AN EFFECTUAL LOGITBOOST BASED ARTIFICIAL BEE COLONY CLASSIFICATION ALGORITHM FOR PREDICTING STUDENTS PERFORMANCE

Author1: Dr. E. Chandra Blessie
Professor, Department of Computer Application
Nehru College of Management
Coimbatore

Author2: Vineetha K R
PhD Research Scholar, Department of Computer Application
Nehru College of Management
Coimbatore

Abstract- This paper explains the efficiency of Artificial Bee colony based classification (ABCC) strategy for classification purpose in the field of Educational Data Mining (EDM). More precisely, it anticipates ABCC to design classification model which has the ability of classifying the performance of students based on Bloom's Taxonomy. To conclude this, this paper anticipates a novel initialization purpose based on Effectual LogitBoost algorithm (ELBA) to mitigate adverse effects of curse of dimensionality on ABCC performance. Moreover, in the design of ELBA based ABCC model of performance classification, various feature selection methods are investigated. To perform this, this work considers online available dataset and real time data set, manually annotated with Blooms' cognitive levels, and transform into Rule discovery, Rule Pruning and prediction strategy. With the available dataset, numerous experiments have been conducted, and the outcomes depict better performance of the proposed ELBA-ABCC owing to its prediction accuracy. Moreover, when anticipated ELBA-ABCC based initialization method is utilized, a significant enhancement in the performance of students is attained. As well, the outcomes specifies feature selection like wrapper based and filter based approaches plays significant role in the performance of ELBA-ABCC. The simulation was performed in MATLAB environment, and the results were compared with the existing methods, which confirm the efficiency of the anticipated ELBA-ABCC approach. The attained accuracy is about 88.9%.
Keywords – Artificial Bee colony based classification; LogitBoost algorithm; Rule discovery; Rule pruning; Prediction strategy.

I. INTRODUCTION

Educational Data Mining (EDM) is an upcoming data mining field that spotlights on the development of approaches for exploring unique sorts of educational data that rises from an educational system or process [1]. In EDM, along with other data mining techniques, classification plays a significant role which emerges in diverse contexts and various techniques have been applied in it [2]. The complete review of EDM classification strategy reports that a specific offshoot of Meta heuristic approach known as ABC is effectual for data classification known as Artificial Bee Colony Classification to construct classification component in tools of teaching effectiveness [3], which is competent of classifying students performance based on cognitive levels of Bloom's taxonomy.

Artificial Bee colony (ABC) is a Meta heuristic optimization method inspired by the characteristics of honey bee and the control parameters. In general, ABC is a simple concept, easier to implement and possess few control parameters; it is extensively used in numerous optimization applications such as student's performance analysis, digital filters, artificial neural networks and other methods [4]. Traditional ABC has been applied to functional optimization problems successfully, moreover in current times a tremendous growth in the application it deals with to resolve various crisis in other areas is also reported [5]. Based on the above mentioned reason, data mining also utilizes it for its extensive application needs, where ABC is applied to crisis such as classification, clustering, outlier detection and feature selection [6]. For classification, ABC has recently gained increasing interest via specific ABC variant, termed Artificial Bee classification, though its primary application to classify the performance. Cumulative evidence as then recommends that ABC is an appropriate and competitive approach, which can be efficiently applied to demand classification problems, specifically when accurate, still comprehensible classifiers are required.

Moreover, to mitigate the difficult effects of curse of dimensionality [7], an inherent crisis of high dimensional dataset, this work anticipates a new specialized initialization method that is sourced on specific information retrieval procedure, called LogictBoost algorithm. Rationale behind anticipating LA-based initialization method can be depicted as follows: In numerous existing investigations, key role of initializing methods for evolutionary procedures [8], e.g. ABC, has been emphasized. Nevertheless, some investigators have initiated to question this high dimensional space. More particularly, it is reported that advanced generic initialization methods performs well in evaluating fitness function, and therefore performance of numerous evolutionary algorithms depreciate significantly [9]. Subsequently, a novel trend has initiated to spot lights on generating specialized initializing methods to enhance performance if evolutionary procedures in fitness value computation [10]. Specialized initialization method develops problem specific knowledge to recognize promising region in search space during the occurrence of evolution. In a specific region, which has been seems to be more promising for initialization of evolutionary procedures.

According to the theory of rule discovery and rule pruning which consists of higher probability of prediction to unknown optimal solution. According to this concept, LA is anticipated to recognize centre region of search space for data classification, and therefore provides a promising initialization of ABC optimization.

As feature selection process plays a significant role in data classification, the influence of feature selection approaches over the proposed ELBA-ABCC is anticipated. In precise, performance of ELBA-ABCC is investigated with the following feature selection and subset feature selection approaches: Wrapper based, Filter based, Filter based subset evaluation and Wrapper based subset evaluation. At last, anticipated ELBA-ABCC approach is validated by comparing the corresponding outputs with existing methods like Naïve Bayes (NB), k-nearest neighbor (kNN), Support Vector Machine (SVM), rule based ML algorithm (RIPPER, JRip), decision tree algorithm (J48), Bayesian Networks (ByesNet) and Adaptive Boosting method (AdaBoost) which have been functional to similar data set under similar settings.

