HFIPO-DPNN: A Framework for Predicting the Dropout of Physically Impaired Student from Education

Marina. B
Research Scholar, Alagappa University, Asst. Prof, St. Anne's Degree College for Women, Ulsoor, Bangalore
Email: marinab45698@gmail.com

A. Senthilrajan
Professor & Director, Dept of Computational Logistics, Alagappa University, Karaikudi, Tamil Nadu
Email: asenthilrajan123@yahoo.com

Abstract: Education plays a significant role in individuals' development and economic growth of the developing countries like India. Dropout of students from their studies is the major concern for any order of education. Some models for predicting the dropout of students are developed with several factors. Many of them lacked consistencies as they backed their studies with the academic performance of the students. Especially, for those students suffered with physical impairment the drop out depends on several external factors. Students drop out of school for a variety of reasons, including financial difficulties, parents' unwillingness, distance and a lack of basic amenities, poor educational quality, an inadequate school environment and building, overcrowded classrooms, improper languages of instruction, carelessness on the part of teachers, and security issues in girls' schools. Hence, this work proposes a novel HFIPO-DPNN to predicting the physically handicapped student's dropout from School also to predict the student dropout rooted on the previous semester marks. The proposed model enclosed the hybrid firefly and improved particle swarm algorithm to optimize the feature selection that influence the dropout of hearing-impaired students. The optimized feature data are used to predict the dropout with the novel DPNN. The optimized data was split and used for training the DPNN. The testing data is used to evaluate the performance of the proposed framework. The outcome for the proposed framework is evaluated on several metrics. The accuracy of the proposed model is about 99.02%. The HFIPO-DPNN framework can be enhanced for predicting the dropout for students with other disabilities. The optimization revealed that factors other than family factors should be taken into account when predicting dropout.

Index Terms: Education, Dropout, Physically Impaired, Feature selection, Accuracy

1. Introduction

Education has the pivotal role in uplifting the people and possesses cutting impact on all aspect of life as it forms an investment for human and economic development [1]. Especially for those peoples are physically impaired, the education is the only way that can uplift them to have a better life. The average dropout of normal students in their secondary is about 36.5%, 38.5% and 8.6% among the Hispanic, Afro American and Asian ethnicity respectively [2]. There are several factors that lead to the dropout of students from schools and colleges. The initial studies concentrated on the factors related to students. However, many researchers suggested that the family factors like poor economy, single parent is found to be the dominant risk factors along with poor student's performance [3]. When the student is being physically impaired, the risk factors multiple in larger magnitudes. The challenges involve learning with teaching response, accommodation and curriculum adaptation [4]. Especially according to 2011 census 45% of disabled populations in India are remained to be illiterate, related to 26% of Indians. Physically disabled students have higher dropout rates, with 59 percent of educated people with disabilities in regular Class X, compared to 67 percent of the population. [5]. Early precise prediction on students' performance with their current performance records is crucial for efficiently performing essential pedagogical interventions for ensuring students' satisfactory graduation within course period [6]. The predicting task gets complex by the ever-expanding data with the student enrolments and continuous shifting of student characteristics [7]. Additionally, this model may guide the students on a particular path to select appropriate options that can fulfil them in their life rooted on the past experience of students [8].

With the earlier prediction on drop out the organization or institutions can take the necessary steps to prevent it from occurring. Rooted on the earlier works, it is observed that the deep learning algorithms can provide better performance in predicting the drop out. Hence a novel HFIPO-DPNN framework is proposed with the hybrid firefly and improved particle swarm optimization (HFIPO) and Dropout Prediction neural network (DPNN).

In the proposed model, the data on the students' performance and their family background is taken for the drop out prediction. So that it will be helpful for the professors to give special training to students rooted on dropout status. In order to anticipate student dropouts based on last semester's grades, this paper suggests a unique HFIPO-DPNN. To maximize the feature selection that affects the dropout of hearing-impaired students, the suggested model included an enhanced particle swarm algorithm and a hybrid firefly. To guarantee that the particle has high exploration capabilities in the early iterations and good development capabilities in the later iterations IPSO is applied.

The foremost contributions of the projected work are listed below:

- The data set on drop out of physically impaired students are obtained from the institutions.
- The pre-processing is performed over the obtained data to get rid of its ambiguities.
- The optimization is carried out on the data for the selected features

Future study will involve expanding the performance prediction to optional courses and advising students on courses based on the forecast's outcomes. The other sections of paper are provided rooted on the proposed HFIPO are abridged as follows: In section 2, the works related to the dropout prediction study and the neural network model used are studied, section 3 explains the mathematical working of proposed HFIPO-DPNN. Section 4 provide facts of HFIPO-DPNN in terms of neural network model and mathematical expressions for feature selection, section 5 exhibits the proposed framework result rooted on the performance. The final section provides the conclusion and lists the future scope.

2. Related Works

2.1 "Some of the Related work related to Dropout of Physically Impaired Student from Education is reviewed in this section"

Ahmed et. al., [9] had suggested that practically all real-world applications, data mining was seemly more and more prevalent. Researchers find classification, one of the data mining techniques, to be seductive subject since it properly and effectively categorizes data for knowledge discovery. The results will assist both instructors and students in raise student achievement levels and it will also attempt to pinpoint the students who required extra help, as well as seek to lower the failure rate and take the necessary precautions for the examinations scheduled for the next semester.

Saa et. al., [10] had suggested the academic achievement of the pupils and investigates a number of variables that are theorized to have an impact on students' success in higher education and identifies a qualitative model that best categorizes and forecasts the students' performance based on relevant personal and societal variables.

Yukselturk et. al., [11] had proposed the use of data mining techniques in an online program and it was chosen from a pool of 189 students who registered for the online information technologies certificate program between 2007 and 2009. Online surveys were used to gather the data. The information was gathered by ten variables: gender, age, educational attainment, prior online experience, occupation, self-efficacy, preparedness, prior knowledge, locus of control, and dropout status as the class designation. In order to categories dropout students, four data mining techniques based on k-NN, Decision Tree, Naive Bayes, and Neural Network were used. With the use of 10-fold cross validation, these procedures were trained and evaluated.

Makhtar et. al., [12] had classified the performance of Maktab Rendah Sains MARA Kuala Berang students via data mining techniques based on how well they do in particular topics. The purpose was to explore the Naive Bayes algorithm, one of the classification methods used in data mining, in order to uncover any hidden relationships between the subjects that may possess impacted students' performance in Malaysian Sijil Pelajaran.

