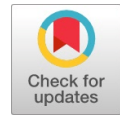


# Ensemble Swarm based Feature Selection (ESFS) and Ensemble Three Classifiers (ETCS) to Predict Student's Academic Performance

C. John Paul, R. Santhi



**Abstract---** Recently Educational Data Mining (EDM) has attracted many researchers in recent years. Many techniques of data mining are formulated to generate the techniques of the knowledge that is hidden within the educational data. The knowledge which is extracted aid the educational institutions to enhance the teaching process and learning methods. These improvements enhance the student performance and the performance of overall outputs. In EDM, Feature Selection (FS) plays a significant role in the improvement of quality of the models used for the purpose of prediction of educational datasets. Single feature selection algorithms do not render enhanced results of prediction. In this proposed work, Ensemble Swarm based Feature Selection (ESFS) and Ensemble Three Classifiers (ETCs) is formulated to classify the performance of students based on the selected features. This work concentrates on ESFS techniques are formulated to select the important and intrinsic features before the process of classification, ETCs are proposed. The samples are selected from the knowledge repository, which is initially pre-processed by means of Min Max Normalization (MMN) and Z Score Normalization (ZCN) method. Then the selected attributes from the technique called Ensemble Swarm based Feature Selection (ESFS) are combined to the learner's communication together with e-learning management system. ESFS algorithm fuses the Fuzzy Membership Genetic Algorithm (FMGA) and Improved Clonal Selection Algorithms (ICSAs). Also, Ensemble Three Classifiers (ETCs) is identified for the prediction of students' performance by combining the qualifiers like Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier and Decision Tree (DT). A widespread ensemble approach namely Bagging is utilized to combine all the results of three classifiers. The results that are obtained are found to have strong relationship among the learner's behaviors and their academic achievement.

**Index Terms---** Student Academic Performance, Educational Data Mining, E-learning, Ensemble, knowledge Discovery, Normalization, Ensemble Swarm based Feature Selection (ESFS), and Ensemble Three Classifiers (ETCs).

## 1. INTRODUCTION

One of the most important aspect is the enhancement of the education's quality. The information is stored in the repository of the educational institution which play a significant role to generate the interesting and hidden patterns to help all the stakeholder of an educational

institution [1]. Many techniques are existing to evaluate the academic performance of the students for building a good future of a student. The prediction process of a student is being a popular area in Educational data mining field. Data mining is one of the good choices for the researchers for the purpose of analyzing the performance of the students. In recent days, data mining technique is broadly used in the mining process of educational data [2,3]. This process is called as educational data mining. Educational Data Mining (EDM) projects the educational data for the better understanding of the issues identified in evaluation of student's performance utilizing the various techniques of data mining [4]. EDM process the educational information in order to aid the educational institutions to plan on the various strategies on education so as to enhance the quality of education. The major area in EDM is Prediction. The important factor for the academic growth of a student is the prediction and analysis of students' performance. The progress of the student academics is captured with the help of prediction models. These models which are used for the prediction uses variety of EDM techniques to analyze the academic performance of a student. It is difficult to differentiate the features that affect the academic performance of a student [4]. The prediction of this evaluation is helpful for many educational institutions to select the students those who are in dire need of financial assistance [4], [5], enhancing the institution enrolment quality [6], aid students in their better planning of their future, also to meet the struggle in their studies. The model to predict the students' performance depends on the selected features from the dataset. The common features that can be chosen by implementing the feature selection algorithm [7]. These algorithms are used to refine the results of prediction [8]. Feature Selection methods can be categorized into 3 groups which are filter, wrapper, and embedded models. Filtering method is based on the characteristics of training data. This step is carried out in the preprocessing phase and is independent on the learning algorithm. Wrapper method utilizes the learning algorithms to validate the features. The specific methods are Embedded technique which are applied on few learning algorithms and these methods are carried out on the classifiers during the training process. Nowadays, the researchers have concentrated on the techniques used for the mining the educational data and becoming popular research field. More works have shown variety of approaches that aid in the rendering of a clear decision for enhancing the performance of the students.

Manuscript published on 30 August 2019.

\* Correspondence Author (s)

C. John Paul, Research Scholar, Alpha Arts and Science College, Porur, Chennai, Tamilnadu. E-mail: [jpcalling@gmail.com](mailto:jpcalling@gmail.com)

Dr.R. Santhi, Research Supervisor, Alpha Arts and science College, Porur, Chennai, Tamilnadu. E-mail: [santhi.educator@gmail.com](mailto:santhi.educator@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## Ensemble Swarm Based Feature Selection (Esfs) And Ensemble Three Classifiers (Etc) To Predict Student's Academic Performance

These strategies focus on constructing the classifier models that identifies the potential of unknown students utilizing some features. Though these techniques produce some effect in the enhancement of the prediction of students' performance, they fail to provide unified strategy that determine the performance of the students accurately. Additionally, a single feature selection method may produce local optimal or sub-optimal feature subset for which a learning method consists of its performance. In ensemble-based feature selection method, numerous feature subsets are amalgamated to pick an optimal subset of features utilizing blend of feature ranking that progresses accuracy of classification. In the initial step of ensemble method, a group of various feature selectors are identified and every selector offers an organized order of features. The next step aggregates the chosen subsets of features using various aggregation techniques [9]. In recent studies, researchers use ensemble-based feature selection techniques and the combination of classifiers to generate efficient prediction models [10].

In this work, explore whether the utilization of ensemble feature selection techniques can be used to produce further strong feature selection techniques, and whether combining multiple methods has any effect on the classification performance. For the better understanding, Ensemble Swarm based Feature Selection (ESFS) is employed for features/attributes selection. Then, use three of the most common data mining methods in this area to construct the academic performance model: Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier and Decision Tree (DT). Then, implemented an ensemble bagging model to enhance the performance of such classifiers. To conclude with, the results are measured using the classification metrics like precision, recollect, F-measure and accuracy.

## 2. LITERATURE REVIEW

Doshi [8] used feature selection attribute algorithms Chi-square, InfoGain, and GainRatio to forsee the important highlights. Then have applied quick correlation base filter on given features. Future classification is done using Naïve Bayesian (NBs) Tree, Multi-Layer Perceptron (MLP), and Instance based -K- Nearest Neighbor (KNN). Results indicated reduction in computational cost and time and upsurge in predictive accuracy for the student model

Amrieh et al [10] proposed another new student's performance prediction model based on data mining techniques with new data attributes/features, which are called student's behavioral features. These types of features are related to the learner's interactivity with the e-learning management system.

The performance of student's predictive model is evaluated by set of classifiers, namely; Artificial Neural Network (ANN), NBs and Decision Tree (DT). In addition, ensemble methods such as Bagging, Boosting and Random Forest (RF) are also applied to improve the performance of these classifiers. The got outcomes results reveal that there is a strong relationship between learner's behaviors and their

academic achievement. This outcomes proves the unwavering quality of the proposed model.

Mueen et al [11] have gathered students' data from two of the undergraduate courses. Three different data mining classification algorithms such as NBs, NN, and DT were utilized on the dataset. The forecasting performance of three classifiers are estimated and analyzed. It was seen that NBs classifier outflanks other two classifiers by achieving overall prediction accuracy of 86%. This investigation will help teachers to improve student scholarly performance.

Zaffar et al [12] exhibited an examination of the presentation of filter feature selection algorithms and classification algorithms on two different student datasets. The outcomes we get from different FS algorithms and classifiers on two student datasets with different number of features will assist researchers to catch the best combinations of filter feature selection algorithms and classifiers. It is very essential to put light on the relevancy of feature selection for student performance prediction, as the constructive educational strategies can be derived through the relevant set of features. The outcomes of our study depict that there is a 10% difference of prediction accuracies between the results of datasets with different number of features.

