IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 15 January 2024, accepted 2 February 2024, date of publication 6 February 2024, date of current version 16 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3363634

RESEARCH ARTICLE

Explainable Machine Learning Prediction for the Academic Performance of Deaf Scholars

N. R. RAJI[1], **R. MATHUSOOTHANA S. KUMAR**[2], **AND C. L. BIJI**[3]

[1]Department of Computer Applications, Noorul Islam Centre for Higher Education, Nagercoil 629301, India
[2]Department of Information Technology, Noorul Islam Centre for Higher Education, Nagercoil 629301, India
[3]Department of Analytics, School of Computer Science and Technology, VIT University, Vellore, Tamil Nadu 632014, India

Corresponding author: N. R. Raji (rajinrsreelakam@gmail.com)

This work was supported in part by the Centre of Excellence for Disability Studies (CEDS), Kerala, India, under Grant R21F3.

ABSTRACT Deaf and Hard of Hearing (DHH) students encounter obstacles in higher education due to language and communication challenges. Although research aims to improve their academic performance, the potential of Machine Learning (ML) remains underutilized in DHH education. The opacity of ML models further complicates their adoption. This study aims to fill this gap by developing a novel ML-based system with eXplainable AI (XAI), specifically utilizing Local Interpretable Model-Agnostic Explainer (LIME) and Shapley Additive Explainer (SHAP). The objective is twofold: predicting at-risk DHH students and explaining risk factors. Merging ML and XAI, this approach could positively impact DHH students' educational outcomes. A dataset of 454 records detailing DHH students is collected. To address dataset limitations, synthetic data and SMOTE are used. Students are categorized into three performance levels. The data is modeled with different ML models, transfer models, ensemble models, and combination models. Among the models, the stacked model with XGBoost, ExtraTrees, and Random Forest exhibited better performance with an accuracy of 92.99%. Results highlight the model's significance, providing insights through XAI into crucial factors affecting academic performance, including communication mode, early intervention, schooling type, and family deafness history. LIME and SHAP values were found to be effective in deriving insights into DHH student performance prediction framework. Communication mode, notably, strongly influences at-risk students. The major contribution of this study is the development of a novel ML-based system and the XAI interpretations whose value lies in its social relevance, guiding stakeholders to enhance DHH scholars' academic achievements.

INDEX TERMS Artificial intelligence, deaf education, explainability, machine learning.

I. INTRODUCTION

The academic progress of DHH scholars is essential in achieving sustainable economic development in the community. Constructive feedback at an early stage can help scholars to improve their productivity. Though deliberate efforts were initiated to address the challenge, DHH scholars needed to be provided with opportunities in the right direction. Quality feedback and action are essential for this self-regulated learning. There are 388 government-funded schools for DHH students in India [1] but the colleges that offer undergraduate programs for this population are only six [2] as per the available information. Policymakers in India are eager to create early detection and intervention programs for DHH youngsters. The stakeholders of education [3] - students, teachers, and policymakers- need valuable insights that can be used to improve the educational conditions of DHH students.

The factors affecting the academic performance of a DHH student are different when compared to his hearing peer. [4] states that hearing loss of any type or degree can be a barrier to learning. In India, hard-of-hearing students usually prefer an inclusive setup in regular schools, and deaf students generally prefer special schools. As per the Rights of Persons With Disability (RPWD) Act, 2016, a person with 60 DB to 70 DB hearing loss in both ears is considered Hard of Hearing. Though deafness is not

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

a barrier to learning, the inaccessible environment makes learning a challenge for a DHH student. The difficulties faced by DHH students are studied in [5] which identifies the need for communication access, classroom modifications, and other accommodations that are needed to improve the success rate of DHH students. The studies show that students even with minimal or mild hearing loss suffer from poor speech recognition which eventually results in poor academic performance [6]. A regression model with variables related to education, hearing devices, sign language proficiency, and home language was suggested in [7] which shows that sign language proficiency has a significant effect on the academic performance of DHH students. The sign language proficiency of teachers is also important in delivering the content effectively in a classroom with DHH students [8].To maximize the academic potential of these students early intervention and usage of assistive devices like hearing aids are also important [9].

The challenges faced by DHH students are very different from those of hearing students. This calls for the need for studies that specifically focus on the performance of DHH students. The low academic performance of DHH students has been attributed to several issues. Low academic performance has resulted in poor employability, low salaries, and substandard quality of life in India, hence research into this performance and the causes associated is critically needed [2]. In various studies the factors affecting the academic performance of DHH students are identified as parent influence, availability of facilities, teaching, reading, and learning materials [10], [11], Mode of communication, the usage of sign language or speech [7], Grades, demographics, geographical region, school, course type, and course score [12]. In addition to these factors, deafness-related factors also need to be identified which can be an influential factor in their academic performance. From the above discussion, the deafness-related factors identified for this study are communication mode, family history, type of schooling, type and degree of hearing loss, usage of hearing aids, and intervention through speech therapy and cochlear implants.

This paper explores the role of explainable AI (XAI) in improving the performance of DHH scholars in education. The significance of this study includes in its social relevance. As reported by World Federation for the Deaf 80% of the 32 million deaf children worldwide do not have any access to schooling. Only 1-2% of deaf youngsters receive instruction in sign language even when there are educational options. The possibilities to enhance the education condition of DHH population using ML techniques is under-explored and this study is a stepping stone towards achieving this goal. The major contributions of this study include:

- Novel ML System with XAI: A novel ML system integrates eXplainable AI (XAI) using LIME and SHAP techniques. It predicts at-risk DHH students and explains academic risk factors
- Enhancing Education: This study aims to improve DHH students' academic achievements. Identifying factors like communication mode, early intervention, and family history guides effective support..
- Addressing Data Imbalance: Synthetic data and SMOTE mitigate DHH dataset limitations, boosting model robustness
- Algorithmic Comparison: Comparative analysis highlights the stacked model with XGBoost, ExtraTrees and Random Forest superior accuracy (92.99%) in predicting at-risk DHH students, aiding algorithm selection.
- Interpretability and XAI: Incorporating XAI explains predictions, enhances model transparency and offers insights into academic success factors.

This paper is organized as follows: Section II highlights the related works in academic performance prediction of students using Machine Learning techniques. Section III explains the method followed in this work. Section IV discusses the results obtained and Section V concludes by summarising the contributions and future scope of this study.

II. RELATED WORKS

Machine learning (ML) is revolutionizing education and altering teaching, learning, and research in profound ways. ML is being used by educators to identify at-risk students earlier and take steps to increase their achievement. Researchers are using machine learning to speed up research and uncover new findings and insights. The ML models are often considered black boxes which do not provide explanations for the predictions made. The emergence of Explainable Artificial Intelligence (XAI) made these black boxes more transparent. In 2016, Roberto et al. suggested an algorithm called Local Interpretable Model-Agnostic Explanations (LIME), a unique explanation technique that learns an interpretable model locally around the prediction to explain any classifier's predictions in an interpretable and faithful manner [13]. In 2017, a unified framework for interpreting predictions, SHAP (SHapley Additive exPlanations) is introduced [14]. These methods justify the prediction made by the ML model by giving an appropriate explanation for it.

In today's world, AI systems and Machine Learning algorithms are widely used in a variety of fields. Data is used in different ways to assist humans in solving issues. The advancement of machine learning, demands transparency in its prediction. Explainable AI, or XAI, is a set of techniques for better understanding and validating machine learning models. Models must be understood not only by data scientists who are building them but also by end-users who want explanations for why certain decisions are made.

In the education domain, the stakeholders of education highly benefit from the use of machine learning models. The classifications and predictions made by these models help the stakeholders to get insight and help them to make more effective decisions. In [15], the major stakeholders who use the models are classified into four. The students are represented as the affected users, the advisors or teachers are represented as the end-users who trust the model, the regulatory bodies who obtain insights from the model, and the AI system builders who train and evaluate the model and ensure its performance. Some of the major applications in the education domain using ML techniques are the classification of students according to their academic performance [16], [17], [18], identification of at-risk students [19], [20], [21] prediction of marks [22], [23], [24] and identification of factors affecting academic performance [25], [26].

The predictions made by machine learning models indeed give valuable insights to the students and other stakeholders of education. Unfortunately, the current approaches usually give only predictions, not the actual explanations of the predictions to the students [27]. An attempt is made to overcome this major limitation by adding explainability to the developed model. The scope of explainability can be local or global. Global-level explanations explain the complete model and local-level explanations provide explanations at an instance level [15]. Local explanations illustrate how the outcomes of the model change when the values of specific features change within a specific interval [28]. This makes the prediction made by the model more trustworthy and useful.

The pilot study of this research [29] revealed that deafness-related factors have a strong contribution to the academic performance of students with hearing loss. According to a thorough study done in Saudi Arabia using data of DHH students, there is a strong correlation between student grades, demographics, geographic region, school, course type, and course score when predicting academic outcomes. Religious curricula, Arabic language, and Mathematics also function as important predictive indicators of success or failure in the outcome [12]. A time series model was proposed in this study to predict the academic performance of DHH students. Another significant study in analyzing the predicting the academic performance of DHH students was in 2009 [30]. Statistical methods are used in this study to predict the academic performance of DHH students. The datasets used in both these studies are not openly available.

The power of XAI is effectively used in many applications like credit risk prediction [31], Network Intrusion detection [32], medical diagnosis [33], and student performance [34]. The papers on using XAI methods in the academic performance analysis of DHH students are very limited. A framework, XAI-ED is proposed in 2022 using which educational AI tools can be designed and developed [35]. It is observed that the development of dashboards that use the XAI techniques can give valuable feedback to all the stakeholders of education [27].