Okubo et. al., [13] had provided a technique by log data kept in educational systems to predict students' final grades by RNN. We tested the predictability of our strategy through the log data from 108 students. It was evident from the experimental findings, when compared to multiple regression analysis, that an RNN was useful for early grade prediction.

Akram et. al., [14] had presented a system for temporal analytics that examines students' problem-solving techniques and used for stealth assessment. The strategy-based temporal analytic framework groups students' problem-solving sequences across related tasks using long short-term memory network-based evidence models. We test our framework for strategy-based temporal analytics by a dataset of student interactions with a game-based learning environment for computational thinking in middle school.

Piech et. al., [15] had investigated the usefulness of modelling student learning with recurrent neural networks. The RNN family of models has significant benefits over earlier approaches in that they can capture more intricate representations of student knowledge and do not need the explicit encoding of human domain expertise. On a variety of knowledge tracing datasets, by neural networks resulted in significant improvements in prediction performance. The learnt model may also be applied to the building of intelligent programs and enables simple interpretation and structure discovery in student work.

Nghe et. al., [16] had offered a unique method for educational data mining that makes use of recommender system capabilities, notably for forecasting student success. Utilizing academic data for intelligent tutoring systems, we contrast recommender system methodologies with conventional regression approaches, such as logistic/linear regression, to verify the methodology. According to experimental findings, the suggested strategy can enhance prediction outcomes.

Portugal et. al., [17] had proposed the algorithms that are used by recommender systems to provide people advice on what goods or services to use. Recently, these systems had included artificial intelligence (AI) machine learning algorithms. Due to the large number of methods presented in the literature, it might be difficult to select a machine learning algorithm that was appropriate for a recommender system. There was a lack of knowledge regarding current techniques in algorithm utilization for researchers and practitioners creating recommender systems. Additionally, in order for software developers to know where to focus research efforts, a recommender system developed utilizing a machine learning algorithm frequently contains issues and unanswered concerns that must be assessed.

Aher et. al., [18] had compared various concomitances of data mining algorithms, including classification and association rule mining, clustering and association rule mining, association rule mining of classified and clustered data, pairing classification and association rule mining into association rule algorithms, and association rule algorithms alone. ADTree classification algorithm, Simple K-means clustering technique, and Apriori association rule algorithm are all taken into consideration and compared the results and discovered that the optimal combination are the cluster, classification, and association rule approach.

Naser et. al., [19] had predicted student performance in the Faculty of Engineering and Information Technology, an artificial neural network model was provided. The feed forward backpropagation technique was used to train the model. Student registration information were used to get the model's factors and demonstrated the artificial neural network's capability to predict student performance.

Zacharis et. al., [20] had proposed to evaluate artificial neural network performance in predicting student accomplishment by information gathered from online student behaviors in web-based blended learning courses. According to the analysis of the literature, neural networks are the most accurate classifiers in terms of prediction. Using a backpropagation technique, a multilayer perceptron neural network was trained to estimate students' chances of passing the course.

Wunsch et. al., [21] had proposed two contributions: first, it partially replicates previously reported findings in a new setting, and second, it investigates the effectiveness of neural networks in resolving some issue. Our results show that two characteristics are crucial, that machine learning methods are generally consistent across contexts, and that neural network-based approaches are just as efficient as the best Bayesian and decision tree techniques. Additionally, neural networks may be trained to be consistently pessimistic, which means they could be useful in conjunction with other methods for identifying children who need help.

Liu et. al., [22] had analyzed the practical issues with college students' innovation and entrepreneurship education based on the current state of college students' innovation and entrepreneurship as well as the current state of the innovation and entrepreneurship education system in colleges and universities. Next, examine practical, workable countermeasures through a thorough analysis of the causes using innovative and entrepreneurial education and find that there are still some issues with the current innovation and entrepreneurship education system based on the analysis of the status quo of college students' education in these areas. The majority of students lack a clear understanding of innovation and entrepreneurial spirit and lack entrepreneurial enthusiasm.

Alom et. al., [23] had developed the techniques to use data originating in an educational context is the focus of the fascinating interdisciplinary research field known as educational data mining (EDM). To analyse educational data and investigate educational issues, EDM employs computational methodologies. The introduction to EDM, descriptions of the various educational data environments, phases of EDM, applications and goals of EDM, and some of the most promising research areas are covered.

3. Preliminaries

3.1 Selecting the features

The selection of supervised features is mostly focused on the problem of labeling, and the significance or association between the function and the class category is used as its basic concept. Relevance evaluations may determine the significance of the features. This model aims to find an optimum function group for a training sample with characteristics and class labeling that provides the maximum accuracy of the classifier.

A general structure for selecting features is the Hilbert-Schmidt dependence criteria as seen in the equation, where $J(S)$ tests the dependence of a number of data on C. The principle of this paradigm is that $J(S)$ should be maximized by the key frame subset, which converts the choice of features into more of an optimization method.

$$D = \arg\max[J(S)]$$

The filter method typically uses assessment criteria to increase the correlation between the function and the class labeling and to decrease the correlation between features. In addition, the association between features is often superseded by redundancy.

Generally, rooted on the type of output, feature selection method is divided into two, weighted ranking method and subset selection model. Apart from this filter model (which considers the relation between features and output labels), wrapper model (takes the error rate or accuracy in the standard of evaluation) and embedded model (selecting features in the training model and generating output) is commonly used. Its performance is measured by machine learning model. Lasso (Least absolute shrinkage and selection operator) is commonly used to reduce the sum of squares of residuals if the absolute values of regression coefficient.

