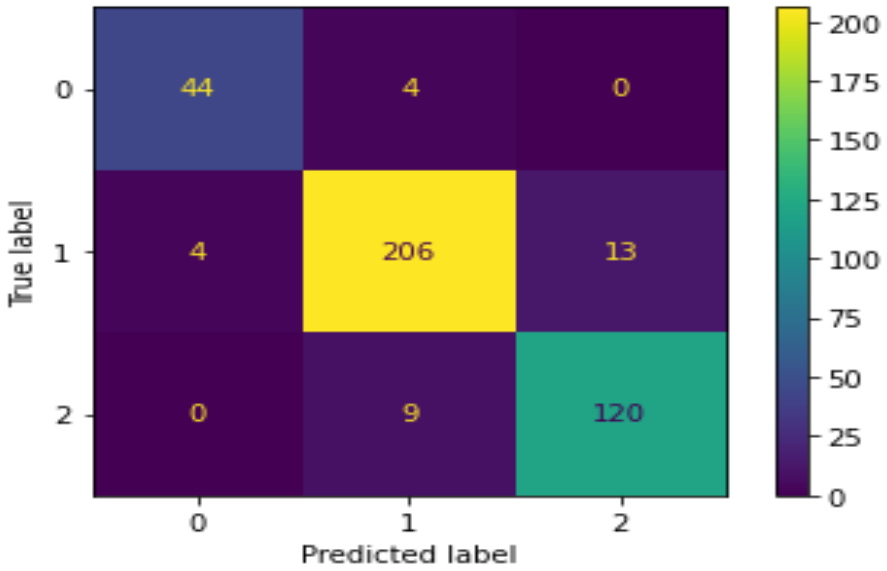

Fig. 7. Confusion Matrix for SVM with Grid Search

For a SVM model with GridSearch: Precision = 0.920, Recall = 0.924, F1-score = 0.922

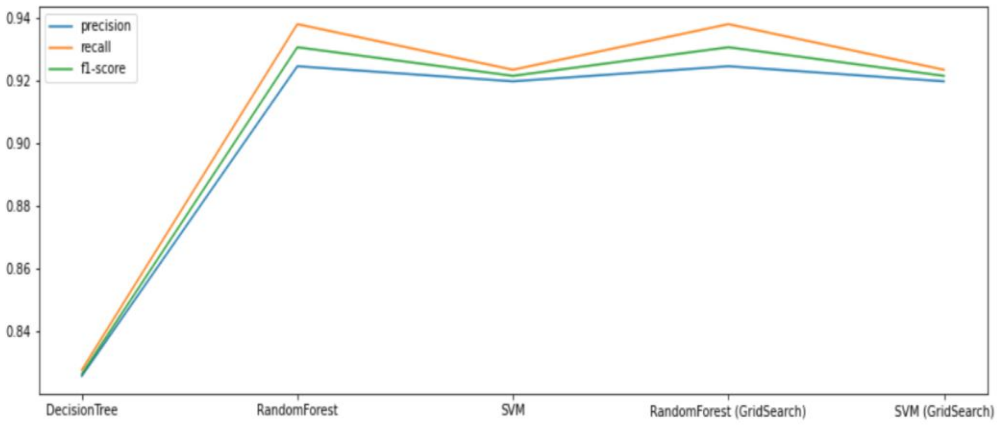

Fig. 8. Line plot to comparison of precision, recall and f1-score of all the models

From the above plot, it is observed that the RandomForest model with GridSearch is the best fit for the given dataset.