III. METHOD

Figure 1 illustrates the performance-predicting framework. Due to the absence of a standard dataset, Phase 1 aimed at the creation of a dataset followed by the relevant features linked with the academic performance of DHH students. Based on the performance of first semester marks, the students are categorized as A (marks between 481 and 600), B (marks between 351 and 480, and C (marks below 351)

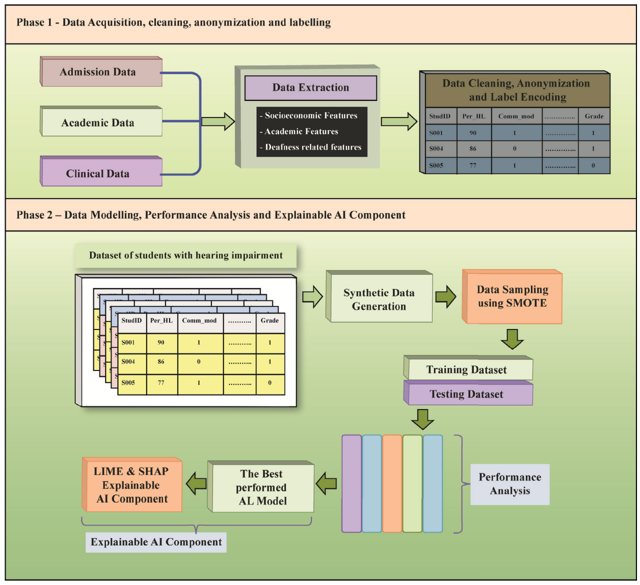

FIGURE 1. Explainable framework for the academic performance prediction of deaf scholars.

grades. To overcome the small size of the dataset, the synthetic data generation method is used, and to overcome the bias due to imbalanced data, SMOTE (Synthetic Minority Oversampling Technique) is used. Finally, in Phase 2, a model was developed using different classifiers, and the important features were displayed using XAI methods. For implementing the proposed model Python programming with LIME and SHAP libraries for XAI are used.

A. DATA ACQUISITION

The data for this study is collected from the higher education institutes for the deaf in different parts of India. The data sources are admission, academic and clinical records of students. The data is also collected through an online survey with their consent. Specially made signed videos were also included in the survey to ensure accessibility. The signed videos were made with deaf signers and the validation is done with another two Sign Language experts who are deaf. A total of 454 students gave their consent and their data is collected. The limited higher education opportunities for DHH students explain the small size of the dataset.

B. FEATURE ENGINEERING

21 features are identified which are further classified into personal details, prior academic details, and deafness-related features. The attribute details in the dataset are given in Table 1 and represented in Figure 2.

The dataset has 35% of female students and 65% of male students. 49.44% of students use sign language and speech for communication while 49.89% use only sign language. Only 0.67% use speech as a mode of communication. About 82% of participants are from the southern states of India,-Kerala, Andhra Pradesh, Karnataka, and Tamilnadu, while the rest

TABLE 1. Feature details of the dataset.

Sl No	Features Selected	Values	Type
Socio-economic features			
1	Gender	(Male, Female)	Categorical
2	Family Income	(Below 5000, 50000-1Llakh,1 -5 Lakhs-5 Lakhs – 10 Lakhs, Above 10 Lakh)	Categorical
3	Education of father	(No Formal Education, Below High School, Plus Two, Under Graduation, PostGraduation, Ph.D.)	Categorical
4	Education of mother	(No Formal Education, Below High School, Plus Two, Under Graduation, PostGraduation, Ph.D.)	Categorical
5	District	(All districts of Kerala, District outside Kerala, Outside India)	Categorical
6	State	(All states of India, Outside India, Union Territory)	Categorical
Deafness related features			
7	Percentage of Hearing Loss	Numeric : 0-100	Numeric
8	Type of Hearing Loss	(Mild, Moderate, Moderately Severe, Severe, Profound)	Categorical
9	Degree of Hearing Loss	(Conductive, Sensorineural, Mixed)	Categorical
10	Speech Therapy received below age 5	(Yes, No)	Categorical
11	Speech Therapy received above age 5	(Yes, No)	Categorical
12	Hearing Aid User	(Yes, No)	Categorical
13	Undergone Cochlear Implant	(Yes, No)	Categorical
14	Communication Mode	(Sign language, Speech, Both Sign and Speech)	Categorical
15	Family History	(Yes, No)	Categorical
16	If yes, relation	(No family member is deaf, Father /Mother/Sister/Brother, Relative, Both)	Categorical
Prior academic features			
17	Type of schooling in class 10	(Special school, Regular School)	Categorical
18	Percentage of marks scored in class 10	Numeric : 0-100	Numeric
19	Type of schooling in class 12	(Special school, Regular School)	Categorical
20	Percentage of marks scored in class 12	Numeric : 0-100	Numeric
21	Semester 1 marks	Numeric : 0-600	Numeric

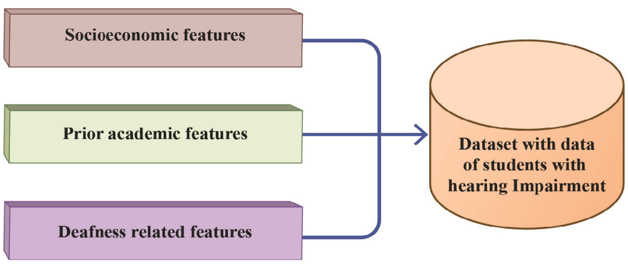

FIGURE 2. Feature engineering for explainable student performance model.

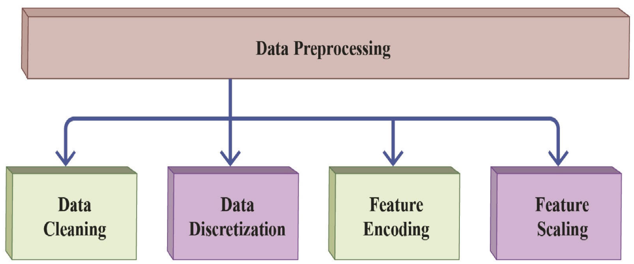

FIGURE 3. Data preprocessing stages.

18% are from various states of India-West Bengal, Uttar Pradesh, Rajasthan, Odhisha, Maharashtra, Madhya Pradesh, Jharkhand, Haryana, Bihar, Gujarat, and Goa. It is found that 57% of participants got early intervention through speech therapy and about 30.38% of participants have a family history of hearing loss. 75.39% of students undergo schooling in special schools. 54.55% are from economically backward conditions with a family annual income below Rupees fifty thousand.

C. DATA PREPROCESSING

The collected data needs to be refined in the following ways (i) data cleaning (ii) data discretization (iii) feature encoding and (vi) Feature scaling were used in the preprocessing stage [29] as shown in Figure 3.

In the raw data, the possibility of incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data was analyzed during the data cleaning phase. There are numerous opportunities for data to be duplicated or mislabeled when combining multiple data sources. The percentage marks are corrected to 2 decimal places. The missing values in these fields were replaced by the average values. The records with no values in the majority of columns were deleted.

A discretization mechanism is used in this study to convert the numerical values of semester 1 marks into nominal values to represent our dataset as a classification problem. The students are classified according to their semester marks as A, B, and C, that is high-level performers, middle-level performers, and low-level performers. The Grade feature is used as the target variable based on which the students are classified.

In this stage, the categorical feature values are converted to numerical form before feeding into the model. The target variable is label encoded with values as follows. The low-level performers are given a grade of C to a value of 0, middle-level performers of Grade B to a value of 1, and High-level performers of Grade A to a value of 2.

In this study, the standard scale technique is used for feature scaling. The mean and standard deviation of each feature are calculated to scale that feature. The scaled value for each feature is calculated as follows.

$$Xi = \frac{x - \mu}{\sigma} \tag{1}$$

where Xi is the scaled feature, μ is the mean and is the standard deviation.

D. MODELLING WITH THE COLLECTED DATA

In standard statistical analysis, the margin of error is first chosen to determine the size of samples from the entire population. If 'n0' is the sample size without a finite population correction factor, the Margin of Error (ME) is given by

$$ME = z\sqrt{(p(1-p)/n_0)} \tag{2}$$

where z is the z-score and p is the initial estimate of proportion. For evaluating the performance of genome compression tools, 'p' may be considered as the estimate of proportion showing the capability of compression tools to reduce the given genome data. From the above equation, sample size 'n0' without a finite population correction factor may be calculated as

$$n_0 = \frac{(z^2 * p * (1-p))}{ME^2} \tag{3}$$

We assume a margin of error of 2.5% and a z-score of 1.96 for 95% confidence. As there is no prior study about the initial estimate of proportion p, we assume by general convention of p=0.5. 'p' in our study is assumed as the pass percentage of deaf students substituting on equation (2) gives n as 1537. The value calculated above is for an unknown population. In our case, we have a finite population (NCBI database). For a finite population size of 'N', sample size 'n' using a finite population correction factor may be computed as

$$Corrected\ Sample\ size,\ n = \frac{n_0}{1 + \frac{n_{0-1}}{N}} \tag{4}$$

As the current statistics on higher education of deaf students are not available, the authors identified 13 institutes situated in various parts of India that offer UG courses exclusively for the deaf with a population size of 1500. Based on it, the corrected sample size, 'n' is 759. A total of 454 students gave their consent and their data was collected. To make it a representative sample, conditional generative adversarial networks are used to scale the data set 1500.

The collected data of 454 records is modeled with transfer learning classification models like Tab Net, VGG 16, and ResNet50 classifiers, and the performance of the model is compared. The performance of these models was found to be not satisfactory.

E. SYNTHETIC DATA GENERATION

The dataset size is one of the crucial factors that affect the performance of a model. In this case, the size of the dataset is small. There are synthetic data generation methods to increase the size of data. The similarity score gives an idea of the similarity between real data and fake data. In this paper, conditional generative adversarial networks from open-source Python libraries called CTGAN and Synthetic Data Vault (SDV) are used for generating fake data. A similarity score of 0.6 is obtained which shows that the dataset is ready for production. After synthetic data generation, the size of the data became 1500.

F. DATA SAMPLING USING SMOTE

The obtained dataset was imbalanced with uneven distributions of target class observations. As shown in Figure 4, the instances of classes are distributed in the training dataset as class A-18.1%, class B - 72.9%, and class C-9.0%.

Here it is observed that class C and class A are the minority classes whereas the majority class is class B. That is, the class distribution is not equal or close to equal. This is a challenging problem that appears in classification problems and leads to poor performance. One of the most important factors in improving model performance is resolving the issue of an imbalanced dataset. This problem leads to the majority class dominating the minority class during the classification. When the dataset is imbalanced, the classifier may get biased toward the prediction.

To overcome the problem of the imbalanced dataset, the oversampling Synthetic Minority Oversampling Technique (SMOTE) is considered a benchmark solution [30]. In this study, this method is used to handle an imbalanced dataset problem. SMOTE examines minority class instances and uses k nearest neighbor to choose a random nearest neighbor, then creates a synthetic instance in feature space at random.