Zaffer et al [13] exhibited an examination of the performance of feature selection algorithms on student data set.

The outcomes of the various FS algorithms and classifiers also help the new scholars in searching the finest combinations of FS algorithms and classifiers. By selecting related features for student prediction model is subtler problem for educational stakeholders, as they have to take conclusions on the basis of outcomes of prediction models. Additionally, our paper is an attempt of playing a positive role in the upgrading of education quality, as well as guides new researchers in making academic involvement.

Anuradha and Velmurugan [14] examined the most applicable subset features for accomplishing high performance accuracy by assuming Correlation based feature Subset Attribute evaluation and Gain-Ratio Attribute evaluation feature selection techniques.

For classification, the NBs classifier is actualized by using WEKA tool.

The result demonstrates the effectiveness in the predictive accuracy with least number of attributes. Also the results reveals that the selected data features have observed to be influenced the classification process of the student performance model.

Zahedifard et al [15] displayed Genetic Algorithm (GA) based feature selection in predicting students' performance. The referenced data of this article are taken from 386 students of high schools in Bushehr province.

Individual, environmental and educational factors affecting successful and unsuccessful students have been examined and bestowing to them efficient models based on c4.5 tree algorithm, Support Vector Machine (SVM) methods and logistic regression.

The outcomes can help directors of educational systems to a correct educational planning an optimizing of educational processes in high schools.

Khasanah [16] led a Feature Selection to choose high impact attributes with student performance in Department of Industrial Engineering Universities Islam Indonesia. Then, two popular classification algorithms, Bayesian Network and DT, were executed and compared to know the best prediction result. The result showed that student's attendance and GPA in the first semester were in the top rank from each and every Feature Selection methods, and Bayesian Network is outperforming DT since it has higher accuracy rate.

Ajibade et al [17] proposed one more new performance prediction model for students which be contingent on data mining methods which integrate new features known as behavioral features of students.

The proposed predictive model is assessed using classifiers like NB, DT, KNN, Discriminant Analysis (Disc) and Pairwise Coupling (PWC). Furthermore, so as to improve the classifiers performance, the ensemble methods such as AdaBoost, Bag and RUSBoost were applied to enrich the accurateness of the performance model of the students. The outcomes demonstrate that there exist a strong relationship between conduct of students and their academic performance.

Soni et al [18] displayed classification Algorithms like DT, NBs and SVM can help us for anticipating student's performance. This forecast helps parents and teachers to keep track of student's performance and provide required counseling.

This Analysis additionally progress in giving scholarship and other required training to the student. It can fetch the benefits & guidance of beginner, teachers and educational institutions. Experimental results show that recommended method significantly outpaces prevailing procedure due to the misuse of family incomes and students' personal data component sets. Results of this inspection can act as policy enhancement technique in higher education

Cheng et al [19] proposed a Synthetic Feature Selection Approach (SFSA), which is incorporated with a SVM to focus patterns and discover the key features that influence students' academic achievement.

For checking the proposed model, two databases, namely, "Student Profile" and "Tutorship Record", were together from an elementary school in Taiwan, and were concatenated into an combined dataset based on students' names as a investigation dataset.

The outcomes specify that the proposed model can enhance the accuracy and facilitate the interpretation of the pattern from a hybrid-type dataset of students' academic accomplishment.

Almasri et al [20] presented in three folds that incorporates the following:

(i) Giving a thorough analysis about the selected features and their effects on the performance value by using statistical examination techniques,

(ii) Building and studying the performance of several classifiers from various families of Machine Learning (ML) techniques,

(iii) Proposing an Ensemble Meta-based Tree model (EMT) classifier technique for predicting the student performance.

Asif et al [21] introduced data mining methods to study the performance of undergraduate students. To start with, predicting students' academic achievement at the end of a four-year study programme. Second, studying typical progressions and combining them with forecasted results. The outcomes demonstrate that by focusing on a small number of courses that are indicators of particularly good or poor performance, it is possible to give auspicious cautioning and care to low achieving students, and provide guidance and opportunities to high performing students.

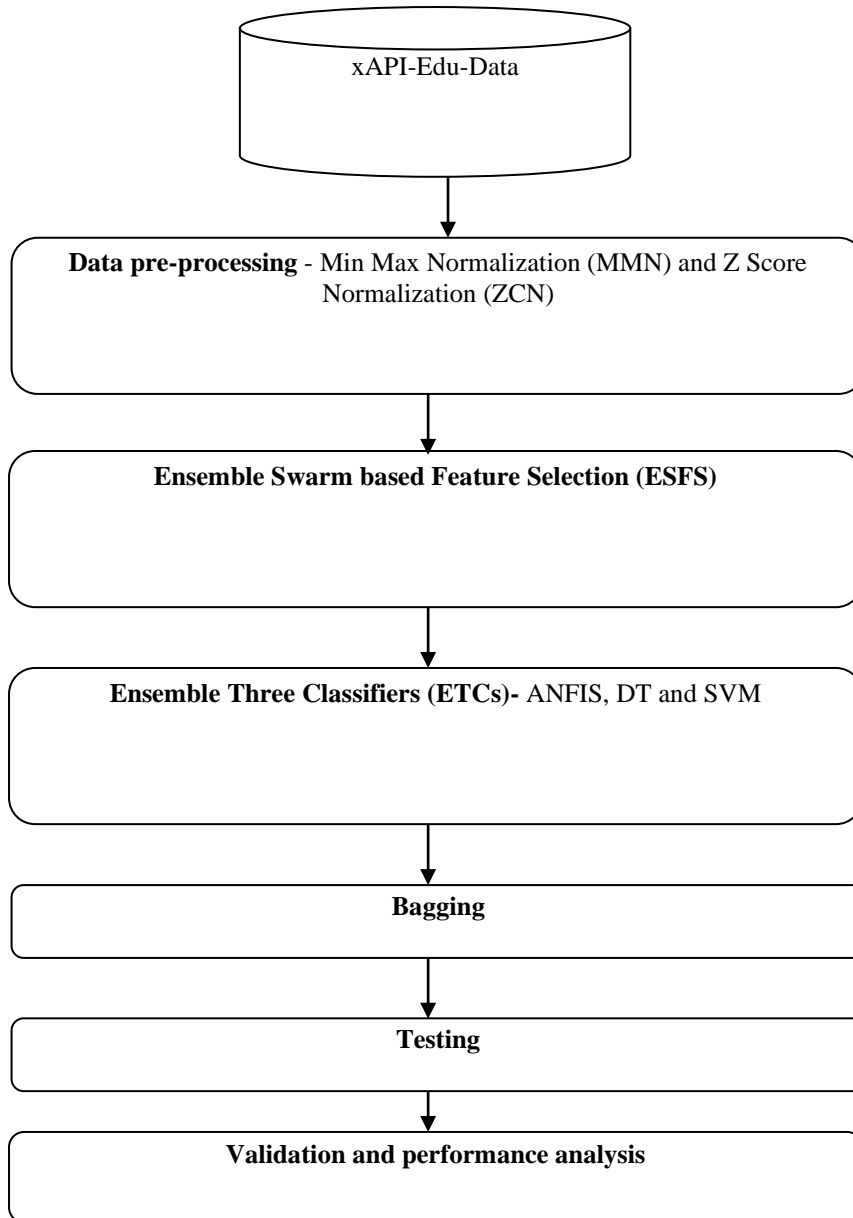
### 3. PROPOSED METHODOLOGY

In this work, Ensemble Swarm based Feature Selection (ESFS) and Ensemble Three Classifiers (ETCs) is proposed to characterize student's performance relies upon with nominated features. Initially the samples are selected from the knowledge repository which is pre-processed through Min Max Normalization (MMN) and Z Score Normalization (ZCN) method.