3.2 IPSO Algorithm

The PSO is the well-known optimization algorithm that is framed rooted on the natural swarm of flocking of birds. The IPSO [24] is constructed on the fundamental concept of positioning the particles and its velocities. The features along with its datasets are initially located in the search space. The data position vector is provided as in equation (1).

$$o_i = (o_{i1}, o_{i2}, o_{i3}, \dots \dots \dots \dots, o_{iS}) \tag{1}$$

In equation (1), S is the Search space dimensionality and i is the feature set. In the PSO, the particles are always under the motion and hence it possess the velocity V and its vector can be represented as in equation (5).

$$v_i = (v_{i1}, v_{i2}, v_{i3}, \dots \dots \dots \dots, v_{iS}) \tag{2}$$

The particle continuously updates its velocity and position rooted on the nearest swarm particles. Each particle records are previous position as the personal best and the population generates the best position to be global best that are termed as the pbesti and gbesti respectively. From the obtained best positions, the search for the optimum solution continues with updating of position and velocity with the following expressions in equation (3) and (4)

$$o_{is}(t + 1) = o_{is}(t) + v_{is}(t + 1) \tag{3}$$

$$v_{is}(t + 1) = wv_{is}(t) + C_1 R_1 (pbest_i(t) - o_{is}(t)) + C_2 R_2 (gbest_i(t) - o_{is}(t)) + C_3 R_3 (gbest_i(t) - pbest_i(t)) \tag{4}$$

Where,
'$C1$'and '$C2$'are acceleration coefficients,
'w' = inertial weight (constant),
'R' = a random number in range [0 1],
$pbest_i(t)$ = best position experienced by particle i until time t,
$gbest_i$'$(t)'$ = best position discovered by the swarm until time t 'i' = variable of a D-dimensional vector of position or velocity.

Since the inertial weight is highly responsible for balancing the exploration and exploitation rates of the algorithm, it is considered to have a high impact on the direction of particles in problem space. Exploration is defined as the property of swarm intelligence and the evolutionary computing method that is popularly preferred at the initial stages of a process to find new solutions in the problems space. The $pbest_i$ and $gbest_i$ will be updated as follows:

$$if\ f(xi(t + 1))\ \geq\ f(pbesti\ (t)),$$
$$pbest_i\ (t + 1)\ =\ pbesti\ (t)$$

otherwise,

$$pbest_i\ (t + 1)\ =\ xi\ (t + 1) \tag{5}$$

where, $f(xi)$ is the fitness value of particle i.

To calculate $gbest$,

$$gbest_i(\text{t} + 1)\ =\ \text{minf}(pbest_i\ (\text{t} + 1)) \tag{6}$$

It is generally the case for diversity of the particles at initial steps to be high and then slowly decrease over optimization processing time.

3.3 Firefly algorithm

Firefly is another nature inspired algorithm developed with its attractive characteristics that are represented as the illuminative function [25]. The Firefly method is a quick and simple to use algorithm. It can also be implemented in parallel. For multimodal situations, research indicates that it is sluggish to converge and quickly falls into a local optimum. The algorithm can be described through its three fundamental functions namely, attraction towards partners, attraction towards prey, and its mechanism of warning. Firstly, the light intensity at a specific distance 'r' from a light source obeys the inverse square law which means, as the distance 'r' from the light source increases, the brightness 'I'

tends to decline, i.e., the light source is inversely proportional to the squared distance. The second aspect influencing the visibility of fireflies is that light is absorbed by the air, which gets weaker as when the distance is increased. Initially the light intensities are formulated and rooted on it the attractiveness of the firefly is generated for the given data as in equation (7) and (8)

$$I = I_0 e^{-\gamma r_{ij}^2} \tag{7}$$

$$I = I_0 e^{-\gamma r_{ij}^2} \tag{8}$$

In equation (7) and (8), I_0, B_0 are the initial light intensity and attractiveness constant of firefly respectively, γ is the light absorption coefficient (=1). The distance between the two fireflies is r_{ij}, which is given as in the equation (9).

$$I = I_0 e^{-\gamma r_{ij}^2} \tag{9}$$

The beat solution obtained from the firefly algorithm is represented in the equation (10) as.

$$o_i = o_i + B_0 e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha(rand - 0.5) \tag{10}$$

Where, the second term of the equation is due to the attraction and the third term is randomization with 'α' being the randomization parameter. 'rand' indicates a random number picked from a uniform distribution in the range [0, 1]. Similarly, the expression (rand − 0.5) represents the range [−0.5, 0.5] to incorporate a positive and negative variation. β0 is always set to 1 and α ∈ [0, 1].

4. Proposed Methodology

The present prediction system is structured with the novel DPNN along with the hybrid improved Particle Swarm and Firefly optimization algorithm. In the proposed methodology, initially the dataset was pre-processed by removing inexactness and normalized through the Min-Max Normalization Technique, where min and max are minimum and maximum of each input features (i.e. 0 and 1 respectively).

Initially, using the collected data, the 3rd semester marks will be predicted. The obtained pre-processed data along with the predicted 3rd semester marks is provided to the novel HFIPO algorithm for optimizing the features. The optimized features are provided to the novel DPNN. The proposed DPNN model is trained with 80% of the dataset. Subsequently, the trained model is then tested with the complete dataset. The proposed HFIPO-DRP-NN framework for predicting the student dropout prior to the final semester exams is given in figure 1.
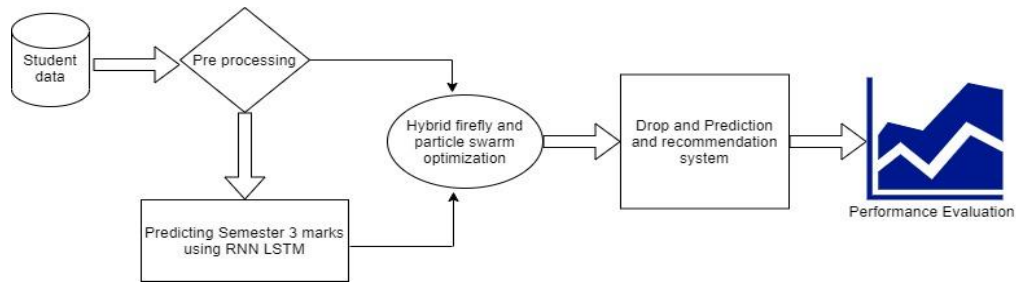

Fig. 1. Proposed HFIPO-DPNN.