G. DATA MODELLING

After balancing the dataset, it is modeled using different machine-learning classification algorithms. The algorithms used are classic ML models - Support Vector Machine (SVM), K Nearest Neighbor (KNN), Naïve Bayes(NB), Tree-Based models-Decision Tree(DT), Random Forest (RF), XGBoost (XG), ExtraTrees Classifier(XT) and Combination models -Stacked model of XG, RF and XT. The performance of models is then compared. It was found that the performance of the Stacked model is superior when compared to others. Further, the evaluation of the model is done using

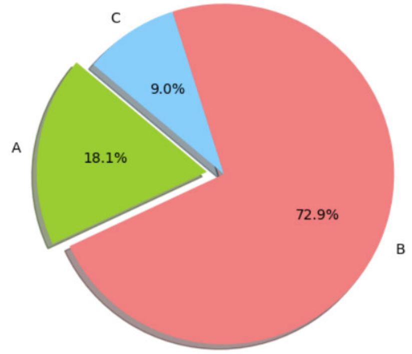

FIGURE 4. Distribution of grades in dataset.

a classification report considering the values of accuracy, F1 score, precision, and recall. The performance of the model is further improved by using the k-fold validation technique. Since the dataset is balanced, this method is suitable for model evaluation. The k-fold validation technique is used to generalize the model. That is, to reduce the variability, cross-validation is performed using different subsets of the same data. The model evaluation is done with accuracy, F1 score, and precision. The sensitivity and specificity scores are also considered for evaluating the performance of the model. The parameters are calculated with True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) from the Confusion matrix.

Cross-validation is used for a robust evaluation of the model's effectiveness. A machine learning model's performance and capacity for generalization are assessed using the cross-validation methodology, which divides the dataset into many subsets systematically This process is performed several times, with a new subset being used for testing each time. The model is trained on a subset and tested on the remaining data. To reduce the possibility of bias or overfitting that comes with a single train-test split, the results are averaged to offer a more reliable evaluation of the model's performance. By averaging results over multiple iterations of training and testing data subsets, this method evaluates the overall performance of the model. Cross-validation maximizes data utilization and identifies potential issues like overfitting and underfitting.

H. EXPLAINABLE AI COMPONENT

To make the model transparent, an attempt was made to include Explainability to improve the student's experience in learning. Two methods are used to add explainability to the model

1) Explainability with SHAP values.
2) Explainability with LIME

1) EXPLAINABILITY WITH SHAP VALUES

SHAP is a more comprehensive and interpretable method for evaluating feature importance in machine learning models, while feature importance is a quick and simple option. SHAP provides a unified and consistent measure of feature importance, accounting for complex relationships between features and the model output. It assigns a contribution value to each feature for a specific prediction, whereas feature importance only provides a single value per feature across the entire dataset. feature importance is based on a linear approximation of the model's output while SHAP can account for non-linear relationships and interactions between features. Being model agnostic, SHAP can be applied to any model regardless of its type, while feature importance is specific to certain types of models, such as decision trees.

TABLE 2. Method to calculate shapley values.

Step 1 :	Generate the power set of F (including the empty set and F itself).
Step 2:	Obtain model predictions for all subsets of features, treating excluded features as missing.
Step 3:	Calculate the marginal contribution of each feature by comparing model predictions with and without that feature.
Step 4:	Calculate the Shapley value of each feature as the average, over all permutations, of its weighted marginal contributions.

Using SHAP, the reliability of the model can be enhanced by explaining the model behavior. SHAP has explainers and model-agnostic methods on how the features used in the model influence the outcome predicted by it. SHAP uses Shapley values to explain to what extent each input features contribute to every model prediction. The average marginal contribution of a feature value across all possible coalitions is referred to as the Shapley value. The Shapley value is calculated for the feature f out of the set of input features F as in the method described in Table 2.

Mathematically Shapley values can be calculated using the formula [31],

$$\phi i(f,x) = \Sigma z' \underline{c} x' \frac{|z'|!(M-|z'|-1)!}{M!}[fx(z')-fx(z'\backslash i)] \quad (5)$$

where f is the model, x is the available features, and x are the selected features. The quantity $fx(z')-fx(z\backslash i)$ expresses, for every single prediction, the deviation of Shapley values from their mean: the contribution of the i-th feature.

SHAP can give both global and local explanations for the model. The summary plot, force plot, and waterfall plot in the SHAP package are used to give explanations for the predictions. The summary plot combines the relevance of features with the effects of features. A Shapley value for a feature and an instance is represented by each point on the summary plot. The feature determines the position on the y-axis, while the Shapley value determines the position on the x-axis. The color denotes the feature's value, which ranges from low to high. Overlapping points are jittered in the y-axis direction to give us a sense of the Shapley value distribution per feature. The features are ranked in order of importance. The force plots in the SHAP package can give both global and local interpretations. The SHAP waterfall plot visualizes how individual localized predictions are made.

2) EXPLAINABILITY WITH LIME

The LIME zooms into the local area of the individual prediction and creates a simple explanation that makes sense in the local region. LIME creates a linear model in the local area which helps to make simple explanations for the prediction. That is a local approximation of a complex model. Using the prior knowledge about the domain, the explanations can be validated and trust can be built. The explanations using LIME are valid locally, in the global perspective, they may not be faithful. The working of the LIME algorithm is shown in Figure 5.

The algorithm of LIME assumes that every complex model is linear on a local scale. LIME seeks to fit a basic model around a single observation that simulates the behavior of the global model at that location. For the instance to be explained, perturb the instance n times to create replicated feature data with slight modifications in values. This perturbed data is created around the instance used to build the linear model locally. Based on the range of possible category values and how frequently they appear in the training dataset, random values are selected for categorical variables. Samples from a normal distribution are used to perturb data for continuous variables, and the perturbed value is added to the original. On the perturbed data, predictions are made using the predictor function of the black box model. These forecasts are utilized to train LIME's local linear model. The original and perturbed data points are compared for each data point, and the Euclidean distance between them is found. A sense of the point's distance from the initial observation can be obtained using Euclidean distance. A smaller distance implies the data point is closer to the observation. LIME typically uses a kernel function to assign weights to the perturbed data points based on their distance, and these weights or similarity scores are then used in the weighted linear regression model. The closer the point to the observation, the higher the similarity score. The model could be run with n features which are selected using feature selection methods like highest weights, forward selection & lasso path. With all the input data prepared, LIME creates the local linear model that can be used to explain the predictions. The coefficients or the weights from the linear regression model, combined with the features of the perturbed instances, are used to explain the local behavior of the instance.

LIME and SHAP have different strengths and weaknesses, and the choice between the two depends on the specific use case, the complexity of the model, and the trade-off between computational cost and interpretability.

The steps involved in the LIME algorithm are shown in Table 3.

The algorithm of LIME assumes that every complex model is linear on a local scale. LIME seeks to fit a basic model around a single observation that simulates the behavior of the global model at that location.

The optimization used in LIME can be represented as

$$\xi(x) = \frac{argmin}{g \varepsilon G} L(f, g, \Pi x) + \Omega(g) \tag{6}$$

x - Input data point
f - Complex model

TABLE 3. Algorithm of LIME.

Step 1 :	Perturb the observation n times to create replicated feature data with minor value changes for the observation to be explained. This disturbed data is a fictitious data set constructed around the observation for LIME to use in constructing the local linear model.
Step 2 :	Predict the outcome of the perturbed data
Step 3 :	Calculate the distance from each perturbed data point to the original observation
Step 4 :	Convert distance to the similarity score
Step 5 :	From the perturbed data, select m features that best describe the predictions
Step 6 :	Fit a simple model to the perturbed data for the selected features
Step 7 :	Feature weights (coefficients) of the simple model are the explanations of the observation

g - Simple interpretable local model such that $g\varepsilon$ G, where G is the family of interpretable models
L (f,g, Πx)- Denotes approximation of complex model f by the simple model g in the neighborhood of data point x
Ω (g) - regularization term

IV. RESULTS AND DISCUSSIONS

The features included in the dataset can be classified as Socioeconomic features, prior academic features, and deafness-related features. EDA on the collected data showed that the distribution of semester 1 marks is normally distributed around the mean value of 408. The majority of students have 100% of disability and they fall in the profound or severe type of hearing loss category. The graphical representation is shown in Figure 6.

Figure 7 shows that there is no significant variation in the distribution of semester 1 marks based on hearing aid usage. The communication mode seems to be an important factor in academic performance. The students who use both sign language and speech scored more than those who depend on sign language alone for communication. Though the number of speech users is very low, their performance is much better than others. There is not much difference in performance based on the type of hearing loss.

It is observed that the majority of students studied in special schools and the majority received speech therapy below the age of 5. The early intervention can be a factor that the majority of students used both sign language and speech for communication which in turn helped in their academic performance. As the majority studied in special schools, there are high chances of them using sign language as the primary mode of communication and speech as a supporting mode of communication. The result is graphically represented in Figure 8.

The feature Grade is added depending on the semester marks. The distribution of features with the grades is shown in Figure 9. It is observed that the majority of students are middle-level performers. Though the mean semester mark is almost the same for male and female students, the high performers are more male students when compared to female