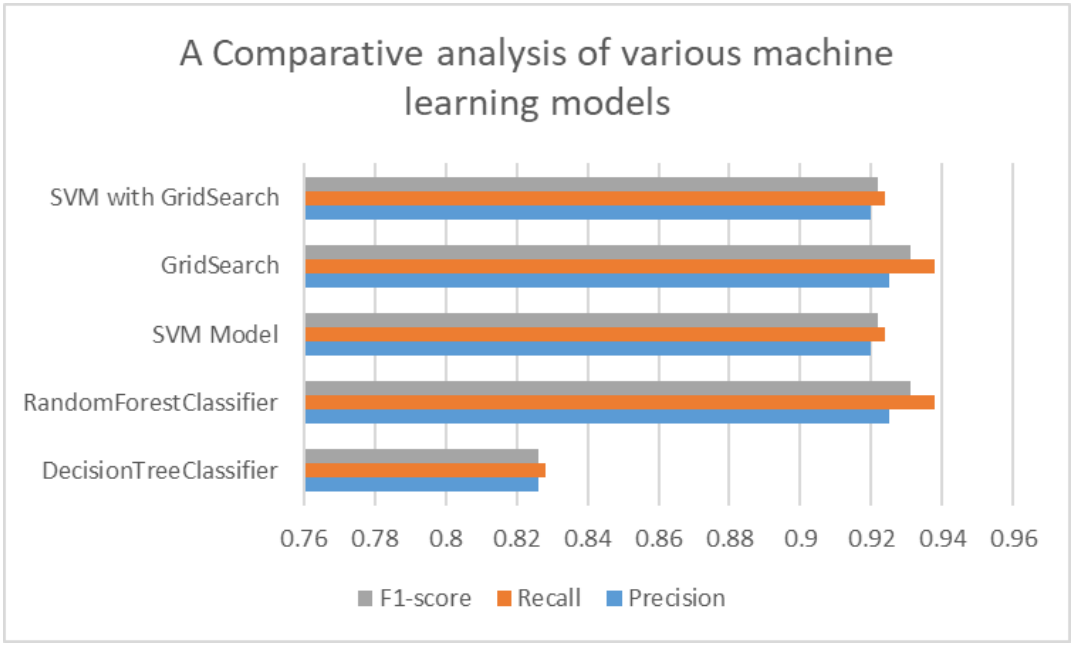

Fig. 9. Comparative analysis of F1-score, Recall and Precision with different machine learning models.

From the analysis, it is observed that RandomForestClassifier performs much better than the other proposed algorithm in terms of F1-score, Recall and Precisions.

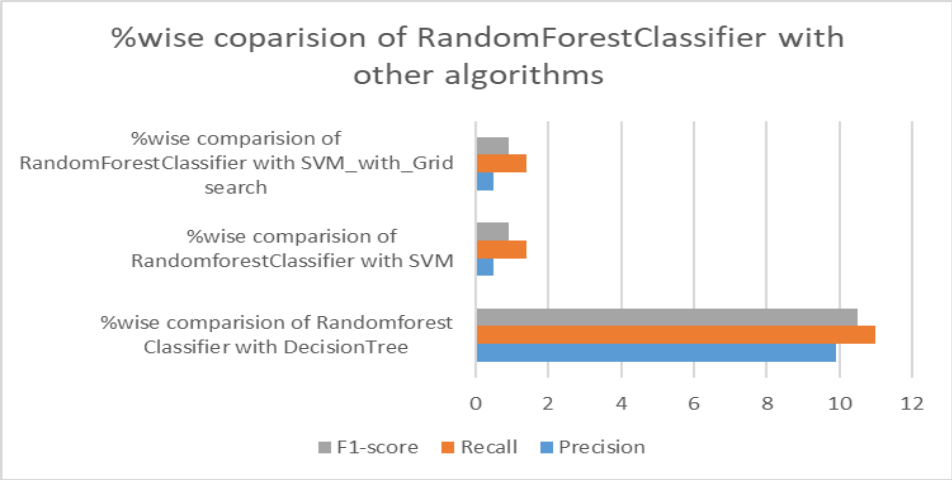

Fig. 10. % Percentage-wise comparison of RandomForest with other algorithms

A comparative analysis of the RandomForestClassifier algorithm is done in the above graph and it is observed from the analysis that RandomForestClassifier is 9.9% better than decision tree, 0.5% better than SVM and 0.5% better than SVMwithGrid in terms of precision. While it shows 11%, 1.4% and 1.4% better than decision tree, SVM and SVMwithGrid algorithms in terms of recall. It is observed from the analysis that it is observed to be 10.5%, 0.9% and 0.9% better than decision tree, SVM and SVMwithGrid in terms of F1score.

5. DISCUSSIONS

Of all the models tested, the Random Forest model with GridSearch had the best precision, recall, and F1-score, demonstrating its improved capacity to correctly categorize people into various dyslexia likelihood groups. In addition, the model demonstrated strong performance, exhibiting a comparatively low error rate of 5.80%. This shows that there are few misclassifications, and consistent predictions are made using the Random Forest model with GridSearch.

The methodology used in the study included several crucial steps, such as gathering data via a thorough survey and quiz, preprocessing the data to ensure consistency and reduce biases, choosing and training models using a variety of machine learning algorithms, optimizing model performance through hyperparameter tuning, and conducting a thorough evaluation of the models using relevant performance metrics. The created model offers a comprehensive approach to dyslexia prediction by utilizing a combination of survey responses and quiz scores. This allows for the consideration of both subjective self-reported data and objective cognitive evaluations. This makes it possible to assess people's cognitive profiles and risk levels for dyslexia in a sophisticated manner, which makes early detection and intervention easier.