For the HFIPO modelling, firstly, a set of particles in D-dimensional space within the range of [0,1] is generated, where 'D' denotes length of the original feature vector. If the value of the decision variable is higher than a predefined threshold value, then the corresponding feature element will be selected; otherwise, it will be removed from the original feature set. Further, the best subset of features is explored by the particle set in the search space. Every iteration involves the updating of velocity and position of each particle, the best experience of each particle and the best experience of swarm. In spite of finding the distance between firefly1 and firefly2, proposed work calculates the distance between firefly1 and $gbest_i$.

$$rij = ||xj - xi||, \tag{11}$$

Where, xj and xi are firefly1 and firefly2 respectively. The distance between any two fireflies xi and xj is expressed as the Euclidean distance by the basic Firefly algorithm.

$$rij = ||xj - gbest_i|| \tag{12}$$

$$r_{ij} = \sqrt{\Sigma_{k=1}^{d}(x_{i,k} - gbest_{i,k})^2} \tag{13}$$

Then the position is calculated as,

$$o_i = o_i + B_0 e^{-\gamma r_{ij}^2}\left(x_j - gbest_i\right) + \alpha(rand - 0.5) \tag{14}$$

In this model, embedded feature selection technique is used. This is the combination of filter and wrapped methods as discussed in section 3. This model reduces over fitting by regularization techniques. This technique works by penalizing the magnitude of feature coefficients and helps in minimizing the error rate over the iterations.

4.1 Data collection

The Physically impaired student data is collected exclusively from the ITI institution in and around Bangalore. The dataset consists up of several features like family background, academic performance during the final schooling and also the first two semester marks obtained by every individual. Additionally, the level of impairment is also specified as in Table 1.

Table 1. Feature from the collected data

S.no	Features
1	Sex
2	Family Size
3	Family Income
4	Age
5	Subject
6	Medium of Instruction
7	Kannada
8	Maths
9	Science
10	Social Science
11	Percentage
12	Sem I Percentage
13	Sem2 Percentage
14	Difficulty level in understanding
15	Disability Status

4.2 Preprocessing and Normalization

The collected data consists up some missing values that provide some ambiguity in utilizing the data for the present work. This method computes the missing value from the measured values from the data through the equation (15) and (16).

$$d_i = \sqrt{(P_{xi} - K_{xi})^2 + (P_{yi} - K_{yi})^2} \tag{15}$$

$$M_{vi} = \frac{\Sigma_{i=1}^{n}\frac{M_i}{d_i^2}}{\Sigma_{i=1}^{n}\frac{1}{d_i^2}} \tag{16}$$

In the equation (16), M_{vi} is the missing value, M_i is the measured value, d_i^2 is the k powered distance (K=2), n is the total data in each feature, P_{xi}, P_{yi} are the position of missing value in x and y axis. Similarly, K_{xi}, K_{yi} are the position of known value in x and y axis

The complete dataset obtained through the IDW is normalized through the min-max normalization technique. The mini-max is generally a linear transformation technique in which the data is transformed with pre-defined boundary [26]. The normalized data is obtained through the equation (17).

$$D' = \left(\frac{D - D_{min}}{D_{max} - D_{min}}\right) * (U - L) + L \tag{17}$$

In equation (11), D' and D denote normalized data and actual data, D_{min} and D_{max} are the minimum and maximum value in the dataset for each feature. U and L are the predefined upper and lower boundary values.

4.3 Recurrent Neural Network with BiLSTM layer

Recurrent neural systems have a directional loop that can retrieve and relate past data to the actual output, which is the key qualification in RNN. Moreover, the output is known as the current output of an arrangement, and the nodes between the hidden layers are never connected. In the present framework to predict the final semester marks Bi--LSTM-RNN network model is used, which includes input layer, output layer, 4 hidden layers (including BiLSTM).

This model is activated by sigmoid activation function and optimized using adam optimizer, which used the magnitude of the gradients and normalizes it. This facilitates the prediction of final semester marks. This helps in increasing the learning rate and stimulates rapid convergence. The mathematical expression for rmsprop optimizer is as follows:

$$b = b - \alpha \cdot \frac{db}{\sqrt{vdb} + \varepsilon} \tag{18}$$

where, db represents acceleration, vdb represents velocity, α and ε are hyper-parameters.

4.4 Optimization in feature selection

HFIPO algorithm.

In the present framework, a novel optimization algorithm is formulated with IPSO and Firefly algorithm to select the optimum feature from the student dataset. The normalized dataset is initially fed into the PSO and its outcome is processed through the firefly algorithm to obtain the optimized feature that can influence the dropout of student from their education. The primary objective of performing this optimized feature selection is to decrease the number of input features taken for DPNN, which can further result in the reduction of training and processing time and improvisation of its recommendation accuracy as well. Figure 2, 3 represents the Flowchart for preprocessing, normalization and feature selection with HFIP. In Figure 2, Implementation begins after the student data loading, if X_{ij} Null is Yes, estimate by (9) eq. and by eq. (10) calculate the missing value. Generate the complete data and set up upper and lower data boundary. Determine the values and normalize by eq. (11), If X_{ij} Null is NO set up upper and lower data boundary. Determine the values and normalize by eq. (11). Then, Stop the process. In Figure 3, Begin the process, then load the normalized data. Launch input parameter for HFIPO and initialize the velocity and position of the particle. With eqn. 3, 4 and 13 estimate the pbest, gbest and fitness value. If fitness value is best estimate if it is Yes estimating the position velocity by eqn. 14 and 15. If maximum fitness is obtained, normalize the data by 11 eqn. stop the process. If fitness value is best estimate if it is NO, repeat with eqn. 3, 4 and 13 estimate the pbest, gbest and fitness value.

Initially, the weight of the IPSO algorithm is estimated through the equation (19) as

$$w = wi - \left(\frac{wi - wf}{iterationmax}\right) \times iteration \tag{19}$$

The data are processed through the IPSO using the equation (1) and (2), the obtained $gbest$ value is compared for its fitness values over the pbest value of individual particles and the final fitness is established with equation (20).

$$f(i,t) = \begin{cases} true, if\ fitness(particlei\ t\ )gbest^{t-1} \\ false, if\ fitness(particlei\ t\ ) > gbest^{t-1} \end{cases} \tag{20}$$

The position of the particle and its velocity is estimated through the following expression in equation (21) and (22) with the saved temp variable (X_{i_temp})

$$O_i\ (t\ +\ 1) = O_i\ (t) + B_0 e^{-\gamma r_{ij}^2}\ (Xi\ (t) - gbest^{t-1}\ ) + a\ \in_i \tag{21}$$