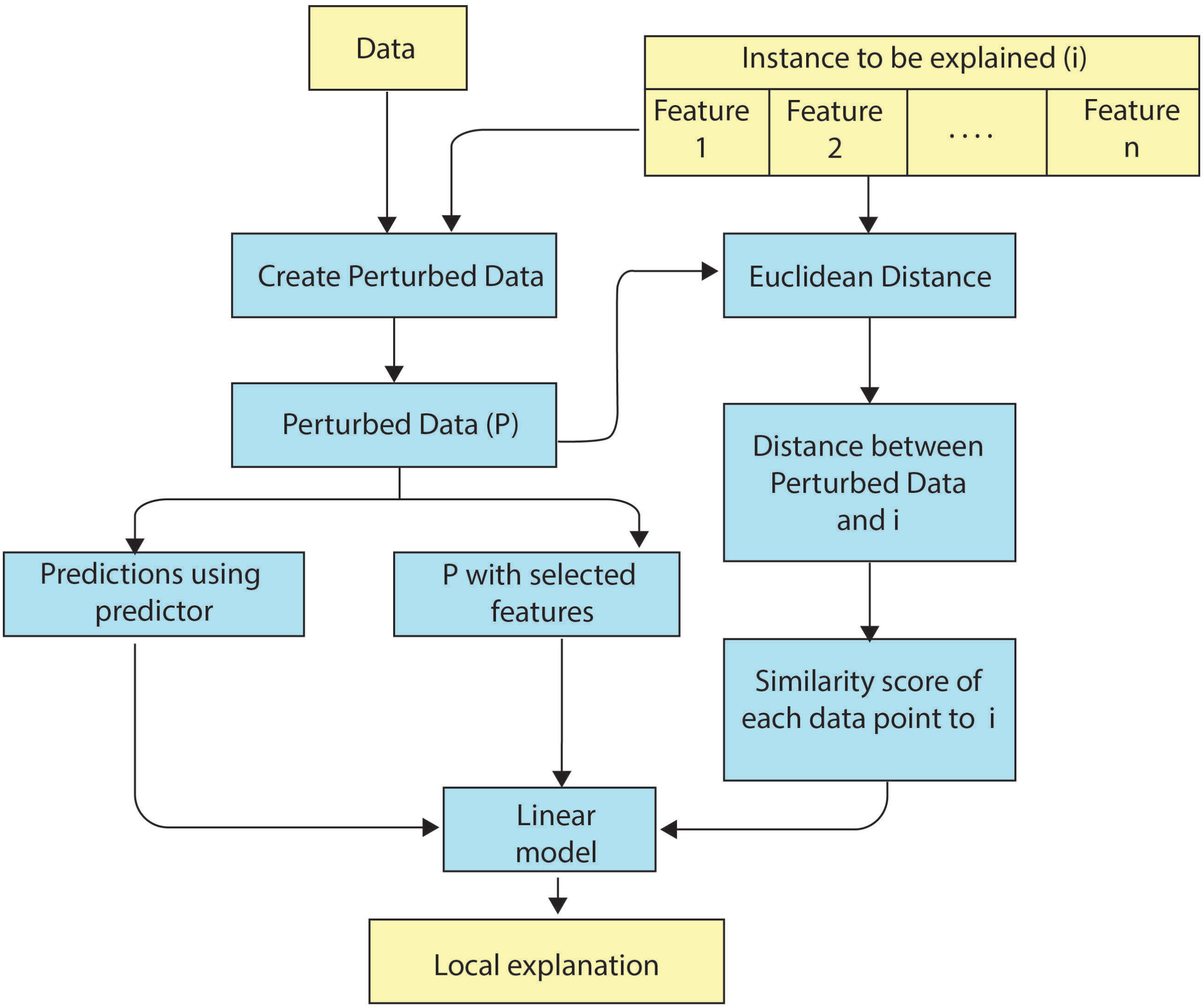

FIGURE 5. Working of LIME.

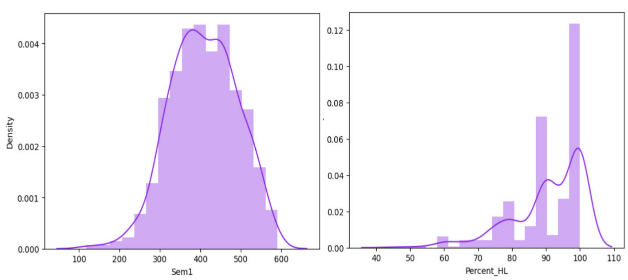

FIGURE 6. Distribution of semester1 marks and percentage of disability.

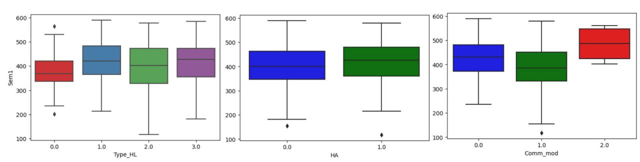

FIGURE 7. Distribution of semester marks with various deafness-related features.

students. The majority of high-level performers use sign language as a communication mode. The importance of early intervention in academic performance is evident from the fact that the majority of middle-level performers have received early intervention. The majority of students do not have a history of deafness in the family. The number of students who have undergone cochlear implants or use hearing aids is very low. The expenses and lack of awareness about the advantages of both can be a reason for this. There are more performers in the sensorineural deafness category and students with profound deafness also perform well compared to the other degrees of deafness.

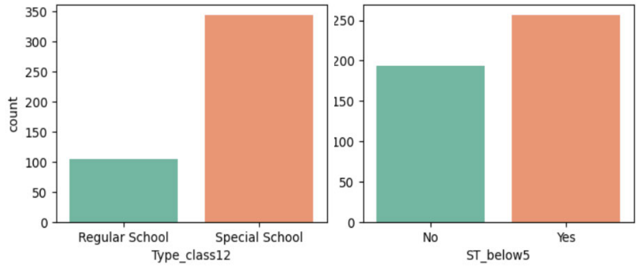

FIGURE 8. Distribution of type of schooling and early intervention.

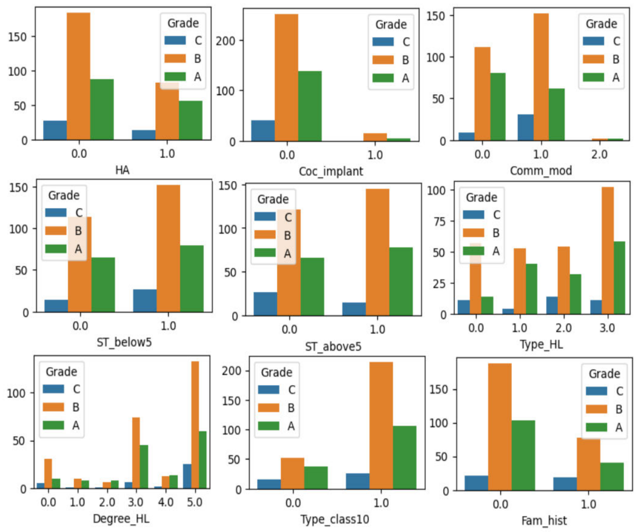

FIGURE 9. Distribution of features with grade.

The balanced data is modeled with ML algorithms - Logistic Regression (LR), K-NN, Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), XGBoost, ExtraTrees Classifier and combination models- SVM and KNN, DT and RF, XGBoost and ExtraTreesClassifier, Stacked model of Random Tree, XGBoost and ExtraTrees Classifier. Also, transfer learning models like TabNet, VGG16, and ResNet50 were used to model the original balanced data. The performance of all models was compared in terms of accuracy, recall, precision, and F1 Score. The results are tabulated in Table 4.

The performance of different models is graphically represented in shown in Figure10. Comparing the performance of different models it is found that the performance of the Stacked model of Random Tree, XGBoost, and ExtraTrees Classifier is better compared to the other models. The performance of the Decision Tree is considerable and comparable to that of KNN and Random Forest. This indicates that the data might have some non-linear relationships. SVM and Naïve Bayes seem to struggle with the dataset. The non-linearity or complexity of decision boundaries might have reduced the performance of SVM. The poor performance of Naïve Bayes can be due to the feature independence assumption. In the case of KNN, the precision is slightly lower. The better performance of Random Forest indicates the benefits of ensemble methods. It has captured complex relationships in the data, outperforming the decision tree. With its gradient-boosting approach, XGBoost is able to capture different patterns than Random Forest. The Extra Trees Classifier also performed well, indicating its effectiveness in capturing complex relationships in the data. There is no significant improvement in performance with the ensemble of SVM and KNN. This shows that the individual weakness of the individual classifier persists. It is observed that the performance of the majority of ensemble models combining different classifiers is better, which indicates that combining these models enhances the predictive capabilities and combines the strengths of individual classifiers. The stacked ensemble of XGBoost, Random Forest, and Extra Trees achieves the highest performance, indicating the effectiveness of combining diverse models. This ensemble model is likely to capture a wide range of patterns in the data.The best parameters of each model in the stacked classifier obtained with GridsearchCV are specified in Table 5.

The accuracy of this stacked model is found to be 92.99% and ROC AUC is .9781. The high accuracy indicates that the model is performing well in terms of overall correctness. The high precision indicates that when the model predicts a positive class, it is often correct. The high recall indicates that the model effectively captures a large proportion of the actual positive instances. The F1 score, being a combination of precision and recall, provides a balanced assessment of the model's performance. The confusion matrix, sensitivity, specificity, and ROC of this model are shown in Figure 11. The performance of this model is further evaluated with Mean Square Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2). The values obtained are shown in Table 6.

The average squared difference between the expected and actual data is represented by the MSE (Mean Squared Error). A smaller number (0.16) denotes better model performance. The square root of the mean square error, or RMSE (0.39), shows better model accuracy. With a mean absolute error of 0.12, the accuracy is also better. The R-squared (R2) value, which ranges from 0 to 1, represents how well the model explains variation. A score of 1 denotes a perfect fit. With an R2 of 0.77, the model is deemed to be satisfactory as it accounts for 77% of the variation. All things considered, the model works effectively, capturing a significant amount of data variation with minor errors.

To evaluate the generalizability of the stacked model, the k-fold cross-validation technique is used with 10 folds. The average accuracy is found to be 92.95 with a standard deviation of 0.012 among the folds.

SHAP summary graphic is a model's overall explanation that combines feature importance and feature effect which is represented in Figure 12. The features are represented on the Y axis and the average impact on the model output is represented on the X axis. In this plot, the mean absolute value for each feature over all the samples is used to

TABLE 4. Performance comparison of different classification models.

Sl No	Model	Accuracy	Precision	Recall	F1Score
Transfer Learning models (Balanced Original Data)					
1	TabNet	75.43	75.05	75.52	75.28
2	VGG16	70.71	70.05	70.68	69.91
3	ResNet50	71.43	70.94	71.40	70.70
ML models (Final Data after Synthetic Data Generation)					
1	Decision Tree (DT)	75.68	75.78	75.92	75.66
2	Support Vector Machine(SVM)	59.75	59.76	59.75	59.76
3	K-NN	75.83	75.49	75.83	75.49
4	Naïve Bayes (NB)	56.00	56.30	56.00	56.30
5	Random Forest(RF)	87.84	88.74	87.84	88.74
6	XGBoost(XG)	85.28	85.99	85.28	85.99
7	ExtraTreesClassifier(XT)	92.66	92.70	92.65	92.71
8	DT+RF	81.54	81.62	83.39	80.70
9	SVM+KNN	67.03	70.22	67.03	70.22
10	XT+XG	89.70	89.70	90.10	89.50
11	RF+XT	90.55	91.03	90.55	91.03
12	XG+RF+XT	92.99	93.34	92.99	93.34

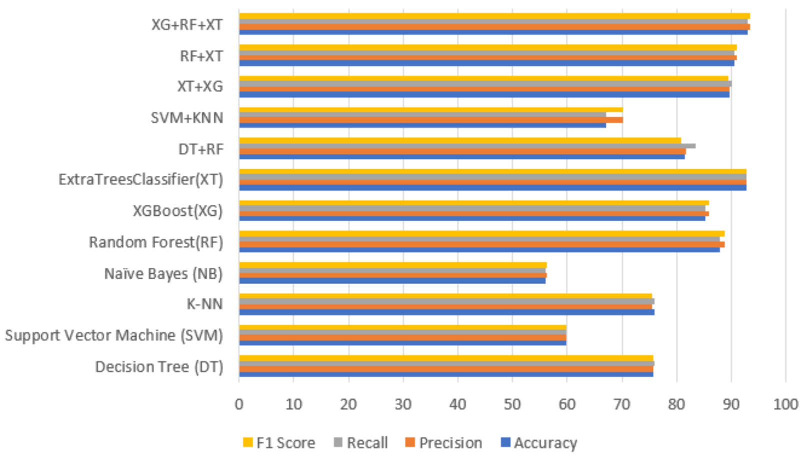

FIGURE 10. Performance comparison of different classification models.