At that point the selected attribute from Ensemble Swarm based Feature Selection (ESFS) are connected to the learner's communication along with the e-learning management scheme.

ESFS algorithm is connected with the procedure of the Fuzzy Membership Genetic Algorithm (FMGA) and Improved Clonal Selection Algorithms (ICSAs). Furthermore, Ensemble Three Classifiers (ETCs) is proposed to predict student's performance by combining classifiers, like Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier and Decision Tree (DT).

# Ensemble Swarm Based Feature Selection (Esfs) And Ensemble Three Classifiers (Ets) To Predict Student's Academic Performance



**Figure 1: Proposed Framework**

### 3.1. Dataset

It is an educational database which is collected from Learning Management System (LMS) named as Kalboard 360 [21-22]. The dataset comprises 305 males and 175 females. The students derive from diverse origins like 179 students are from Kuwait, 172 students are from Jordan, 28 students from Palestine, 22 students are from Iraq, 17 students from Lebanon, 12 students from Tunis, 11 students from Saudi Arabia, 9 students from Egypt, 7 students from Syria, 6 students from USA, Iran and Libya, 4 students from Morocco and one student from Venezuela.

### 3.2. Data pre-processing

After the data gathering task, it concerns some pre-processing method to progress the quality of the data set. Data pre-processing is expected an essential stage in the knowledge discovery process, which comprises information cleaning, feature selection, data reduction and data transformation. Data preprocessing is the accomplishment before relating data mining method, it changes the original

data into a appropriate shape to be utilized via a specific mining algorithm.

#### 3.2.1. Min-Max Normalization

Min–Max normalization is an easy method where the system can particularly fit the information in a pre-defined boundary along with a pre-defined boundary [23-24]. As per Min-Max normalization method

$$N = \left( \frac{A - \min_A}{\max_A - \min_A} \right) * (D - C) + C \quad (1)$$

Where, N includes Min-Max Normalized information one, if pre-defined boundary is [C, D], if A is the range of original data & B is the mapped one data.

### 3.2.2. Z Score Normalization

Z-score Normalization method is provided the normalized values or range of data from the original unstructured data through the concepts such as mean and standard deviation [24]. So the unstructured data can be normalized via z-score parameter, as per specified formula

$$v'_i = \left( \frac{v_i - \bar{E}}{\text{std}(E)} \right) \quad (2)$$

### 3.3. Data Cleaning

Data cleaning is one of the foremost preprocessing tasks, is useful on this data set to eliminate unrelated items and absent values. The data set includes 20 missing values in different attribute from 500 records, the records along with absent values are removed from the given dataset, and the data set after cleaning becomes 480 records.

### 3.4. Ensemble Swarm Based Feature Selection (ESFS)

Ensemble Swarm based Feature Selection (ESFS) is a group of single Feature Selection methods are qualified, and the output of the ensemble is gained by aggregating the outputs of the single models.

The first step includes generating a set of dissimilar feature selectors, each giving their output, while the next step combines the results of the single models. Aggregating the different feature selection results can be done by weighted voting, e.g. in the case of deriving a consensus feature ranking, or by counting the most frequently selected features in the case of deriving a consensus feature subset. In this work, focus on ESFS that work by aggregating the feature rankings provided by the single feature selectors into a final consensus ranking [25].

### Fuzzy Membership Genetic Algorithm (FMGA)

Genetic Algorithms (GA) is a group of transformative algorithms that use growth as a source of motivation to determine the solution for grouping issues. The chromosomes have been measured as number of samples used for student's performance prediction and each dimension of the student performance prediction features are to be considered to be a gene.

Every generation has a specific number of chromosomes also called as the population. The most important procedure of GA is the fitness function; here the fitness function is calculated based on the classification accuracy of each features. Each chromosome from the generation is accepted through the fitness function and thus, they get their fitness value. The fitness value then determines the proximity of the prediction performance from chromosome to the highest fitness value.

The student performance prediction features from chromosomes with improved fitness value is used for reproduction. The means of reproduction are mostly depending on crossover and mutation. Crossover is the interchange of two features between the classifier model and

mutation is the random change in the two features. Mutation is usually done on comparatively weak sensor nodes from graph model, so that it adds diversity to the two features (population) without actually impeding the progress towards the higher fitness value.

The chromosomes that have reproduced are replaced by the new chosen features, irrespective of the fitness values of the new features in the student's performance prediction model. The outcomes in the formation of the new feature generation. The features in this generation, which were the offspring of the previous generation, are now the features of the next generation.

These features (chromosomes) are now passed through the fitness function again and the strongest features are selected to reproduce, which results in a new feature generation, with a new set of chromosomes and ideally nearer to the optimal solution with optimal selected features and fitness value. The following are the most important GA operators:

The Selection operator will chooses features in the population for reproduction. The selection function is commonly stochastic and intended to select features the improved fitness value of the chromosomes from the graph model.

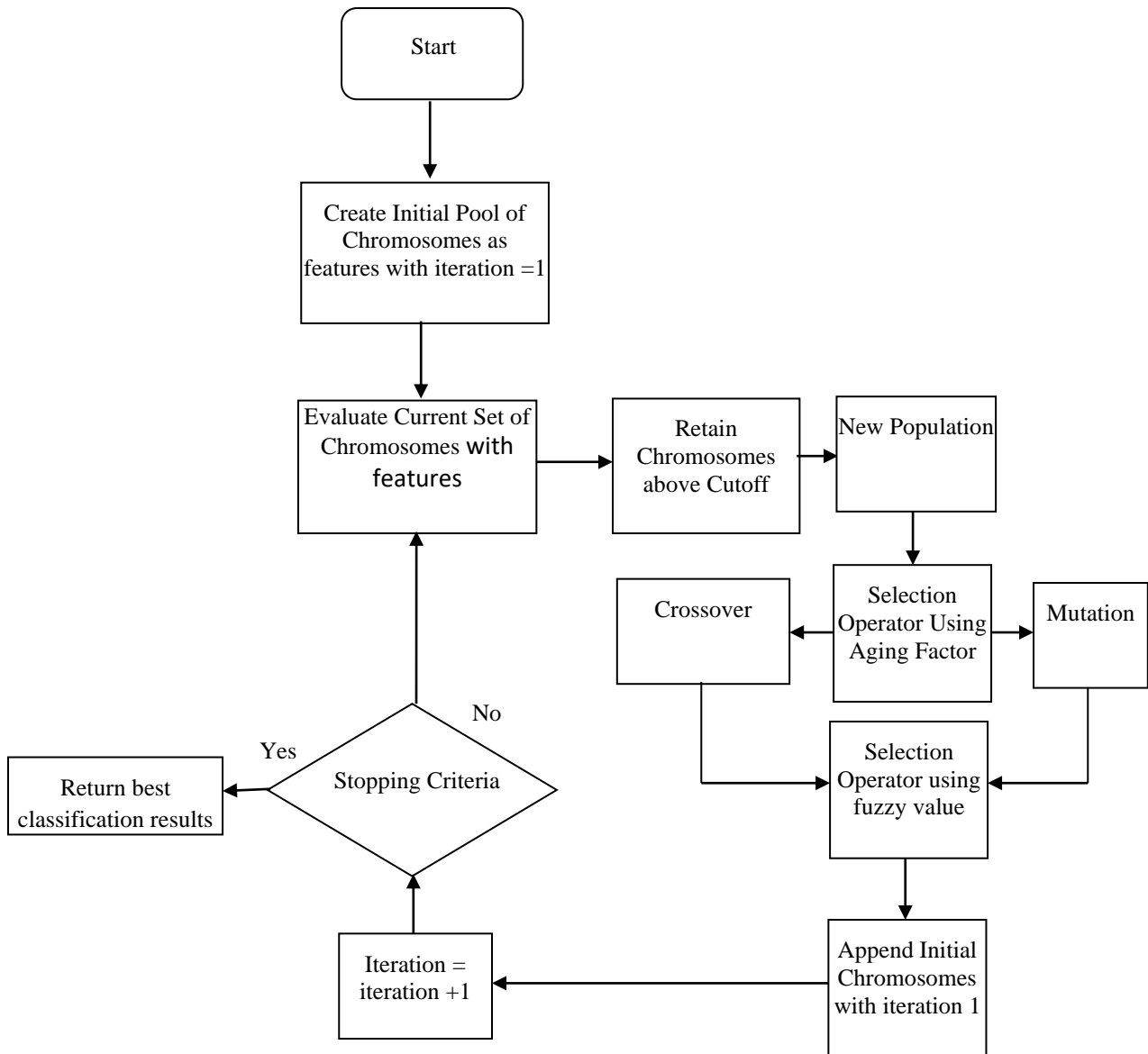
The Crossover operator chooses features and exchanges the features in the chromosomes before and after the features to form new selected features of prediction model offspring.