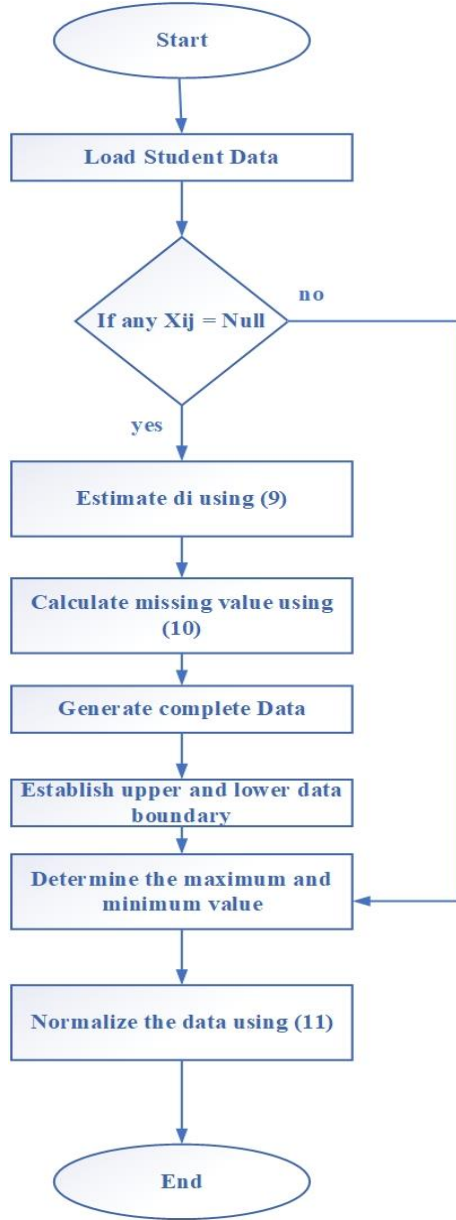$$V_i\ (t\ +\ 1)\ =\ O_i(t\ +\ 1) - X_{i\_temp} \tag{22}$$

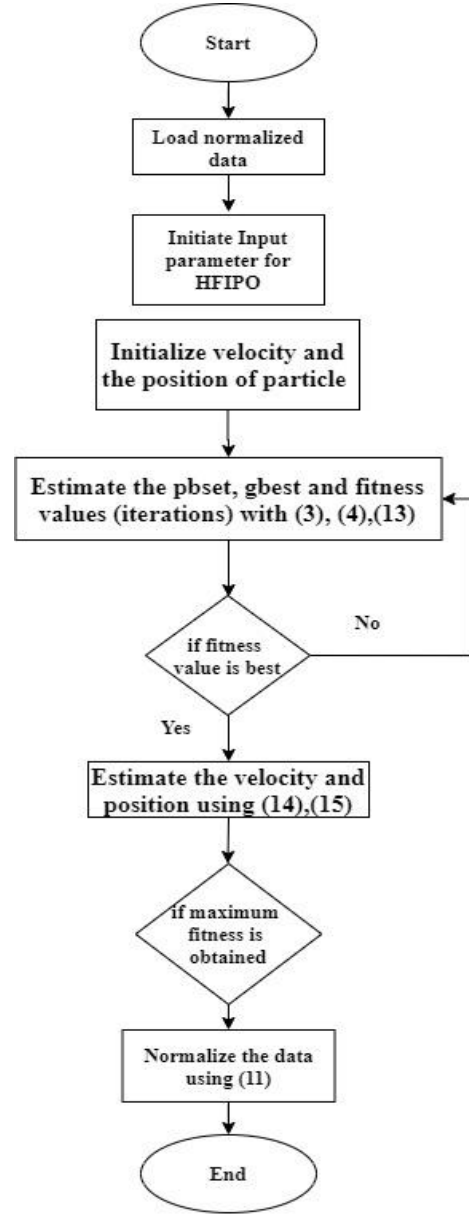Fig. 2. Flowchart for preprocessing, normalization and feature selection

Fig. 3. Flowchart representing implementation for preprocessing, normalization and feature selection

4.5 Prediction of dropout

The optimized data is segmented in the ratio of 80% and 20%. The 80% of data is fed into the DPNN for training the prediction of student dropout. In the proposed model DPNN has input layer (Il) one hidden layer and one output layer (Ol). Two different activation functions are used for the hidden layers as $F1$ and $F2$. Relu and Sigmoid activation function and adam optimizer is used to optimize the predict accuracy. The mathematical representation for each layers are given as,

Input layer:

$$I_l^t(t) = H_j^1(t-1) \tag{23}$$

Hidden layer 1:

$$H_j^1(t) = F_1\left(\sum_i^n V_{ij}^1 x_i(t) + \left(u_{lj}^1 I_1^t(t)\right)\right) \tag{24}$$

$$H_k^2(t) = F_2\left(\sum_i^h V_{jk}^1 H_j^1(t) + \left(u_{jk}^1 H_1^t(t)\right)\right) \tag{25}$$

Output layer:

$$O_k(t) = F_1 \sum_g^m W_{gk} \, d_g^2 (t) \tag{26}$$

The fitness functions are given by

$$F_1 = \frac{1}{1+e^{-i}} \tag{27}$$

$$F_2 = \frac{e^{2i}-1}{e^{2i}+1} \tag{28}$$

Where V, u, are the weight of the first and second hidden nodes and W is the weight between the hidden and output layer and i, j, k be the nodes at hidden layer1, hidden layer 2 and output layer. At the termination of training process, the novel DPNN Prediction model is constructed. The test data is provided to the generated model to evaluate the performance of the dropout prediction. Figure 4 shows the flowchart of DPNN.
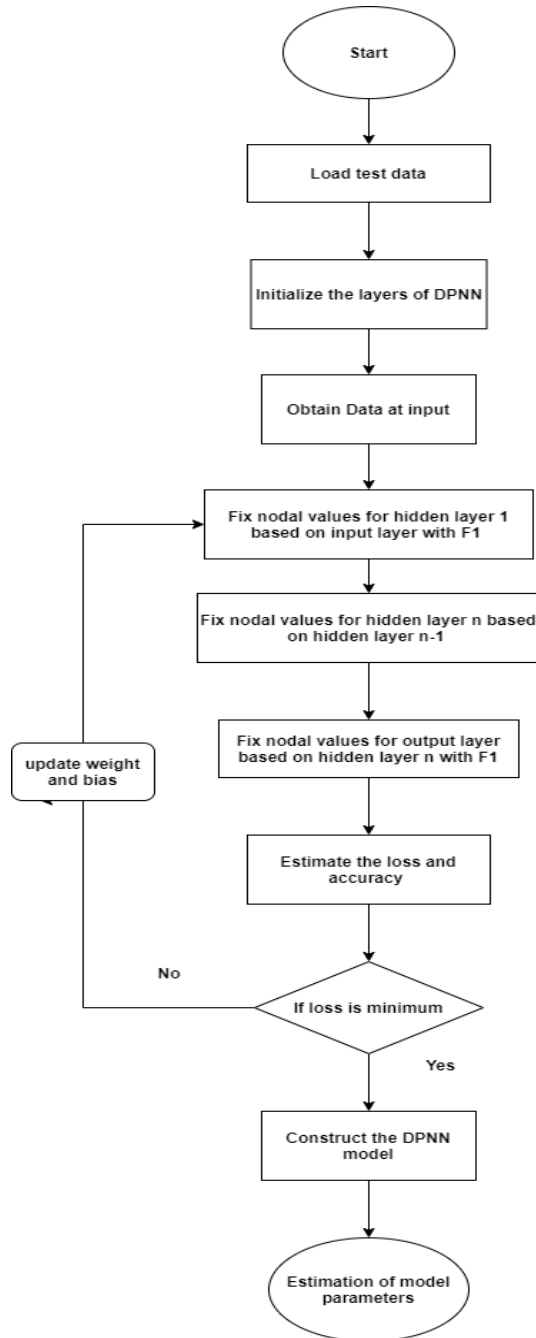

Fig. 4. Flowchart for DPNN

4.6 Evaluation

This section presents a comparative analysis of the performances of the proposed and existing methods. The classification accuracy is calculated by using:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

$$f1\ score = 2 * \frac{Sesitivity * Precision}{Sesitivity + Precision}$$

In HFIPO, in-order to produce an algorithm convergence in early generations, a random number is set to the inertia weight.

Where, TP – True Positive: the number of students that were predicted to be passed and actually passed the exams
TN – True Negative: the number of students that were predicted to be passed but gets dropped out
FP – False Positive: the number of students that were predicted to be dropped out and actually dropped out
FN – False Negative: the number of students that were predicted to be dropped out but pass the exam

5. Result and Discussion

The synthetic data is used for the current prediction model that contain 1,00,000 instances of student details over the feature presented in Table 1. The proposed model is implemented in python language. Using Bi-LSTM RNN model, the semester 3 marks are predicted with RMSE of 0.17. The pre-processed data along with predicted semester 3 marks is given to hybrid optimization model. Family size is 74 %, subject is 59% impact on dropout, Medium of Instruction is 72 % kannada is 63%, 10th percentage is 88%, sem 2 percentage is 92%, social science and Sem 3 percentage is 100%. On Implementing Hybridized Firefly and Swarm Optimization the total number of features get reduced from 15 to 8 and are given in Table 2.

Table 2. Feature selected through the optimization algorithm

Features	% Impact on dropout
Family Size	74%
Subject	59%
Medium of Instruction	72%
Kannada	63%
Social Science	100%
10th Percentage	88%
Sem 2 Percentage	92%
Sem 3 Percentage	100%

Table 3. shows the comparison of sensitivity and Specificity with the proposed.

Author	sensitivity	specificity
Proposed	99.76	97.38
[27]	99	0.907
[28]	84.43	----

Table 3 shows the comparison of sensitivity and Specificity with the proposed. The proposed HFIPO-DPNN framework was evaluated for its performance on various performance metrics as sensitivity, specificity, accuracy and F1 score that are computed through the confusion matrix. The obtained results are plotted according to their values in Figure 5.
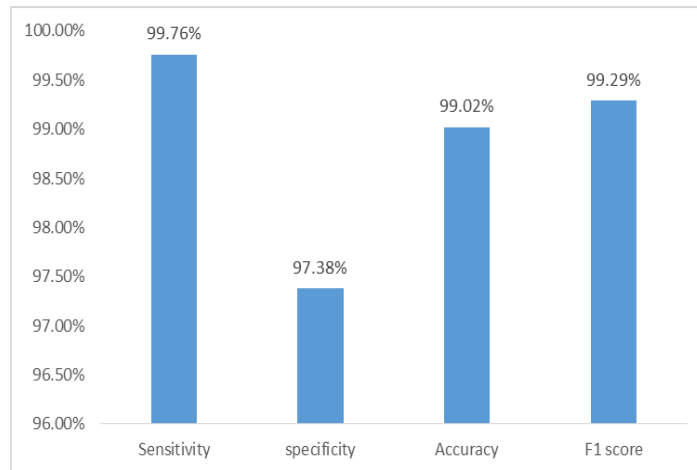
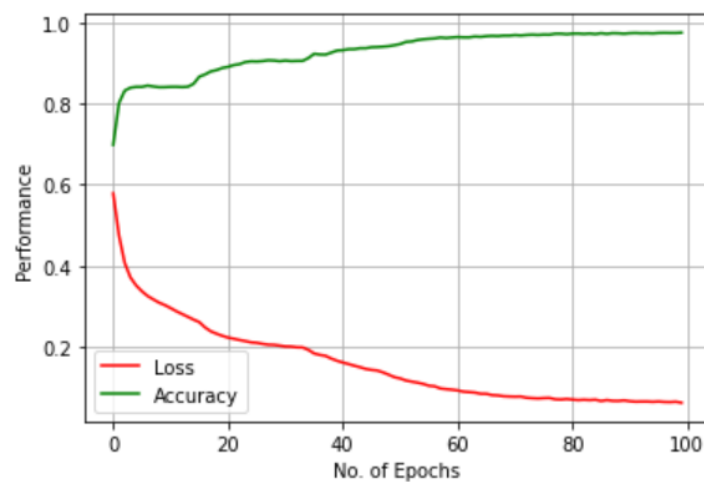Fig. 5. Performance of the proposed HFIPO-DPRNN

Fig. 6. Loss and Accuracy of the proposed HFIPO-DPRNN model

The accuracy of the proposed Prediction model is about 99.02%. The specificity and the sensitivity of the proposed framework is 97.38 % and 99.76% respectively. The F1 score for the proposed framework is about 99.29%. where, sensitivity is greater than other performance matrix. Figure 6 shows the graphical representation of Loss and Accuracy of the proposed HFIPO-DPRNN model. Number of Epochs and their performances is mentioned.