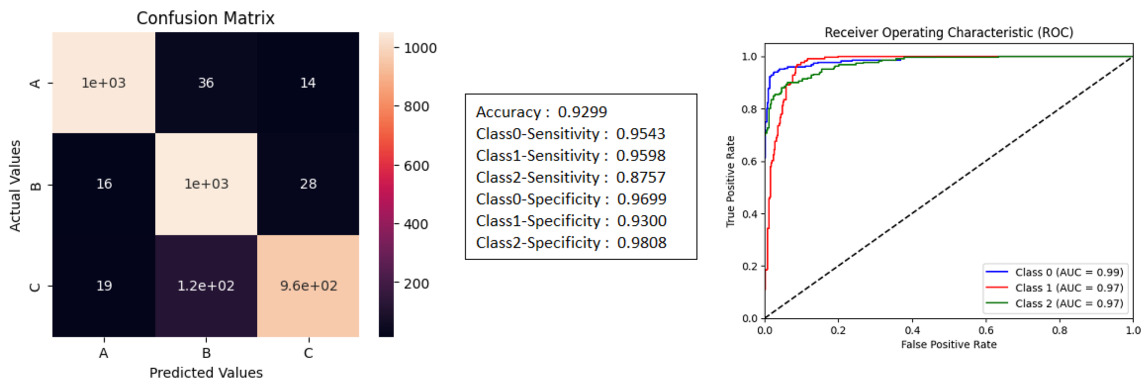

FIGURE 11. Confusion matrix, sensitivity, specificity and roc of stacked model.

determine its global relevance [37]. The feature relevance to each class is also shown. It is seen that the communication mode, speech therapy obtained after age 5, and family history are the most relevant features. The relevance of these features in identifying different levels of performers is different. The importance of the type of intervention through speech therapy is more in identifying middle-level performers while the relevance of family history is equal

TABLE 5. Best Parameters obtained for Stacked Classifier.

Random Forest Model	
max_depth	32
min_samples_leaf	1
n_estimators	256
XGBoost	
learning_rate	0.1
max_depth	5
n_estimators	300
ExtraTreeClassifier	
min_samples_leaf	1
min_samples_split	2
n_estimatorsv 300	

TABLE 6. Performance metrics of best performing stacked model.

Sl No	Metrics	Value
1	Root Mean Square Error(RMSE)	0.39
2	Mean Square Error(MSE)	0.16
3	Mean Absolute Error(MAE)	0.12
4	R-Squared(R2)	0.77

in identifying low and high level performers. Similarly, the impact of communication mode is more in classifying the high-level performers but also significant in classifying low and middle-level performers. From Figure 12, it is evident that deafness-related factors are influencing the classification of DHH students in a significant way.

In Figure 13, a point on a summary plot represents a shapely value for a feature and a specific sample. The Y axis represents features, while the X axis represents shapely values. Low and high values are represented by colors. The negative values in the X-axis show a negative impact while the positive values show a positive impact. The features in the summary plot are organized in order of relevance, with the top feature being the most essential and the bottom feature being the least [32]. Figure 14 shows the impact of features on the classification of the model towards class B. It can be observed that the Communication mode is the most relevant feature in classification. It is also seen that family history, the intervention methods like speech therapy also play an important role in classifying the student as a middle-level performer. The lower values of communication mode have a positive impact on the prediction of middle-level performance. That means that students who use both sign and speech for communication tend to perform better.

SHAP local explanations evaluate only one occurrence at a time and generate an explanation, indicating which feature values lead to good decisions and which lead to negative decisions for that particular instance. In Figure14 three force plots, one for each target class are shown. The model predicted the class with the highest score that is,4.06. The model classified the student as a middle-level performer or class B performer. Features pushing the prediction higher are shown in red and those pushing predictions lower are shown in blue. The explanation changes as we change the input instance. For this particular student, the model prediction and

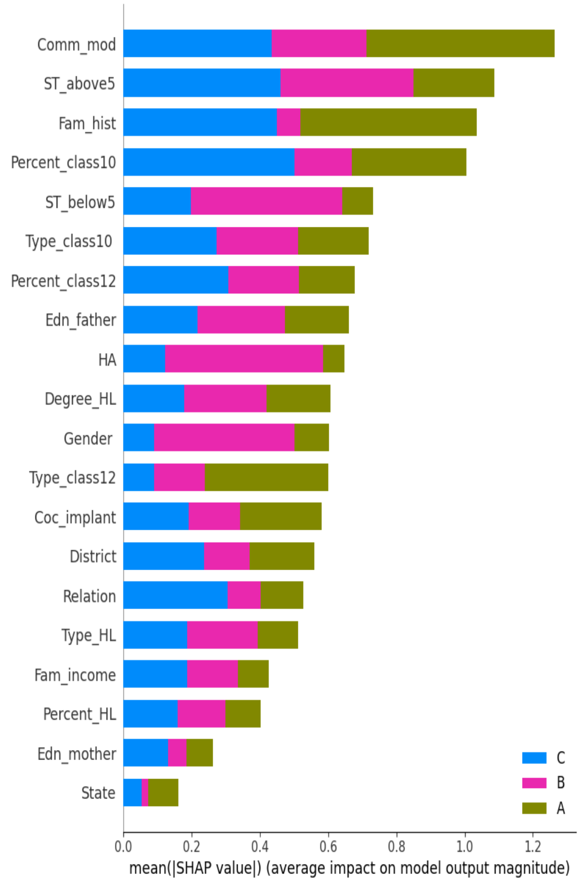

FIGURE 12. Summary plot using stacked model indicating communication mode as a significant feature.

the ground truth label are the same. That is, the prediction made by the model is correct. The major features that affected the prediction are speech therapy, education of the father, Communication mode, and Percentage of class10. For this instance, most of the features impact positively to classify the student as a middle-level performer.

Analyzing further three students were considered who belong to high level, middle level, and low level performance. For the particular instances, the features contributed to the classification are shown in Table 7 and the force plot is represented in Figure 15.

For an instance that falls in the Class B category (f(x)-4.06), the communication mode has a positive value(1.037) which indicates a positive impact on the student's performance, reducing the prediction below the baseline. This suggests that the communication mode used by the student is optimal for his learning and may be positively affecting his performance. On the other hand, the speech therapy, education of the father, and percentage of class 10 have a negative value which indicates that these factors are contributing to his performance negatively. This recommends that policymakers pay attention to this and should trigger more research avenues on this area need to be more focused

TABLE 7. Feature contributions - selected instances from each class.

Sl No	Instance 1: Grade A	Instance 1: Grade B	Instance 1: Grade C
f(x)	5.22	4.06	5.27
Features Contributes			
	Type of Schooling	Speech Therapy above age 5	Speech Therapy above age 5
	Communication Mode	Education of father	District
	Type of Hearing Loss	Communication Mode	Speech Therapy below age 5
	Percentage of class10	Percentage of class10	Education of mother

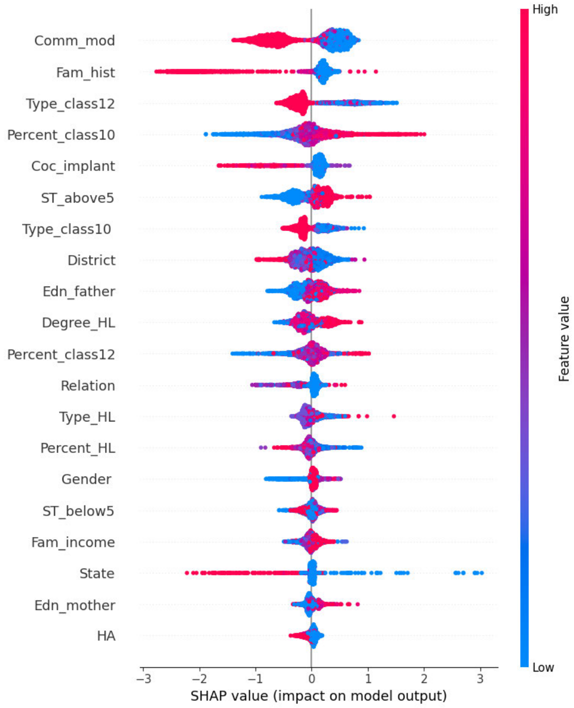

FIGURE 13. SHAP summary plot - high value of communication mode increases the impact on model output.

to solve the underlying issues. The teachers need to review the teaching pedagogies followed in school and this student should give adequate support to perform well in school which will help to perform better later in his higher education. On the other hand, the feature, type of hearing loss has a positive impact (1.783) on the high-level student performer, increasing the prediction above the baseline. This suggests that the student's hearing abilities may be positively affecting their performance. The recommendations can be given to policymakers so that more assistive support systems can be provided to students with hearing loss.

Though the explanation of specific instances is considered, it is observed that the deafness-related factors contribute highly to the classification. Considering the fact that the major barrier for a DHH student is communication and as the mode of communication is a deciding feature, the communication system followed in colleges needs to be studied and the measures to improve the pedagogies need to be considered.

LIME is used to generate local explanations for the classification. Figure 16 shows the explanation for determining whether the classification result is 'B' or 'Not B,' as well as the probability and instance values. Colors are used to draw attention to which features contribute to which class. The orange features contribute to the 'B' category, whereas the blue features contribute to the 'Not B' category. ML models should be transparent to end users. They should receive answers to all of their questions, such as why the model reached a particular decision, what circumstances led to this decision, how the model's decision can be revised, and so on [38]. Here the 10 features that contributed significantly to the decision are considered. For this particular student, the usage of the hearing aid, speech therapy, percentage of hearing loss, family income, percentage of class 10, and family history are the significant factors that contributed to the classification decision. In this case, the speech therapy is one of the very significant features. The usage of hearing aids also contributes to this prediction. As the Hearing aid usage is having a positive impact on this student's performance, this needs to be encouraged. In this particular instance, the student uses speech as his communication mode. This could be one of the reasons the usage of hearing aids is beneficial. As technology has advanced, this student may benefit more in classrooms with FM Systems which are usually used to delete unwanted noise. He may also benefit from the advanced digital hearing aid. The appropriate recommendations can be given to the teacher of this student to adopt suitable pedagogies in class. Even this student's performance is also connected to his family economy. It was found that his family income is between 5,000 and 1 Lakh rupees, which restricts him from getting better resources for his development. So policymakers can take measures to support the student with financial assistance to procure a digital hearing aid.