6. CONCLUSION

The results of this study have important ramifications for efforts to diagnose and treat dyslexia. The developed model gives parents, educators, and clinicians valuable insights into an individual's dyslexia risk, enabling timely support and intervention to mitigate the long-term consequences associated with undiagnosed dyslexia. This tool for dyslexia prediction is dependable and easily accessible. Ultimately, the Random Forest model combined with GridSearch proves to be the most successful method for predicting dyslexia, providing precise and trustworthy predictions based on people's cognitive evaluations and self-reported information. In order to improve results for those with dyslexia, this study emphasizes the significance of utilizing machine learning approaches to address the important need for early dyslexia detection and intervention.

References

1. X. L. Y. Guo, "An Ensemble Deep Learning Model for Dyslexia Detection in Children," in IEEE International Conference on Bioinformatics and Biomedicine, 2019.
2. H. L. C. Wang, "Dyslexia Prediction Using Convolution Neural Networks with Image-Based Features," in IEEE International Conference on Multimedia and Expo, 2020.
3. L. ,. C. Y. Zhang, "Feature Extraction and Selection for Dyslexia Detection Based on EEG Signals," in International Conference on Biomedical Engineering and Informatics, 2018.

4. H. ,. Z. W. Chen, "Exploring EEG-based Dyslexia Detection Using Deep Learning Techniques," in IEEE International Conference on Neural Networks, 2018.
5. J. W. Q. Liu, "A Deep Learning Approach for Dyselexia Detection using Functional MRI Data," in IEEE International Conference on Medical Imaging, 2019.
6. J. W. X. Li, "Deep Learning for Dyslexia Detection using Brain Imaging Data," in IEEE International COnferenc on Engineering in Medicine and BIology Society, 2021.
7. X. a. W. Z. Zheng, "Dyslexia Detection in Multimodal Data using Deep Learning," in IEEE International Conference on Pattern Recognition, 2021.
8. L. X. S. Yang, "Exploring Machine Learning Techniques for Dyslexia Identification from Eye Movement Data," in IEEE International Conference on Systems, Man and Cybernetics, 2018.
9. Z. Z. S. Wang, "Dyslexia Detection Using Machine Learning from Speech Processing Data," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2019.
10. Y. L. H. Zhang, "Predicting Dyslexia Risk Using Machine Learning and Behavioural Data," in IEEE International Conference on Cognitive Computing, 2020.
11. A. J. P. Ramteke, "A Comprehensive Study on Dyslexia Detection using Deep Learning Techniques," in International Conference on Computational Intelligence and Data Science, 2020.
12. S. S. R. a. S. A. Gupta, "Machine Learning Approaches for Dyslexia Detection: A Review," Lecture Notes in Networks and Systems, Springer, 2019.
13. M. T. Johnson, "A Comparative Study on Machine Learning Algorithms for Syslexia Detection," in International Conference on Intelligent Computing and Communication, 2021.
14. F. A. S. Khan, "Feature Selection Techniques for Dyslexia Detection Using Machine Learning," in International Conference on Artificial Intelligence and Big Data, 2022.
15. R. C. M. K. O. a. M. M. O. L. Usman, "Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and Challenges," IEEE Access, vol. 9, pp. 36879-36897, 2021.
16. M. K. K. A. A. Y. A. G. Fateme Asghari Tolami, "An intelligent linguistic error detection approach to automated diagnosis of Dyslexia disorder in Persian speaking children," in 11th International Conference on Computer and Knowledge Engineering (ICCKE 2021), October 28-29, 2021, Ferdowsi , Mashhad, Iran, 2021.
17. U. G. V. U. M. K. S. Tushar B T, "Automated Detection of Dysgraphia Symptoms In," in 2024 International Conference on Emerging Smart Computing and Informatics (ESCI) , AISSMS Institute of Information Technology, Pune, India. Mar 5-7, 2024, 2024.
18. X. Z. R. J. S. Kaiming He, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition, Microsoft Research, 2016.
19. V. K. P. M. S. M. J. Ivan Vajs, "Dyslexia detection in children using eye tracking," in IEEE, EUSIPCO 2022, This research was supported by the Ministry of Education, Science and Technology Development of Serbia, Belgrade, Serbia, 2022.
20. G. K. K. Y. N. T. Atmakuri Sasidhar, "Dyslexia Discernment in children based on Handwriting images using Residual Neural Network," in 2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Dayananda Sagar University, Bangaluru, 022.
21. S. S. A. S. ROSE, "Enhanced Special Needs Assessment: A Multimodal Approach for Autism Prediction," Enhanced Special Needs Assessment: A Multimodal Approach, vol. 12, no. 2024, pp. 121688-121699, 2024.
22. Z. L. F. X. Xin Li, "Eye-tracking based Detection of Developmental Dyslexia in Children Using Convolutional-Transformer Network," in Proceedings of the 2024 27th International Conference on Computer Supported Cooperative Work in Design, Dayananda Sagar University, Bangaluru Karnataka, India.
23. S. D. a. S. Muntaha, "Hybrid Deep Learning for Dyslexia Identification through Heterogeneous Cognitive and Behavioral Data Analysis," in 5th International Conference on Computing Communication and Networking Technologies (ICCCNT), Dayananda Sagar University, Bengaluru, Karnataka India, 2024.
24. S. S. a. L. A. D. P, "Detection of Potential Specific Learning Disabilities in Children through Handwriting Analysis Using Machine Learning," in 2024 IEEE Region 10 Symposium (TENSYMP), New Delhi, India, 2024, Delhi, 2024.
25. L. B. A. M. K. M. A. A. A. M. a. M. M. K. Gasmi, "Optimal Ensemble Learning Model for Dyslexia Prediction Based on an Adaptive Genetic Algorithm," IEEE Access, vol. 12, pp. 64754-64764, 2024.
26. H. C. B. B. A. D. a. G. P. Y. K. Meena, "Detection of Dyslexic Children Using Machine Learning and Multimodal Hindi Language Eye-Gaze-Assisted Learning System," IEEE Transactions on Human-Machine Systems, vol. 53, no. 1, pp. pp. 122-131, 2023.
27. G. S. K. T. M. P. a. M. M. J. A. Vajs, "Eye-Tracking Image Encoding: Autoencoders for the Crossing of Language Boundaries in Developmental Dyslexia Detection," IEEE Access, vol. 11, pp. 3024-3033, 2023.
28. C. Rajan and B. A. G. B. G. K, "Artificial Intelligence Enabled Hybrid Machine Learning Application for Dyslexia Detection using Optimized Multiclass Support Vector Machine and Personalized Interactive and Assistive tools using Adaptive Reinforcement," in 4th International Conference on Innovative Practices in Technology and Management (ICIPTM) IEEE, 2024.
29. A. K. S. A. G. S. a. B. J. C. A. Senthilselvi, "Enhancing Dyslexia Awareness: A ML Model for Early Identification and Support," in 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT) IEEE, Chennai, 2024.