The proposed HFIPO-DPNN is compared with the rule and decision tree [29] based drop out prediction model. The proposed framework has the higher accuracy than the existing models in predicting the drop out of students. The comparison on performance is given in figure 7. Table 4 shows the comparison of student dropout finding and its causes.
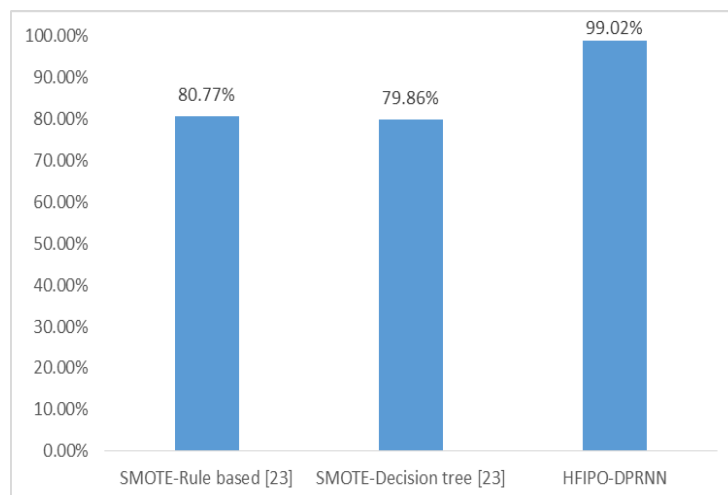

Fig. 7. Comparison on accuracy over existing model

Table 4. Comparison of author finding & the causes.

AUTHOR	FINDINGS	CAUSES
Sarker et. al., [30]	Factors For Academic Failure in Bangladesh	Poor Physical Condition, Biased Social Practice, Lack of Quality of Education, Economic Hardship, Geographic Isolation, Parental Education and Family Factor, Uncontrolled Population Growth and Unequal Access to Education, Early Marriage and Pregnancy of School Going Girl
Xenos et. al.,[31]	An analysis of the student dropout rates and factors contributing to them during the informatics course at the Hellenic Open University	incorrect time estimates for the students' jobs, which reduces the amount of time available for studying, lack of assistance from the tutor or less assistance than the student had initially anticipated., death of a family member, Health issues involving the student or another family member of the student, other factors besides those already mentioned that students classified as personal and refused to discuss with the interviewer.
Pierrakeas et. al., [32]	An analysis of the dropout rates and factors that contribute to them for two different distance learning courses	Professional reasons, Family and personal reasons, Health reasons, educational purposes
Alika et. al., [33]	Edo State: Implications for Counseling Related to School Drop-Out Among Girls	lack of resources poor performance in school, harassment from the other sex, unfriendly school culture, distance between home and school, early marriage or pregnancy, poor health, insufficient instruction, parents' passing

6. Conclusion

Early marriages, safety issues, and the inability to pay for college costs are the main reasons why students drop out of school. The primary factor identified by all studies is the poor financial and economic standing of families, which makes it difficult for them to pay for educational expenses, as well as the absence of essential school amenities like computer labs, science laboratories, libraries, electricity, clean drinking water, and furniture. The suggested HFIPO-DPNN model was used to forecast student dropout prior to tests in order to reduce dropout rate by providing customized instruction. Through the use of synthetic data, the study is specifically conducted for physically challenged students. The data set is first preprocessed to eliminate any data ambiguities, then normalized. The HFIPO algorithm is used to identify the optimum feature from the normalized data in order to forecast student dropout. The unique DPNN is fed with the optimized feature set data and has two distinct active functions for the four hidden layers. The neural network's output enables pre-cautionary measures to be modified to improve pupils' exam performance. The effectiveness of the suggested framework is assessed using a number of metrics. The proposed model had a 99.02% accuracy rate.

Future Research

The robust recommender model with higher accuracy and various other factors that may affect the performance and drop-out rate of physically challenged students from their education may be included in future work.

References

[1] Latif, A., Choudhary, A. I., & Hammayun, A. A.: "Economic effects of student dropouts: A comparative study". *Journal of global economics*. 3(2), 137. (2015).

[2] Stillwell, R.: "Public School Graduates and Dropouts from the Common Core of Data: School Year 2007-08. First Look". NCES 2010-341. *National Center for Education Statistics*, (2010).

[3] Swanson, C. B., & Schneider, B.: "Students on the move: Residential and educational mobility in America's school". *Sociology of Education*.54-67. (1999).

[4] Ahmad, Wasim.:"Higher Education for Persons with Disabilities in India: Challenges and Concerns". *Journal of Disability Management and Rehabilitation* 2, no. 1 ,pp : 1-4, (2017)

[5] Census 2011

[6] Xu, J., Moon, K. H., & Van Der Schaar, M.: "A machine learning approach for tracking and predicting student performance in degree programs". *IEEE Journal of Selected Topics in Signal Processing*, 11(5), 742-753. (2017).

[7] Sweeney, M., Rangwala, H., Lester, J., & Johri, A.: "Next-term student performance prediction: A recommender systems approach". arXiv preprint arXiv:1604.01840. (2016).

[8] Santos, O. C.& Boticario, J. G.: "Practical guidelines for designing and evaluating educationally oriented recommendations", *Computers & Education*, vol. 81, pp. 354-374. (2015).

[9] Ahmed, A.B.E.D. and Elaraby, I.S.: "Data Mining:A prediction for Student's Performance Using Classification Method". *World Journal of Computer Application and Technology*, 2(2), pp.43-47.(2014).

[10] Amjad Abu Saa.: "Educational Data Mining & Student's Performance Prediction". *International Journal of Advanced Computer Science and Applications*. Vol. 7, No. 5, (2016).

[11] Yukselturk E, Ozekes S, Türel YK.: "Predicting dropout student: an application of data mining methods in an online education program". *European Journal of Open, Distance and E-learning*. Jul 1;17(1):118–33 (2014).

[12] Makhtar, M., Nawang, H., and Shamsuddin, S. N. W.: "Analysis onstudents performance using Naïve Bayes classifier" *J. Theor. Appl.Inf. Technol.* 95(16):3993–3999, (2017).