The Mutation operator arbitrarily flips the nominated features thereby creating a new features offspring. Mutation adds diversity to the prediction database (population).

While substituting chromosomes from prediction model in the iteration N to iteration N+1, selected features with a fairly worthy classification value may be exchanged by selected features with a poor classification accuracy value. Hence, this may consequence in the choice of classification. In exclusiveness, the top features of each generation are classified as elite individuals. They will contribute in the reproduction, but will not be replaced by any features from the next generation.

This is called Simple Elitism. In Global Elitism, each feature from generation N+1 can replace its parent from generation N, if and only if, its performance is found to be superior. The weakness in this case is that, the comparison is still being done only on features to predict performance model on the basis and not on a generation to generation basis (see figure 2).

## Ensemble Swarm Based Feature Selection (Esfs) And Ensemble Three Classifiers (Ets) To Predict Student's Academic Performance



**Figure 2: Flow chart of a FMGA**

**Trapezoidal Fuzzy membership function:** Then Trapezoidal Fuzzy membership function is introduced to automatic representation of classification value into equal ranges [0-1]. The trapezoidal curve is a function of a classification value,  $Y$ , and depends on four scalar parameters  $a$ ,  $b$ ,  $c$ , and  $d$ , as given by

$$f(y, a, b, c, d) = \begin{cases} 0, & z \leq a \\ \frac{z-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq z \end{cases} \quad (1)$$

**Dynamic Population Size:** The fundamental issue with the traditional GA is the static size of population. So the computational complexity drastically increases if out of  $k$  chromosomes, the fitness values of  $k/2$  chromosomes are below par. GAs would deliberate these chromosomes for reproduction using crossover and mutation, thus increasing the time complexity. In the modified GA, a cut-off on the classification value has been measured and every features that has a fitness value less than this cut-off classification value is discarded. If at any point after the cutoff, the number of features from the prediction model is more than

the initial population size, the size is reset to initial population size with the less fit features being disposed of. Thus, in this way the number of features at any point will never be greater than the size of the initial population from classification model, thus ensuring computational efficiency.

**Dynamic Elitism:** The worldwide elitism is either done on highlights to classification basis is considered as elite individuals. FMGA is being used with the number of elite features is active, i.e. it is changing from generation to generation. The upside of this FMGA method is that, the life of the sensor nodes is directly proportionality with the wellness.

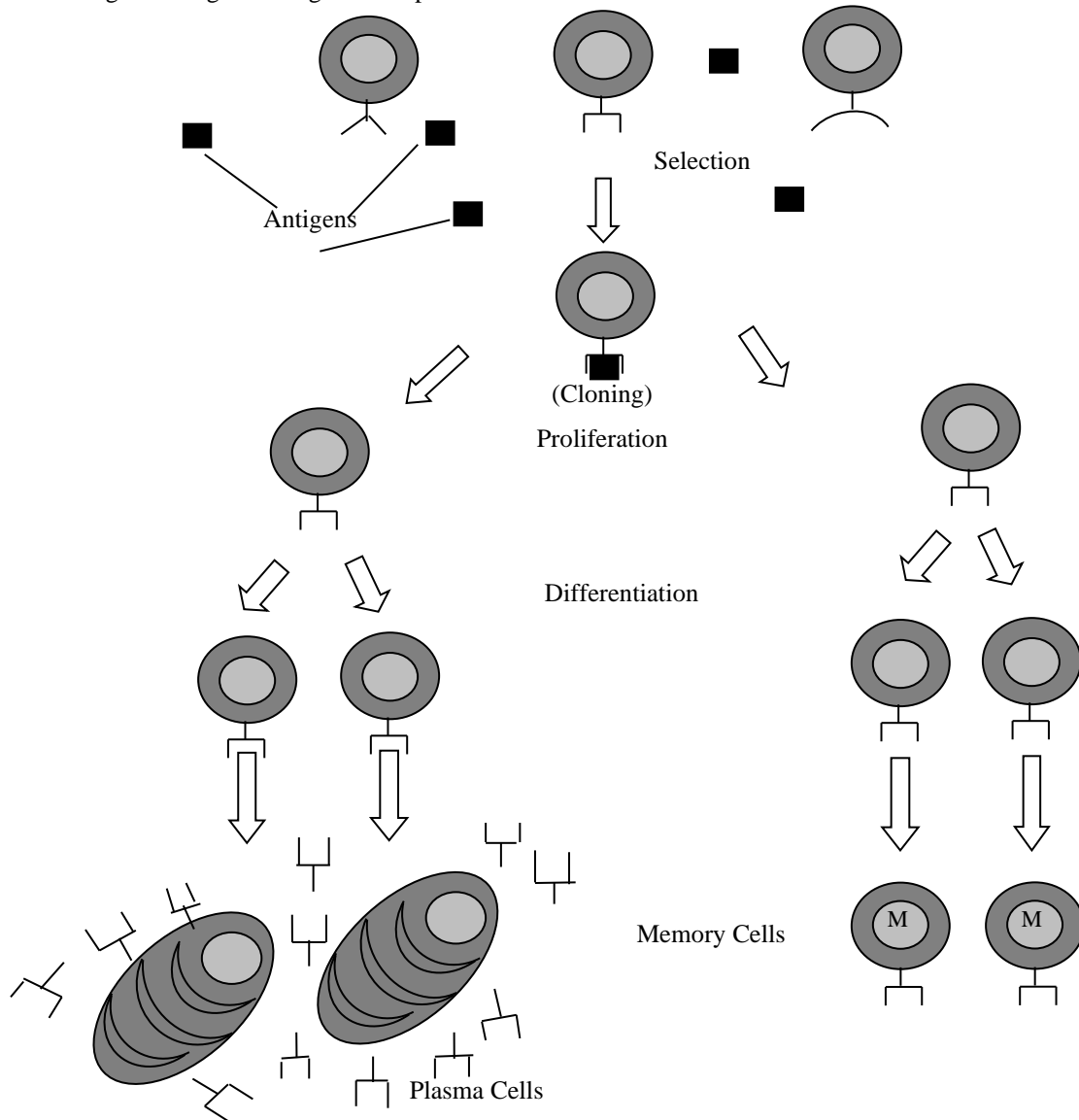
**Aging factor:** A yet another new parameter called the age of the features has been introduced. The original principle behind the inclusion of this parameter is that, the features that are fit to live on for a more number of generations have already reproduced in the earlier generations. Thus, allowing these features to reproduce again will decline the diversity of the population to graph model hence causing premature convergence.