The ML-based model developed to classify students with deafness according to their academic performance will be useful if it is deployed and monitored among the user population which includes teachers, policymakers and researchers. A prototype of such an end-user interface is proposed as shown in Figure 17. The interface is developed using the open-source Python library, Streamlit which is commonly used to develop web applications using Machine Learning models. As a prototype, there is a huge chance of modification with valuable feedback from the end users. Based on the feedback obtained, the modifications to the web application like the option to upload a set of students' data as. csv file, Search facility, and database connectivity are identified as the future scope of this project.

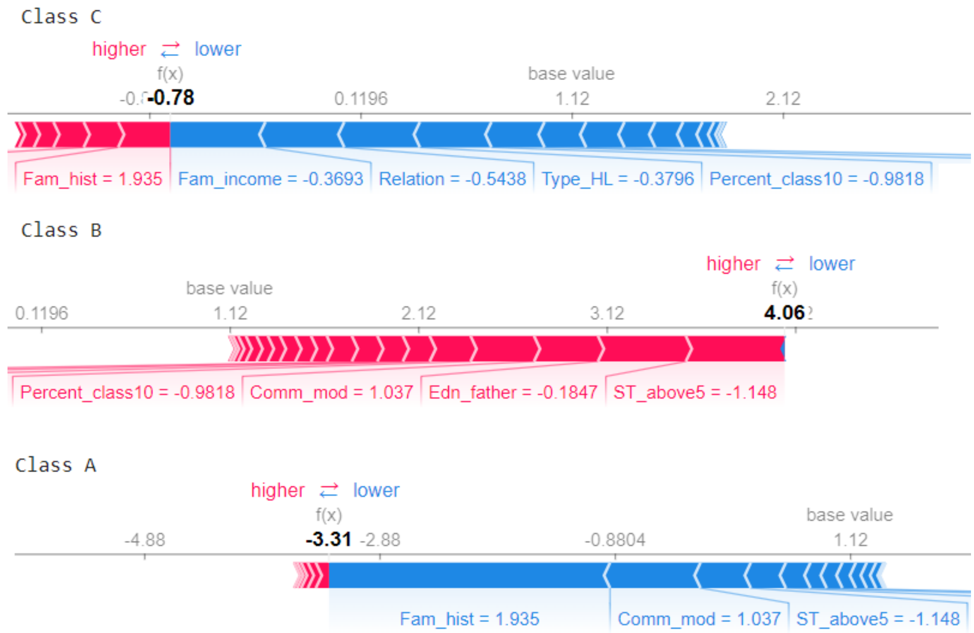

FIGURE 14. SHAP force plot of a middle-level performer.

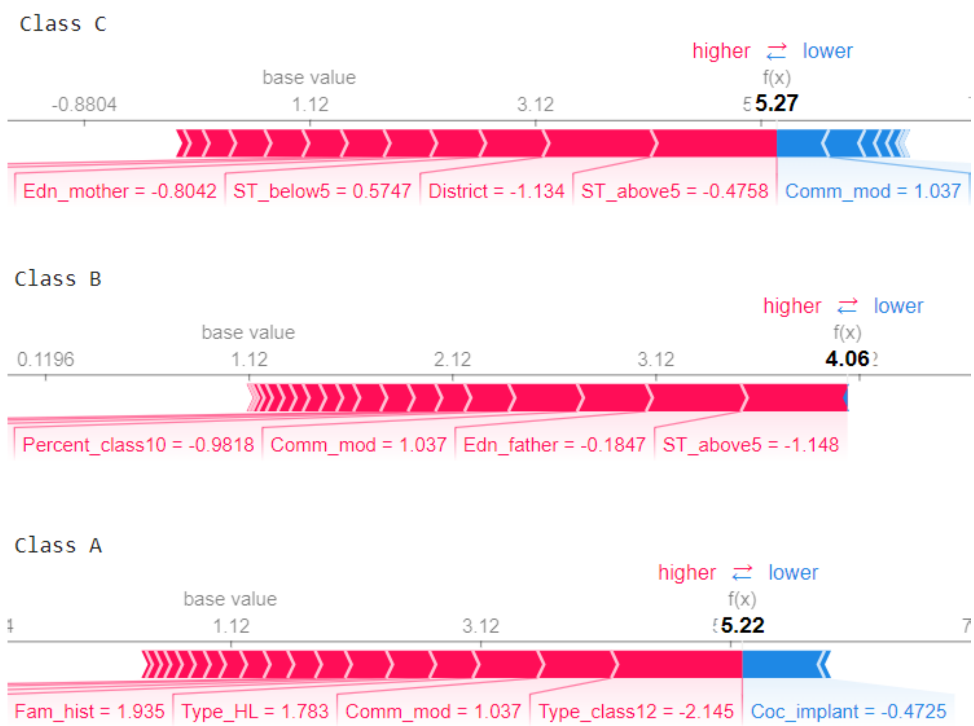

FIGURE 15. SHAP force plot of a high-level, middle-level, and low-level performer.

Using the Dashboard developed, a pilot study with the domain experts is done. According to the stakeholders in an educational context, there are several benefits in assigning students to groups according to their academic standing. With the use of this technique, teachers may provide teaching pedagogies that are specifically suited to the requirements and skills of each group. More homogenous, smaller groups allow for more focused interventions and individualized

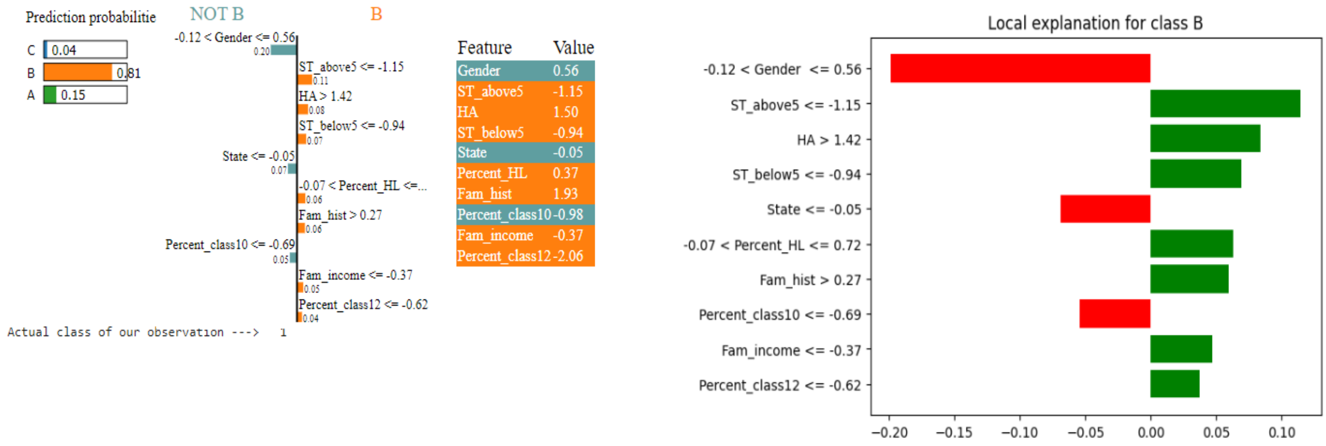

FIGURE 16. Explanation using LIME.

attention, which promotes a positive learning environment. Students with comparable intellectual backgrounds gain from more interactions with their peers and could feel more confident. Positive effects include effective resource allocation, differentiated curriculum, better classroom management, and increased student involvement. It is also important to check for potential limitations that may arise to ensure flexibility and inclusion in the teaching methodology. It is crucial to regularly re-evaluate students' academic standing to meet their changing demands.

For further analysis, one instance from each class is chosen and the feedback from an educator, a clinician, and an administrator is collected. The instances chosen are shown in Table 8. The feedbacks from the above experts are explained below.

Instance 1- Low-level Performer

Educator's recommendation -

- Assess the student's willingness to adapt and potential benefits from hearing aids through an audiological evaluation.
- Recognize the student's dual communication mode involving both sign language and speech.
- Consider employing a combination of vocalization and sign language in instruction to cater to the student's communication needs effectively.

Clinician's recommendation -

- The degree of hearing loss alone is not a direct predictor of academic performance; the area and extent of damage to the auditory system play a crucial role.
- Periodically evaluate mixed hearing loss and consider pharmacological or surgical interventions to manage the conductive component, leading to improved auditory perception, communication, and scholastic skills.
- Prioritize the use of assistive devices, such as hearing aids or cochlear implants, for rehabilitation, considering the individual's characteristics of hearing loss.
- Implement intensive speech and language stimulation to aid the participant in acquiring age-appropriate communication skills.

Administrator's recommendation -

- Considering the education and socioeconomic status of parents, as higher education and socioeconomic status may contribute to better speech and language skills. Parents with better resources are more likely to choose assistive devices and seek professional support. Financial support for the student through scholarships or sponsorship can be considered.
- Government grants to procure assistive devices can be recommended

Instance 2- Middle-level Performer

Educator's recommendation -

- Consider medical or surgical treatment for the mid-level performer with conductive hearing loss, especially if the middle ear is damaged.
- Explore the option of fitting suitable hearing aids based on the audiological assessment.
- Recognize the student's dual communication mode involving both speech and sign language, necessitating the use of vocalization and sign language in instruction.
- Ensure accessibility by captioning audio and video presentations.
- Provide transcriptions for live presentations to accommodate the student's needs.
- Verify the potential benefits of an FM system through audiological assessment for optimal support.

Clinician's recommendation -

- Periodically assess participants with conductive hearing loss for potential pharmacological or surgical interventions, as effective management can alleviate the degree of hearing loss and improve auditory perception, communication, and scholastic skills.
- Emphasize the use of assistive devices, such as hearing aids or cochlear implants, for rehabilitation, especially for candidates with conductive hearing loss.
- Implement intensive speech and language stimulation to aid participants in acquiring age-appropriate communication skills.

TABLE 8. Features of the instances selected from each class.