[13] Okubo, F., Yamashita, T., Shimada, A., and Ogata, H.: "A neuralnetwork approach for students' performance prediction". *In:Proceedings of the Seventh International Learning Analytics &Knowledge Conference*, pp. 598-599, (2017).

[14] Bita Akram, Bradford Mott, Wookhee Min, Kristy Elizabeth Boyer, Eric Wiebe,and James Lester.: "Improving Stealth Assessment in Game-based Learningwith LSTM-based Analytics". *In Proceedings of the 11th International EducationalData Mining Conference*. Pp. 208–218, (2018).

[15] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami,Leonidas J. Guibas, and Jascha Sohl-Dickstein.: "Deep knowledge tracing". *neural information processing systems*, 505–513 (2015).

[16] Thai-Nghe N., Drumond L., Krohn-Grimberghe A and Schmidt-Thieme L.: "Recommender system for predicting student performance". *In Procedia Computer Science*, Volume 1,Issue 2, pp 2811-2819(2010).

[17] Portugal I, Alencar P, Cowan D.: "The use of Machine Learning Algorithms in Recommender Systems". *A Systematic Review*. arXiv:1511.05263v4 [cs.SE].

[18] Aher, S. B. and L.M.R.J., L.: "Best Combination of Machine Learning Algorithms for Course Recommendation System in", *International Journal of Computer Applications* (0975, 41(6), pp. 2–11. (2012). doi: 10.5120/5542-7598.

[19] Abu-Naser, S. S., Zaqout, I. S., Abu Ghosh, M., Atallah, R. R., & Alajrami, E.: "Predicting student performance using artificial neural network": *Faculty of Engineering and Information Technology*. (2015).

[20] Zacharis, N. Z.: "Predicting student academic performance in blended learning using Artificial Neural Networks". *International Journal of Artificial Intelligence and Applications*, 7(5), 17-29 (2016).

[21] Castro-Wunsch, K., Ahadi, A., & Petersen, A.: "Evaluating neural networks as a method for identifying students in need of assistance". *ACM SIGCSE technical symposium on computer science education* (pp. 111-116) (2017, March).

[22] Liu, A., 2016, June. "A Study on the Current Situation of Innovation and Entrepre neurship of Chinese College Students". *In 2017 2nd International Conference on Machinery, Electronics and Control Simulation* (MECS 2017) (pp. 9-12). Atlantis Press.

[23] Alom, B.M. and Courtney, M., 2018. "Educational data mining: a case study perspective from primary to university education in Australia". *International Journal of Information Technology and Computer Science*, 10(2), pp.1-9.

[24] Lv, X., Wang, Y., Deng, J., Zhang, G., & Zhang, L:. "Improved Particle Swarm Optimization Algorithm Rooted on Last-Eliminated Principle and Enhanced Information Sharing". *Computational intelligence and neuroscience*, (2018).

[25] Xu, H., Yu, S., Chen, J., & Zuo, X.: "An Improved Firefly Algorithm for Feature Selection in Classification". *Wireless Personal Communications*. doi:10.1007/s11277-018-5309-1 (2018).

[26] Patro, S., & Sahu, K. K.: "Normalization: A preprocessing stage". 2015 arXiv preprint arXiv:1503.06462.

[27] Sivakumar, S., Venkataraman, S. and Selvaraj, R., 2016. "Predictive modeling of student dropout indicators in educational data mining using improved decision tree". *Indian Journal of Science and Technology*, 9(4), pp.1-5.

[28] Kostopoulos, G., Kotsiantis, S. and Pintelas, P., 2015, October. "Estimating student dropout in distance higher education using semi-supervised techniques". *In Proceedings of the 19th Panhellenic Conference on Informatics* (pp. 38-43).

[29] Meedech, P., Iam-On, N., & Boongoen, T. "Prediction of Student Dropout Using Personal Profile and Data Mining Approach". *Intelligent and Evolutionary Systems*, 143–155. 2015. doi:10.1007/978-3-319-27000-5_12

[30] Sarker, M.N.I., Wu, M. and Hossin, M.A., 2019. "Economic effect of school dropout in Bangladesh". *International journal of information and education technology*, 9(2), pp.136-142.

[31] Xenos, M., Pierrakeas, C. and Pintelas, P., 2002. "A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University". *Computers & Education*, 39(4), pp.361-377.

[32] Pierrakeas, C., Xenos, M., Panagiotakopoulos, C. and Vergidis, D., 2004. "A comparative study of dropout rates and causes for two different distance education courses". *International Review of Research in Open and Distributed Learning*, 5(2), pp.1-15.

[33] Alika, I.H. and Egbochuko, E.O., 2009. "Drop out from school among girls in edo state: implications for counselling". *Edo Journal of Counselling*, 2(2), pp.135-141.

Authors' Profiles

Ms. Marina.B is currently working as a Asst. Prof in the department of Computer Applications, St. Anne's Degree College for Women, Ulsoor, Bangalore. She is pursuing her Ph. D at Dept of Computer Science, Alagappa University, Karaikudi, Tamil Nadu in part time. She has 12 years of teaching experience in the field of Computer Science. She has Completed M. sc., M. Phil in computer Science. Also, she has completed her UGC NET 2019 and TN SET 2016. Her research area includes machine learning and Deep Learning. She has Presented Papers in the International / National conferences and she received the Best Paper Award in the international conference.

Dr. A. Senthilrajan, Professor & Director, Dept of Computational Logistics, Alagappa University, Karaikudi, Tamil Nadu. He has completed his BE., MBA., M.sc (IT)., M.Phil., Ph. D. He has 19 years of Teaching Experience and 7 years of Research Experience. His area of research includes Image Processing, Networks, Artificial Intelligence. He is Guiding many Research Scholars at Alagappa university. He has been an Editorial board member in IEEE since 2006 to till date and also a Editorial Board member IAENG (International Association of Engineers) member from 2011 till date He is a Technical Team member in MOOC courses on SWAYAM, MHRD and also a Coordinator "Swayam Course" for UGC, New Delhi. Nodal officer of National Academic Depository (NAD). 4. He is a Nodal officer for conducting spoken tutorial, NMECIT, IIT, Mumbai. He has been a reviewer in many International Journals. He has published 23 papers in international Journals and 25 Papers in the conferences and Organized many International and National Conferences. His Achievements are Nominee for world who is who book for year 2008 and 2009.