Thus, the fitness values of the features that are measured for the sake of reproduction are indirectly proportional to the age of the highlight.

*Improved Clonal Selection Algorithms (ICSAs)*

The main aim of CSA theory is in the event where B cell acts to invaded antigen through altering the receptor called

antibody is illustrated in figure 3. Generally CLOGNALG [26], is one of the description for CSAs. Three main operations are cloning, hyper mutation, and choice is used for choosing most the features in student performance prediction dataset.



**Figure 3: Clonal selection principle**

Those physiognomies bring out shortages once facing the high-dimensional, nonconvex multimodal and multi-objective optimization problems. At first, the cloning operator occupies more calculation time. Secondly, hypermutation is insufficient to accept the burden of balancing the global search and the local search, which lead to the premature convergence and ungratified accuracy. Thirdly, the inadequate deliberation on communication between the samples in the population could lead to the hunt missing global awareness, which is, continuously searching many areas of the search space leaving the others unvisited. These above mentioned problems are solved by using the following functions.

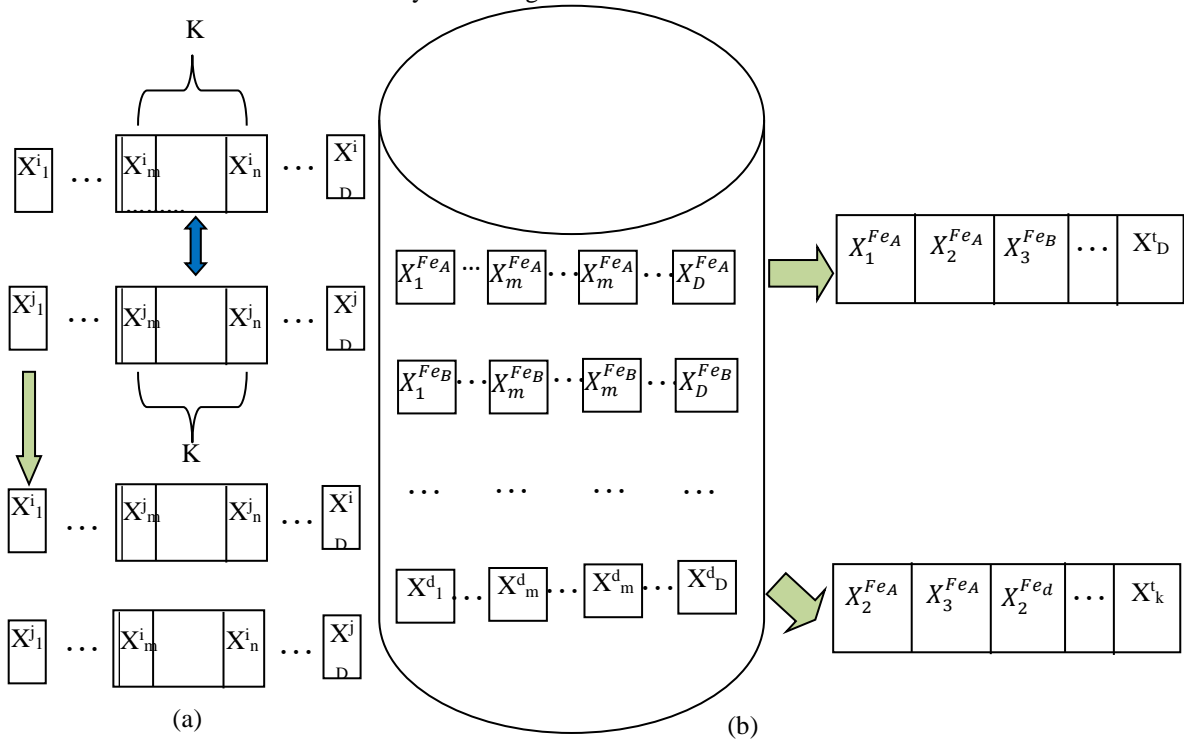
In immunology, these grouping and somatic mutation are up-to-date which takes the control for the feature selection of antibody genes. The recombination of the

immunoglobulin gene segments is the first step when the cells are first projected to antigen. In the view of optimization, recombination functions as the coarse-grained exploration while hypermutation functions in the alike way as fine-grained exploitation. To overcome the prediction error of the ensemble model in student performance prediction and the decrease the time complexity of the coding, Improved Immune Clonal Selection (ICSA) algorithm is coded in real number and every attributes of prediction samples are deliberated as a gene segment. The complete student performance dataset samples forms the gene fragments library.

## Ensemble Swarm Based Feature Selection (Esfs) And Ensemble Three Classifiers (Ets) To Predict Student's Academic Performance

According to the recombination in immunology, any orderly rearrangement of gene segments would establish a new B cell. As shown in Figure 4, the recombination could be (a) between two features as crossover or (b) between many features as the combination of randomly selected gene

segments. With the use of normalization, the combination of gene segments will be in the determined order, such as in Figure 4(a) or could be randomly arranged as shown in Figure 4(b) in the computational respective,



**Figure 4: The Way of Rearrangement of Gene Segments (a) between Two Student Performance Prediction Samples (b) Among Several**

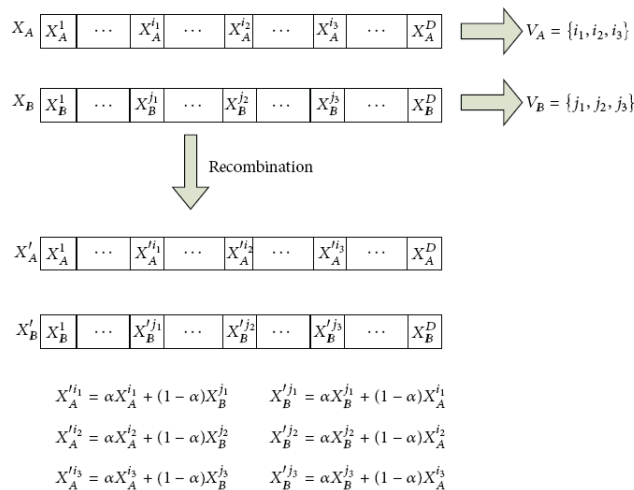
First identify two students samples attributes unsystematic from the education data, denoted by \$X\_A\$, \$X\_B\$, and then randomly select \$m\$ number of features \$m \in [1, D]\$, from each of them, in which feature index could be noted as vectors \$Fe\_A\$ and \$Fe\_B\$, respectively. The new selected features samples are generated by

$$X_A^{FeA} = \alpha X_A^{FeA} + (1 - \alpha) X_B^{FeB} \quad (2)$$

$$X_B^{FeB} = \alpha X_B^{FeB} + (1 - \alpha) X_A^{FeA} \quad (3)$$

Where \$\alpha\$ is a randomly selected number between 0 and 1. It should be noted that the range of each students features of

classification results should be normalized at first. The new ICSA algorithm the recombination operator is characterized in Figure 5. Two new students' samples are generated through the combinational recombination which is illustrated by Figure 5. In the students, \$m\$ equals 3. It requires to be known that \$i\_1\$ could be different from \$j\_1\$, the same as in \$i\_2\$ and \$j\_2\$ and \$i\_3\$ and \$j\_3\$ as long as the normalization has been done.