Features	Low Level Performer	Middle Level Performer	High Level Performer
Percentage of Disability	80	60	90
Percentage marks-Class 10	55	99	84
Percentage marks-Class 12	65	95	63
Gender	Male	Male	Female
Family Income	Below 50000	Below 50000	50000-1 Lakh
Education of Father	Below High School	Below High School	Plus Two
Education of Mother	Below High School	Below High School	Under Graduation
District	Thrissur	Wayyanad	Pathanamthitta
State	Kerala	Kerala	Kerala
Type of Hearing Loss	Mixed	Conductive	Sensorineural
Degree of Hearing Loss	Severe	Moderately Severe	Profound
Speech Therapy below Age5	No	No	No
Speech Therapy above Age5	No	No	Yes
Hearing Aid Usage	No	No	Yes
Cochlear Implant	No	No	No
Communication Mode	Both Speech and Sign Language	Both Speech and Sign Language	Both Speech and Sign Language
Family History	Yes	Yes	No
Relation	Father /Mother/Sibling	Father /Mother/Sibling	No Family Member is Deaf
Type of Schooling-Class10	Special School	Special School	Regular School
Type of Schooling -Class12	Special School	Regular School	Regular School

- Recognize the potential hindrance to speech and language skill development in the absence of early auditory stimulation and professionally guided interventions.

Administrator's recommendation -

- Financial aid to support the student needs
- Implementation of classroom accommodations needed for the student.

Instance 3- High-level Performer

Educator's recommendation -

- This instance of the high-level performer has profound sensory deafness and lacks a family history of hearing loss. The communication of the student with the family needs to be checked.
- When choosing the best teaching method, evaluate the student's proficiency in sign language and combine vocalization and sign language as necessary.
- Examine the possibility of using an FM system by conducting an audiological evaluation, keeping in mind that the student may wear hearing aids.
- By offering captions and live transcriptions, you may make use of the students' great reading and writing abilities to improve communication and involvement in class.

Clinician's recommendation -

- In light of the participant's age at identification and the specifics of their hearing loss, emphasize the use of assistive equipment, such as hearing aids, as a critical component of rehabilitation.
- Encourage participants to learn age-appropriate communication skills by providing them with extensive speech and language stimulation, especially if they are wearing hearing aids.
- Recognise the benefits of speech therapy in improving speech and language abilities, which may lead to better academic performance and literacy skills, even if it is started after the age of five.
- As this student do not have a family history of hearing loss, he may have good verbal communication role models. After evaluating this, more support and resources can be provided.

Administrator's recommendation -

- Teachers training programs recommended to meet the s of student
- As the student has 90% disability and highly benefits from intervention methods, recommendations to promote intervention methods can be made.

An attempt was made to assess the faithfulness of the interpretations obtained. Using XAI methods, significant features that contribute to the academic performance of deaf students are obtained and represented as the global interpretation of the stacked model using SHAP values. To quantitatively analyze this interpretation, 16 educators who have experience more than 10 years were asked to rank the features that contribute to deaf students' performance in general. The Spearman rank correlation coefficient, , is calculated in each case as below. The values are plotted in Figure 18.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (7)$$

where di is the difference between the ranks of ith feature and n is the number of features.

A Spearman correlation value of 0.4238, on average among 16 experts, indicates a modest positive monotonic connection between the model's permutation significance rankings and the rankings supplied by the experts. The value of 0.42308 implies that the monotonic association is somewhat strong. It is not a high association, but it is not weak. The positive Spearman correlation shows that the model's judgment of feature importance and expert opinions are generally in agreement. Individual coefficients offered by each expert may differ, but the fact that the average is positive

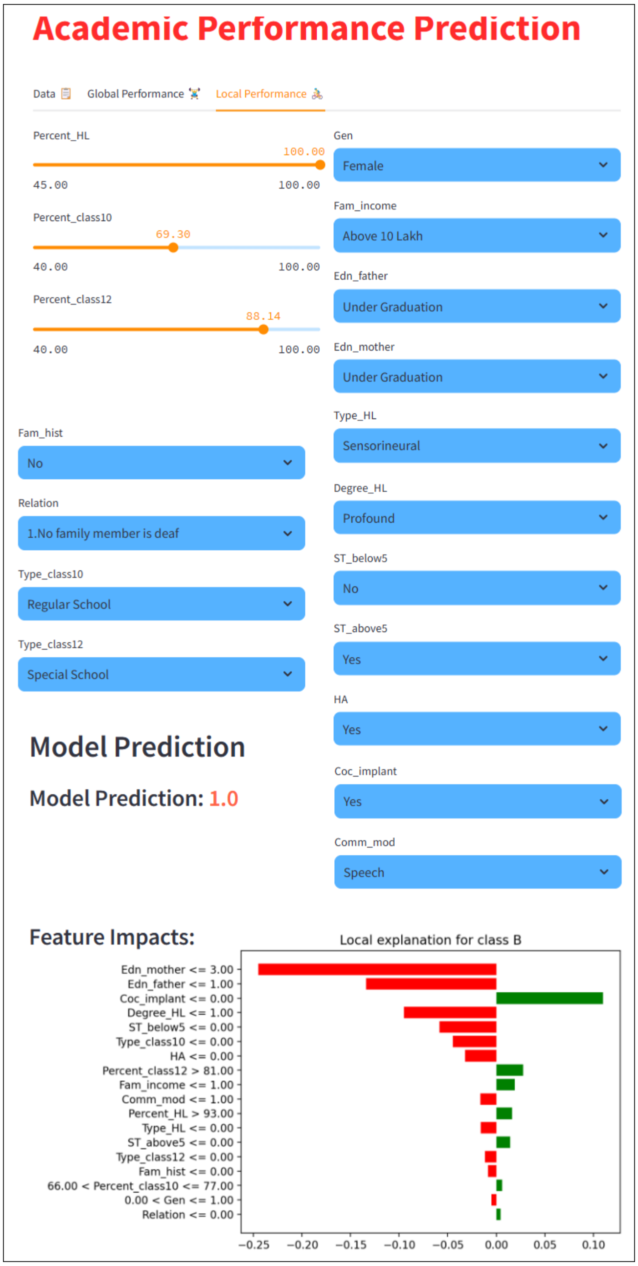

FIGURE 17. Web application for predicting the performance indicator using streamlit.

implies that the ranks are aligned in the right direction. The positive association suggests that characteristics scored higher in permutation importance by the model are also ranked higher by domain experts. This consistency is good since it indicates that the model's understanding of feature value aligns relatively well with human expert opinions.

The similarity score parameter is used to determine how effectively. LIME describes specific instances. Ten instances are selected randomly for this investigation. The cosine similarity is calculated between each instance's explanation feature weights and the model's feature permutation importance. Figure 19 depicts the results. The average similarity score for these ten examples was found to be 0.7518. The scores were consistently similar across instances, showing that the model's general behavior matches the instance-level explanations effectively. However, there were two instances where the ratings were inconsistent, indicating that a feature is generally important but has a varied impact on specific instances.

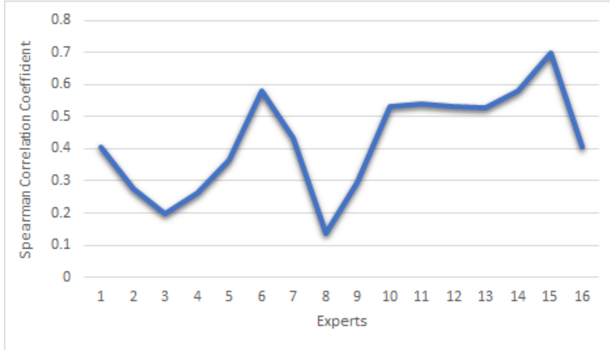

FIGURE 18. Spearman correlation coefficients obtained with domain experts.

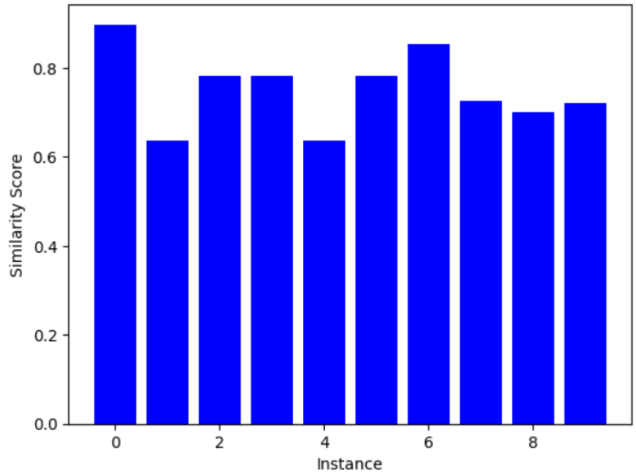

FIGURE 19. Similarity scores - 10 instances.

V. CONCLUSION

In this paper, an attempt is made to propose a model that can classify a DHH student according to his academic performance. Unlike the students with hearing, the factors related to deafness have a significant role in the academic performance of DHH students. This leads to many decision-making policies that will result in more opportunities for the higher education of these students. This work also examines the possibility of multi-label classification interpretability. By adding explainability, the model is made transparent. The stakeholders of the education domain can get trustable predictions using this model. An interactive dashboard development can be considered as a scope of future work for this research. As the next phase of this work, more domain experts can be involved and the modification of the dashboard can be done. The model interpretations can be evaluated with more domain experts and can be compared to see the validity of interpretations.

The contribution of this study includes the development of a dataset comprising details of students with deafness. From the results, it is observed that the existing gap identified which includes the small data size - imbalance problem can be solved to an extent by synthetically generating data and SMOTE technique. It is found that the proposed stacked model along with XAI can give valuable insights to the stakeholders of deaf education which can improve the education opportunities and conditions of students with hearing loss. The ML model which is usually considered as a black box can be made transparent using XAI techniques. Individual predictions can be evaluated and analyzed which can help the stakeholders to make better decisions. In the deaf education domain, such models are rarely used. This study is a stepping stone to introduce possibilities of ML and XAI in deaf education which can change the lives of many students with deafness.

COMPETING INTEREST

The authors declare that they have no competing interests.

CONTRIBUTIONS

All the authors contributed equally to research design, implementation, result analysis, and manuscript writing

REFERENCES

[1] R. K. Singh and S. K. Mahapatra, "Asian journal of distance education education of deaf learners through open schooling system in India," *Asian J. Distance Educ.*, vol. 14, no. 2, pp. 26–31, 2019.

[2] K. Mandke and P. Chandekar, "Deaf education in India," in *Deaf Education Beyond the Western World*. London, U.K.: Oxford Univ. Press, 2019, pp. 261–284, doi: 10.1093/oso/9780190880514.003.0014.

[3] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Exp. Syst. Appl.*, vol. 33, no. 1, pp. 135–146, Jul. 2007, doi: 10.1016/j.eswa.2006.04.005.