30. R. C. M. K. O. a. M. M. O. L. Usman, "Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and Challenges," IEEE Access, vol. 9, pp. 36879-36897, 2021.
31. Khan Mubeen Ahmed, “Performance research of WiMAX networks with frame periods in strict priority”, International Journal of Recent Technology and Engineering, vol 8 Issue 2(11), 2019
32. Khan Mubeen Ahmed, “Gender Classification Based on Machine Learning Models”, 2025 4th Opju International Technology Conference on Smart Computing for Innovation and Advancement in Industry 5 0 Otcon 2025, IEEE.

BIBLIOGRAPHY

1. Dr. Mubeen Ahmed Khan is an Associate Professor, Department of Computer Science and Engineering NIMS University Rajasthan, Jaipur with an overall experience of more than 20 years.He has done his B.E. (IT:2005) and M-Tech (CSE:2012) from Rajiv Gandhi Proudyogiki Vishwavidyalaya and PhD Sangam University (CSE:2024). He has research experience in the fields of Wireless Networking, Ad-Hoc Networking, Cyber Security and Machine Learning. He is having a vast experience of Academics with more than 20 years. Mail id: makkhan0786@gmail.com, mubeen.khan@nimsuniversity.org

2. Dr. Latika Jindal is an Associate Professor in the Department of Computer Science and Engineering Medicaps University Indore, M.P. She has more than 15 years of experience. Her Research interest includes Block Chain Technology and Data mining. Mail id: latika19mehrotra@gmail.com

3. Dr. Piyush Chouhan is an Assistant Professor in Department of Electronics and Communication Engineering Medicaps University Indore, M.P. He is having more than 13 years of experience. Mail id: pchouhan6@yahoo.com

4. Prof. Ankita Chourasia is Assistant Professor in Medicaps university from 2022. She has completed her BE in CSE from DAVV Indore and ME from DAVV Indore. She has and overall experience of 13 years. Her area of specialization is networking, Ad-hoc networks, and machine learning. Mail id: ankita.chourasia29@gmail.com

5. Dr. Pankaj Malik is an Assistant Professor in Medicaps University, Indore, since 2022. He has an overall experience of 22 years. He has completed his Bsc. In computer science, MSc in CS, MCA and PhD in Artificial Intelligence. Email id: pankajmalik1979@gmail.com

6. Dr. Owais Ahmad Shah is an Assistant Professor in the Department of Electronics and Communications Engineering at Dayananda Sagar University, Bengaluru, Karnataka, India. He has more than 14 years of experience in education. Mail id: mail_owais@yahoo.co.in