**Figure 5: Recombination Process of ICS**



Evaluation is performed on the fitness of the originally selected student's performance dataset. Together with the original student performance samples, two with the higher fitness will survive, and the other two individuals are rejected; Select two features with high fitness from the education data is measured as chosen features. Hypermutation operator transports diversity for the population by hosting perturbation for each clone. Although there exists many methods to implement this operator [27-28], alternatively proportional strategy remains to be in main basis. The concept of the operator [28] is followed in this research work, where every selected feature of the student samples are subject to  $M$  mutations without explicitly using a mutation probability. The inversely proportional law is used to calculate the number of the mutations  $M$ :

$$\alpha = \exp(-\rho f^*(X_i)) \quad (4)$$

$$M = \lfloor (\alpha \times n) + 1 \rfloor \quad (5)$$

Where  $f^*(X_i) \in [0,1]$  is the normalized fitness of  $X_i$ ,  $\rho$  is the decay constant which evaluates the shape of the mutation rate, and  $\lfloor \cdot \rfloor$  returns the lower bound integer.

$$X_i^j = \begin{cases} X_{r1}^j + \lambda(X_{r1}^j - X_{r2}^j) & \text{if } j \in \text{rand}M(n) \\ X_i^j & \text{elsewhere} \end{cases} \quad (6)$$

Then,  $M$  mutation is done on every candidate solution: ( $j$ ) is the  $j^{\text{th}}$  student features of the  $i^{\text{th}}$  education dataset,  $\text{rand}(n) \in \{1, \dots, n\}$  is randomly chosen  $M$  indexes without replication, and  $\lambda$  is a random number in the range of  $[-1, 1]$ .  $r1, r2 \in \{1, 2, \dots, N\}$  are randomly chosen numbers; The amplitude of the hypermutation is dogged automatically by the discrepancy of randomly chosen features in the dataset. The mutation equation (6) could be occupied as the variant of differential evolution The  $M$  strategy determines the direction counting the number of feature dimensions, while the equation governs the distance of the mutated clones with their parents. With union of both, the amplitude of the hypermutation is spontaneously determined with regard to the distribution of the population is shown in algorithm 1.

Algorithm 1. Proposed ICSEA algorithm

1. Initialization of the education dataset  $ab = X_i = (X_i(1), X_i(2), \dots, X_i(D))$ ,  $i = 1, \dots, N$  be the number of dataset samples,  $j = 1$  to  $D$  be the number of features are produced randomly within the range of boundaries of the decision space:
 
$$X_i^j = X_{\min}^j + \text{rand}(0,1)(X_{\max}^j - X_{\min}^j)$$
 $X_{\min}^j$  &  $X_{\max}^j$  are the lower and upper bound value of features  $j$  respectively.
2. Evaluation of fitness function (Classification accuracy) of the antibody population as their fitness
3. Generate new copies of the student features as antibodies. The recombination rate is set to  $N_r$ .
4. Mutate all the generated copies(hypermutation)
  - (i) Cloning: each individual  $X_i$  generates  $N_c$  copies  $\{X_i^1, X_i^2, \dots, X_i^{N_c}\}$ , where  $N_c$  is the clone number, which is a user defined constant
  - (ii) Hypermutation: each clone  $X_i^j, j = 1, \dots, N_c$ , goes through the

- hypermutation and creates the hypermutated clones  $X_i^j$ .
- (iii) Selection: select features the individual with highest fitness among  $X_i$  and hypermutated clones  $\{X_i^1, X_i^2, \dots, X_i^{N_c}\}$ .
5. Select the features with highest classification accuracy to endure(Selection)
6. Repeat Steps 2-5 until a termination criterion is happened(met)

In this work the fitness value is updated based on the weight values of the student features, a fitness value ( $fitness_i$ ) for a student features selection problem can be allocated to the solution.

$$fitness_i = \begin{cases} \frac{1}{(1 + fit_i \cdot we_i)} & \text{if } fe_i \cdot we_i \geq 0 \\ \frac{1}{(abs(fit_i \cdot we_i))} & \text{if } fe_i \cdot we_i < 0 \end{cases} \quad (7)$$

where  $fit_i$  is the classification accuracy. Assigned a weight  $we(fe_i)$  to each feature  $fe_i$ . The value of weight  $we(fe_i)$  for each  $fe_i$ , which is set to zero initially, is calculated sequentially throughout the whole matrix using the mean value of the attribute and update using the following formula when a new entry  $a_i$  is met in the discernibility matrix:

$$we_i = we(fe_i) \cdot \mu(fe_i) \quad (8)$$

When the optimization problem involves more than one objective function is described in equation (7-8), the task is to find more optimal student features solutions regarding each classifier accuracy.

Decision Support

Usually we use step for dealing with ESFS is selecting from majority voting, where the most agreed student feature is nominated as the final ensemble prediction. Similarly, a majority voting scheme with threshold may be implemented for ESFS. With the help of the notations introduced prior, for a given ensemble  $E$ , the consequences of the ensemble components can be systematized in a  $K \times M$  Boolean decision matrix  $D$ , where  $K$  is the size of the ensemble, and  $M$  is the total number of student features. In this representation, the horizontal row  $D_i$  denotes the student feature subset  $fe_i$ , and the binary cell value  $D_{ij}$  indicates whether  $fe_j \in fe_i$ . Borrowing the terminology of ensemble ESFS, the ensemble agreement  $j$  for the student features  $fe_j$  can therefore be calculated by:

$$\gamma_j = \frac{\sum fe_{ij}}{K} \quad (9)$$

An agreement threshold  $\alpha, 0 < \alpha \leq 1$ , can then be definite to control the number of physiological signal features being included in the final result  $fe^*$ , such that:  $fe^* = 1$ ; if  $\gamma_j > \alpha$ . From this, the common majority vote can be assimilated by setting  $\alpha = 0.5$ .

# Ensemble Swarm Based Feature Selection (Esfs) And Ensemble Three Classifiers (Etc) To Predict Student's Academic Performance

The value  $\alpha$  may be adjusted according to the problem at hand, if the amount of agreement is very high, a higher  $\alpha$  value can be used to control the size of the resultant feature subset. Alternatively, if a highly diverse 1, ESFS is obtained, there may exist no feature with  $y_j > 0.5$  to combat this, it may be necessary to employ a lowered  $\alpha$  value. At last, least error values are scheduled using intersection operation in numerical structure.

## Ensemble Three Classifiers (ETCS)

An ensemble of classifiers [29] is a group of classifiers whose distinct decisions are amalgamated in some manner frequently through weighted voting to categorize new student prediction samples. One of the most energetic fields of work in supervised learning is to revise techniques for building good ensembles of classifiers. The behavioral of student's classification method is associated through group of classifiers, called as Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier and Decision Tree (DT).

## 4. RESULTS AND DISCUSSION

Run the experiments on the PC containing 2GB of RAM, 4 Intel cores. For the experiments, used MATLAB we evaluate the proposed classification models and comparisons. This research student's collected data through two educational semesters: First and second, in which 245 students record collected during the first semester and 235 student's record collected during the second semester. In this experiments, use four common different measures for the evaluation of the classification quality: Accuracy, Precision, Recall and F-Measure [30-31]. Measures calculated using Table 1, which shows classification confusion matrix based on the Equations (10-13) respectively.

Table 1: Confusion Matrix

		Detected	
		Positive	Negative
Actual	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Precision is the ratio of the appropriately classified cases to the entire number of misclassified cases and correctly classified cases. Recall is the ratio of correctly classified cases to the total number of unclassified cases and correctly classified cases. In addition, we used the F-measure to combine the recall and precision which is considered a good indicator of the relationship between them [31].

Accuracy is the proportion of the total number of predictions where suitably valued.