[4] B. S. Wilson, D. L. Tucci, M. H. Merson, and G. M. O'Donoghue, "Global hearing health care: New findings and perspectives," *Lancet*, vol. 390, no. 10111, pp. 2503–2515, Dec. 2017, doi: 10.1016/s0140-6736(17)31073-5.

[5] S. D. Antia, P. B. Jones, S. Reed, and K. H. Kreimeyer, "Academic status and progress of deaf and hard-of-hearing students in general education classrooms," *J. Deaf Stud. Deaf Educ.*, vol. 14, no. 3, pp. 293–311, Jul. 2009, doi: 10.1093/deafed/enp009.

[6] R. Elbeltagy, "Prevalence of mild hearing loss in schoolchildren and its association with their school performance," *Int. Arch. Otorhinolaryngol.*, vol. 24, no. 1, pp. e93–e98, Jan. 2020, doi: 10.1055/s-0039-1695024.

[7] I. Hrastinski and R. B. Wilbur, "Academic achievement of deaf and hard-of-hearing students in an ASL/English bilingual program," *J. Deaf Stud. Deaf Educ.*, vol. 21, no. 2, pp. 156–170, Apr. 2016, doi: 10.1093/deafed/env072.

[8] S. Literate and J. I. Indonesia, "Factors influencing academic performance of hearing impaired students in inclusive education: A case of moshi techni," *Digit. Libr. Open Univ. Tanzania*, vol. 1, no. 2, pp. 274–282, 2020.

[9] K. Kuppler, M. Lewis, and A. K. Evans, "A review of unilateral hearing loss and academic performance: Is it time to reassess traditional dogmata?" *Int. J. Pediatric Otorhinolaryngol.*, vol. 77, no. 5, pp. 617–622, May 2013, doi: 10.1016/j.ijporl.2013.01.014.

[10] E. E. M. Agyire-Tettey, M. Cobbina, and E. S. Hemanoo, "Academic challenges of students with hearing impairment (Shis) in Ghana," *Disab. CBR Inclusive Develop.*, vol. 29, no. 3, pp. 127–150, 2017, doi: 10.5463/dcid.v28i3.646.

[11] H. Knoors and M. Marschark, "Language planning for the 21st century: Revisiting bilingual language policy for deaf children," *J. Deaf Stud. Deaf Educ.*, vol. 17, no. 3, pp. 291–305, Jul. 2012, doi: 10.1093/deafed/ens018.

[12] N. Alsraisri, "Predicting academic performance of deaf and hard of hearing students using predicting academic performance of deaf and hard of hearing students using educational data mining techniques," Ph.D. thesis, Dept. Spec. Educ., King Saud Univ., Riyadh, Saudi Arabia, Jun. 2021, doi: 10.13140/RG.2.2.21887.94881.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[14] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec./Nov. 2017, pp. 4766–4775.

[15] R. Alamri and B. Alharbi, "Explainable student performance prediction models: A systematic review," *IEEE Access*, vol. 9, pp. 33132–33143, 2021, doi: 10.1109/ACCESS.2021.3061368.

[16] R. S. Yadav, "Application of hybrid clustering methods for student performance evaluation," *Int. J. Inf. Technol.*, vol. 12, no. 3, pp. 749–756, Sep. 2020, doi: 10.1007/s41870-018-0192-2.

[17] M. Christina, "Predicting student performance using data mining," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 172–177, Oct. 2018, doi: 10.26438/ijcse/v6i10.172177.

[18] S. Al-Sudani and R. Palaniappan, "Predicting students' final degree classification using an extended profile," *Educ. Inf. Technol.*, vol. 24, no. 4, pp. 2357–2369, Jul. 2019, doi: 10.1007/s10639-019-09873-8.

[19] S. Ikbal, A. Tamhane, B. Sengupta, M. Chetlur, S. Ghosh, and J. Appleton, "On early prediction of risks in academic performance for students," *IBM J. Res. Develop.*, vol. 59, no. 6, pp. 5:1–5:14, Nov. 2015, doi: 10.1147/JRD.2015.2458631.

[20] M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, and S. U. Khan, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," *IEEE Access*, vol. 9, pp. 7519–7539, 2021, doi: 10.1109/ACCESS.2021.3049446.

[21] M. Haridas, G. Gutjahr, R. Raman, R. Ramaraju, and P. Nedungadi, "Predicting school performance and early risk of failure from an intelligent tutoring system," *Educ. Inf. Technol.*, vol. 25, no. 5, pp. 3995–4013, Sep. 2020, doi: 10.1007/s10639-020-10144-0.

[22] L. Ramanathan, G. Parthasarathy, K. Vijayakumar, L. Lakshmanan, and S. Ramani, "Cluster-based distributed architecture for prediction of student's performance in higher education," *Cluster Comput.*, vol. 22, no. S1, pp. 1329–1344, Jan. 2019, doi: 10.1007/s10586-017-1624-7.

[23] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of students' academic performance based on courses' grades using deep neural networks," *IEEE Access*, vol. 9, pp. 140731–140746, 2021, doi: 10.1109/ACCESS.2021.3119596.

[24] A. Zollanvari, R. C. Kizilirmak, Y. H. Kho, and D. HernáNdez-Torrano, "Predicting students' GPA and developing intervention strategies based on self-regulatory learning behaviors," *IEEE Access*, vol. 5, pp. 23792–23802, 2017, doi: 10.1109/ACCESS.2017.2740980.

[25] C.-Y. Ko and F.-Y. Leu, "Examining successful attributes for undergraduate students by applying machine learning techniques," *IEEE Trans. Educ.*, vol. 64, no. 1, pp. 50–57, Feb. 2021, doi: 10.1109/TE.2020.3004596.

[26] M. Zaffar, M. Ahmed, K. S. Savita, and S. Sajjad, "A study of feature selection algorithms for predicting students academic performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, pp. 541–549, 2018, doi: 10.14569/ijacsa.2018.090569.

[27] M. Afzaal, J. Nouri, A. Zia, P. Papapetrou, U. Fors, Y. Wu, X. Li, and R. Weegar, "Explainable AI for data-driven feedback and intelligent action recommendations to support students self-regulation," *Frontiers Artif. Intell.*, vol. 4, pp. 1–20, Nov. 2021, doi: 10.3389/frai.2021.723447.

[28] H.-C. Chen, E. Prasetyo, S.-S. Tseng, K. T. Putra, Prayitno, S. S. Kusumawardani, and C.-E. Weng, "Week-wise student performance early prediction in virtual learning environment using a deep explainable artificial intelligence," *Appl. Sci.*, vol. 12, no. 4, p. 1885, Feb. 2022, doi: 10.3390/app12041885.

[29] R. N. Rajamoni, R. M. S. Kumar, and B. C. Leela, "Factors affecting the academic performance of students with hearing impairment," *Revue d'Intelligence Artificielle*, vol. 36, no. 4, pp. 569–574, Aug. 2022, doi: 10.18280/ria.360408.
[30] C. M. Convertino, M. Marschark, P. Sapere, T. Sarchet, and M. Zupan, "Predicting academic success among deaf college students," *J. Deaf Stud. Deaf Educ.*, vol. 14, no. 3, pp. 324–343, Jul. 2009, doi: 10.1093/deafed/enp005.
[31] A. Gramegna and P. Giudici, "SHAP and LIME: An evaluation of discriminative power in credit risk," *Frontiers Artif. Intell.*, vol. 4, pp. 1–6, Sep. 2021, doi: 10.3389/frai.2021.752558.
[32] S. Mane and D. Rao, "Explaining network intrusion detection system using explainable AI framework," 2021, *arXiv:2103.07110*.
[33] P. Chen, W. Dong, J. Wang, X. Lu, U. Kaymak, and Z. Huang, "Interpretable clinical prediction via attention-based neural network," *BMC Med. Informat. Decis. Making*, vol. 20, no. S3, pp. 1–9, Jul. 2020, doi: 10.1186/s12911-020-1110-7.
[34] H. Sahlaoui, E. A. A. Alaoui, A. Nayyar, S. Agoujil, and M. M. Jaber, "Predicting and interpreting student performance using ensemble models and Shapley additive explanations," *IEEE Access*, vol. 9, pp. 152688–152703, 2021, doi: 10.1109/ACCESS.2021.3124270.
[35] H. Khosravi, S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, and D. Gašević, "Explainable artificial intelligence in education," *Comput. Educ., Artif. Intell.*, vol. 3, 2022, Art. no. 100074, doi: 10.1016/j.caeai.2022.100074.
[36] S. D. A. Bujang, A. Selamat, R. Ibrahim, O. Krejcar, E. Herrera-Viedma, H. Fujita, and N. A. M. Ghani, "Multiclass prediction model for student grade prediction using machine learning," *IEEE Access*, vol. 9, pp. 95608–95621, 2021, doi: 10.1109/ACCESS.2021.3093563.
[37] S. Chen, "Interpretation of multi-label classification models using Shapley values," 2021, *arXiv:2104.10505*.
[38] N. B. Kumarakulasinghe, T. Blomberg, J. Liu, A. S. Leao, and P. Papapetrou, "Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models," in *Proc. IEEE 33rd Int. Symp. Computer-Based Med. Syst. (CBMS)*, Jul. 2020, pp. 7–12, doi: 10.1109/CBMS49503.2020.00009.

N. R. RAJI has been serving as an Assistant Professor in the Department of Computer Science at the National Institute of Speech and Hearing, India, since 1998. Additionally, she is a Research Scholar in the Department of Computer Applications at Noorul Islam Centre for Higher Education, India. Her primary research focus lies in machine learning applications within deaf education, and she actively participates in various research activities in this field. She is also a recipient of the National level Innovative Lecture Award during the 24th UGC-CEC Educational Video Competition, conferred by the Consortium for Educational Communication under the of India.

R. MATHUSOOTHANA S. KUMAR received the M.E. degree in computer science from the Arulmigu Kalsalingam College of Engineering, India, and the Ph.D. degree from Anna University. Currently, he is a Professor and the Head of the Department of Information Technology, Noorul Islam Centre for Higher Education, India. His research interests include image processing, data mining, and machine learning.

C. L. BIJI received M.Eng. degree in communication systems from the Mepco Schlenk Engineering College, Anna University, India, in 2006, and the Ph.D. degree in computational biology from the University of Kerala, in 2017. Currently, she is an Associate Professor with the Department of Analytics, School of Computer Science and Technology, Vellore Institute of Technology, India. She has modest number of publications in peer reviewed journals. Her current research interests include prompt engineering, deep learning, machine learning, and computational biology.