**Precision (P):** Precision is defined as the percentages of predicted class which belongs to positive class that were correct, as determined using the equation:

$$\text{Precision} = \frac{A}{A+C} \quad (10)$$

**Recall (R):** Sensitivity is defined as the percentage of predicted and actual class which belongs to positive cases that were correctly identified, as determined using the equation:

$$\text{Sensitivity (Sen)} = \frac{A}{A+B} \quad (11)$$

**F-measure (F):** A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure

$$\text{F-measure} = 2 \cdot P \cdot R / (P + R) \quad (12)$$

**Classification Accuracy (CA):** Classification accuracy is defined as the percentage of the total amount of predictions which belongs to both positive and negative cases that were correctly identified, as determined using the equation:

$$\text{Classification accuracy (CA)} = \frac{A+D}{A+B+C+D} \quad (13)$$

Where A=True Positive (TP), B=True Negative(TN), C=False Positive (FP) and D=False Negative (FN). In this section, evaluate the impact of features on student's academic performance using different classification techniques such as DT, ANN, ANFIS and SVM. After relating the classification techniques on the data set, the results are unique based on different data mining measurements. Table 2 shows the classification results using the several classification algorithms. Each classifier introduces two classification results: (1) classification results with Students Features (SF) and (2) classification results with Students Selected Features (SSF) from ESFS.

Table 2: Classification results with respect to classifiers

Metric	DT(%)		ANN(%)		ANFIS(%)		SVM(%)		ETCs	
	SF	ES FS	SF	ES FS	SF	ES FS	SF	ES FS	SF	ES FS
Precision (%)	56.45	71.91	57.08	82.29	68.88	86.15	71.60	87.51	81.80	93.63
Recall (%)	54.78	72.85	56.78	83.37	69.48	86.89	71.86	88.23	82.51	93.87
F-measure (%)	55.60	72.38	56.93	82.83	69.18	86.52	71.73	87.87	82.15	93.75
Accuracy (%)	54.79	72.70	57.50	83.12	68.95	86.87	72.91	88.12	82.50	93.95

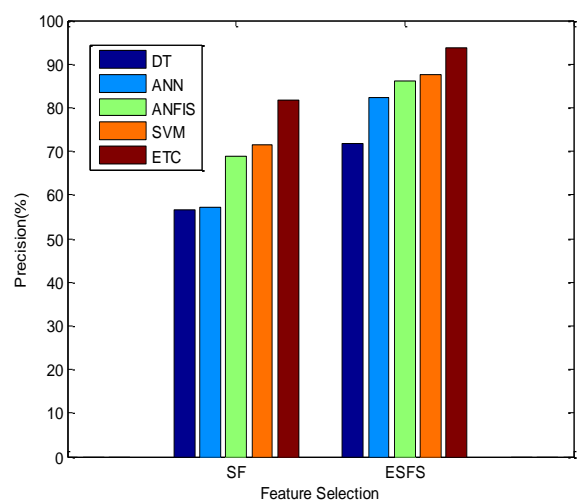
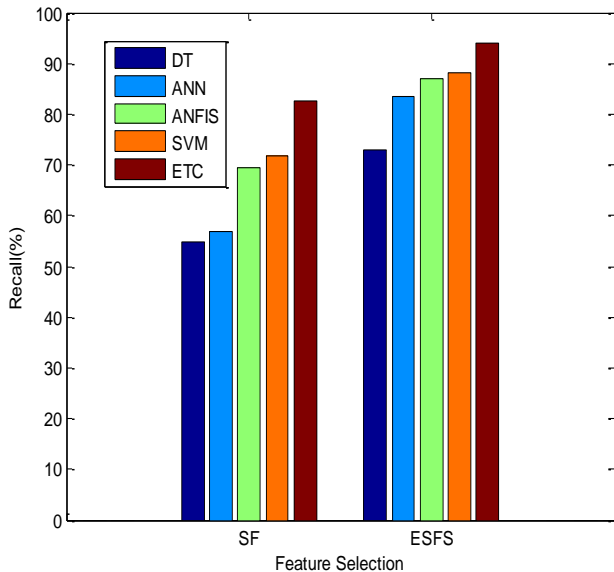


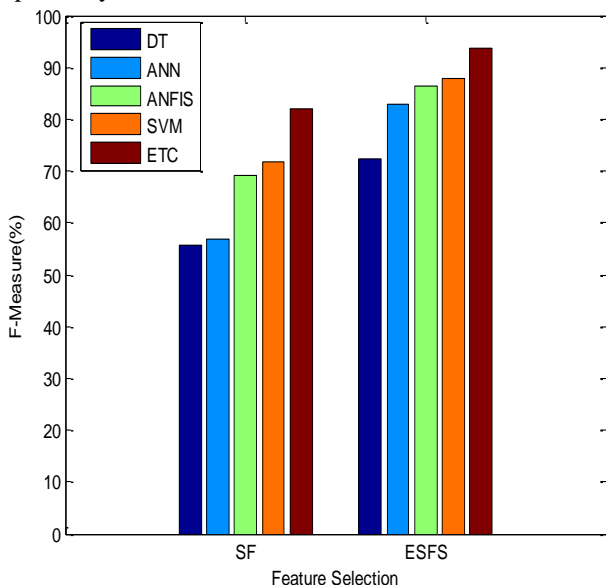
Figure 6: Precision comparison vs. prediction methods

Figure 6 shows the precision comparison results of the five different classifiers such as DT, ANN, ANFIS, SVM and ETC with respect to SF and ESFS algorithm. The results before and after feature selection are shown in figure 6. The proposed ETCs classifier has provides higher precision results of 93.63% after feature selection, whereas other DT, ANN,ANFIS, and SVM provides lower precision results of 71.91%, 82.29%, 86.15% and 87.51% respectively.



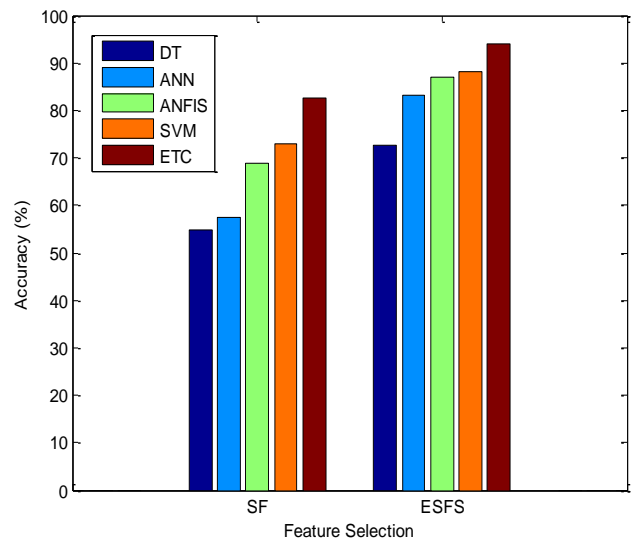
**Figure 7: Recall comparison vs. prediction methods**

Figure 7 shows the recall comparison results of the five different classifiers such as DT, ANN, ANFIS, SVM and ETC. The proposed ETCs classifier has provides higher recall results of 93.87% after feature selection, whereas other DT, ANN,ANFIS, and SVM provides lower recall results of 72.85%, 83.37%, 86.89% and 88.23% respectively.



**Figure 8: F-measure comparison vs. prediction methods**

Figure 8 shows the f-measure comparison results of the five different classifiers such as DT, ANN, ANFIS, SVM and ETC. The proposed ETCs classifier has provides higher f-measure results of 93.75% after feature selection, whereas other DT, ANN, ANFIS, and SVM provides lower recall results of 72.38%, 82.83%, 86.52% and 87.87% respectively.



**Figure 9: Accuracy comparison vs. prediction methods**

Overall prediction performance comparison results of the five different classifiers such as DT, ANN, ANFIS, SVM and ETC are shown in the Figure 9. The proposed ETCs classifier has gives higher accuracy results of 93.95% after feature selection, whereas other DT, ANN,ANFIS, and SVM provides lower accuracy results of 72.7%, 83.12%, 86.87% and 88.12% respectively.

## 5. CONCLUSION AND FUTURE WORK

In educational mining, the data of the student requires prediction methods to mine valuable knowledge. It has been renowned that a high performance accuracy of student prediction model will support for student and stakeholders. This paper proposes an Ensemble Swarm based Feature Selection (ESFS) is applied to the education data composed by learning management system (LMS), and the selected discriminative features are utilized to construct a student performance prediction model based on Ensemble Three Classifiers (ETCs). ESFS algorithm combines the procedure of the Fuzzy Membership Genetic Algorithm (FMGA) and Improved Clonal Selection Algorithm (ICSA). In the FMGA, global elitism is done on features to classification basis. FMGA is being used with the number of elite features is dynamic, i.e. it is changing from generation to generation. ICSA algorithm the recombination operator is used to improve the classification results, by recombination of two new students' samples are generated through the combinational. Finally ESFS algorithm, aggregating the feature selection results from FMGA and ICSA can be done by weighted voting, e.g. in the case of deriving a consensus feature ranking. Moreover, Ensemble Three Classifiers (ETCs) is proposed to predict student's performance by consolidating classifiers, like Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier and Decision Tree (DT). The results of the experiment show that proposed work have been compared with various classifiers algorithms such as DT, ANN, ANFIS and SVM.



## Ensemble Swarm Based Feature Selection (Esfs) And Ensemble Three Classifiers (Etc) To Predict Student's Academic Performance

The performance prediction model based on the ESFS algorithm has better prediction performance. Additionally, some influences and rules affecting student performance can be prolonged as scope of future work, which gives a technical reference for teachers, education management staffs and schools to predict and analyze the students' performances. The lopsidedness of the data might be the reason of expectation execution. Therefore, we will study this issue as future work.

### REFERENCES

- Osmanbegović E., M. Suljić, and H. Agić, "Determining Dominant Factor For Students Performance Prediction By Using Data Mining Classification Algorithms," *Tranzicija*, vol. 16, pp. 147-158, 2015.
- Shahiri M. and W. Husain, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414-422, 2015.
- Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, pp. 601-618, 2010.
- Ramaswami M. and R. Bhaskaran, "A study on feature selection techniques in educational data mining," *journal of computing*, volume 1, issue 1, pp.7-11, 2009.
- Nghe, N.T., Janecek, P. and Haddawy, P., 2007, A comparative analysis of techniques for predicting academic performance. In *2007 37th annual frontiers in education conference-global engineering: knowledge without borders, opportunities without passports*. pp. T2G-7.
- Vandamme, J.P., Meskens, N. and Superby, J.F., 2007. Predicting academic performance by data mining methods. *Education Economics*, 15(4), pp.405-419.
- Harb H. M. and M. A. Moustafa, "Selecting optimal subset of features for student performance model," *International Journal of Computer Science Issues*, Vol. 9, Issue 5, No 1, pp.253-262.
- Doshi M., "Correlation Based Feature Selection (CFS) Technique To Predict Student Performance," *International Journal of Computer Networks & Communications (IJCNC)* Vol.6, No.3, pp.197-206, 2014.
- Wang H, Khoshgoftaar TM, Napolitano A (2010) A comparative study of ensemble feature selection techniques for software defect prediction. In: *2010 Ninth International Conference on Machine Learning and Applications (ICMLA)*, pp 135-140.
- Amrieh, E.A., Hamtini, T. and Aljarah, I., 2016. Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application*, 9(8), pp.119-136.
- Mueen, A., Zafar, B. and Manzoor, U., 2016. Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science*, 8(11), pp. 36-42.
- Zaffar, M., Iskander, S. and Hashmani, M.A., 2018. A study of feature selection algorithms for predicting students academic performance. *Int. J. Adv. Comput. Sci. Appl*, 9(5), pp.541-549.
- Zaffar, M., Hashmani, M.A. and Savita, K.S., 2017, Performance analysis of feature selection algorithm for educational data mining. In *IEEE Conference on Big Data and Analytics (ICBDA)*, pp. 7-12.
- Anuradha, C. and Velmurugan, T., 2016, Feature Selection Techniques To Analyse Student Academic Performance Using Naïve Bayes Classifier. In *The 3rd International Conference on Small & Medium Business*, pp. 345-350.
- Zahedifard, M., Attarzadeh, I., Pazhokhzadeh, H. and Malekzadeh, J., 2015. Prediction of students' performance in high school by data mining classification techniques. *International Academic Journal of Science and Engineering*, 2(7), pp.25-33.
- Khasanah, A.U., 2017, A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. In *IOP Conference Series: Materials Science and Engineering* (Vol. 215, No. 1, p. 012036). IOP Publishing.
- Ajibade, S.S.M., Ahmad, N.B. and Shamsuddin, S.M., 2018, A Data Mining Approach to Predict Academic Performance of Students Using Ensemble Techniques. In *International Conference on Intelligent Systems Design and Applications* (pp. 749-760). Springer, Cham.
- Soni, A., Kumar, V., Kaur, R. and Hemavath, D., 2018. Predicting student performance using data mining techniques. *International Journal of Pure and applied Mathematics*, 119(12), pp.221-227.
- Cheng, C.H. and Liu, W.X., 2017. An appraisal model based on a synthetic feature selection approach for students' academic achievement. *Symmetry*, 9(11), pp.1-18.
- Almasri, A., Celebi, E. and Alkhalwaleh, R.S., 2019. EMT: Ensemble Meta-Based Tree Model for Predicting Student Performance. *Scientific Programming*, Vol. 2019, no.3610248, pp.1-13.
- Asif, R., Merceron, A., Ali, S.A. and Haider, N.G., 2017. Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, pp.177-194.
- G. Kakasevski, M. Mihajlov, S. Arsenovski and S. Chungurski, "Evaluating usability in learning management system Moodle", *30th International Conference on IEEE Information Technology Interfaces*, 2008. ITI 2008., pp. 613-618
- Yadav, S.K., Bharadwaj, B. and Pal, S., 2012. Mining Education data to predict student's retention: a comparative study, *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 10, No. 2, pp.113-117.
- Patro, S. and Sahu, K.K., 2015. Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462.
- Saeys, Y., Abeel, T. and Van de Peer, Y., 2008, Robust feature selection using ensemble feature selection techniques. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 313-325). Springer, Berlin, Heidelberg.
- Campelo, F., Guimarães, F.G., Igarashi, H. and Ramírez, J.A., 2005. A clonal selection algorithm for optimization in electromagnetics. *IEEE Transactions on Magnetics*, 41(5), pp.1736-1739.
- T. Jansen and C. Zarges, "Analyzing different variants of immune inspired somatic contiguous hypermutations," *Theoretical Computer Science*, vol. 412, no. 6, pp. 517-533, 2011.
- M. Pavone, G. Narzisi, and G. Nicosia, "Clonal selection: an immunological algorithm for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 53, no. 4, pp. 769-808, 2012.
- C. John Paul and Dr.R. Santhi, EBAFS-ETCs: Enhanced Bat Algorithm based Feature Selection and Ensemble Three Classifiers (ETCs) to Predict Student's Academic Performance, Vol.11, no. 2, pp. 268-280.
- D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC", *informedness, markedness and correlation*, (2011).
- T. Y. Chen, F. C. Kuo and R. Merkel, "On the statistical properties of the f-measure. In Quality Software, 2004. QSIC 2004", Proceedings. *Fourth International Conference on IEEE*, (2004), pp. 146-153.