The rest of the work is structured as follows: Section II explains in detail about the existing techniques involved in Educational Data Mining. Section III describes in detail about the proposed ELBA-ABCC method. The dataset description based on the online and real time availability student data, then wrapper and filter based and subset of these feature selection approaches are demonstrated in detail. The above mentioned techniques are used for pre-processing purpose, followed by this is the effectual ELBA-ABCC method for prediction strategy. Section IV explains about the numerical results and discussion of the anticipated ELBA-ABCC approach in terms of fitness value computation, cover percentage. Section V is the conclusion of the proposed work with the assistance for future direction.

II. LITERATURE REVIEW

In [11], Norlida Buniyamin et al. anticipated a classification technique known as Neuro-Fuzzy based classification for academic performance/achievements for electrical engineering students. This investigation illustrates about the selection of system to determine the intelligent information like mentor, advice, selecting course outline; adaptive learning model should categorize present situation and for this cause classifier is required- model can identify class value from diverse attributes. Classification techniques would permit more flexibility to moderate on group/single of student performance and Neuro-fuzzy linguistic is a value demonstrating probability of students to attain admirable grade even if student attained weak in specific subject/course.

In [12], Ashish Dutt et al. outlined numerous future insights on educational data clustering sourced on prevailing reviews and further opportunity for further investigation are recognized. As a key advancements of application of clustering procedure to data analysis is that, it offers relatively an unambiguous strategy of learning students style provided a number of variables such as time spent on effecting learning tasks, learning in groups, learner class behaviour, student motivation and classroom decoration towards learning. Clustering can offer pertinent insights to variables that are appropriate in separating clusters. Educational data is characteristically multi-level hierarchical and non-independent in nature, as recommended by the Baker & Yacef, henceforth investigator must carefully select clustering procedure that fulfils research question to attain reliable and valid outcomes.

In [13], Khokhoni Innocentia Mpho Ramaphosa et al, spotlights on examining prediction accuracy of academic performance of learners utilizing diverse classification procedures which are Naive Bayes, BayersNet, JRip and J48. The experimental outcomes proved that attributes selected from original dataset are extremely influential and it is sufficient to enhance prediction accuracy of unknown classes. J48 algorithm proved to be superior prediction model when contrast with other models with 99.13% accuracy on perfectly classified data with model data. This investigation will assist schools to determine academic status of learners in advance and concentrate on weak learners to enhance their academic outcomes.

In [14], Vinayak Hegde et al. spotlights on reason behind student dropout. Therefore, data collection acts as an important role in this investigation. Collected data are examined by various methods under data pre-processing. Data collected by diverse resources demonstrates numerous factors such as Demographical factors, Academics, Health issues, Psychological factors etc. plays a significant role in student dropout. As mentioned above, student dropout is a foremost threat to every educational institution. This replica will assist to recognize students who are going to drop out the registered course. While recognizing it, an early stage will avoid student drop out and can examine and provide valuable counselling to change the mind of student from dropout. It will as well demonstrate an appropriate path for student to attain their dreams.

In [15], Maryam Zaffar et al. depicted diverse feature selection procedures have been analyzed and evaluated. The outcomes on the student dataset have demonstrated that there is no significant change in feature selection performance algorithms accessible in WEKA tool. Moreover, with all these available feature selection techniques, principal components have demonstrated superior outcomes using Random Forest classifier. This investigation has also demonstrated that MLP classifier carried out slightly superior than subsequent classifier on student data set. The outcomes depicts that, there is an essential of subtle factor tuning of these feature selection techniques, to attain superior performance. Further studies on feature selection and their corresponding fusion can also be estimate; furthermore, student's datasets of diverse sizes can also be utilized for evaluations.

In [16], Xiaofeng Ma et al. explains the dependency amongst student features, initiating features dependencies and expert outlook to initiate coefficient, thus the algorithm can be converged quickly. In specific, to determine initial values of co-efficient, the author anticipates initialization of co-efficient rules. Moreover, to enhance the precision of the model, the author utilizes SVM procedure and DT algorithm optimized by grid search procedure to identify student pass rates and utilize information gain to discover student pass rate and utilize information gain to discover the features that possess higher influence on student performance. Experimental outcomes demonstrate that algorithm have attained better outcome, thus the model has been utilized to recognize significant features of students and pass rate of students.

In [17], Yang You et al. anticipated MIC-SVM, which is a highly effectual parallel support vector machine for x86 sourced multi-core and numerous core architectures like Intel KNC MIC and Intel Ivy Bridge CPUs. The author anticipates diverse novel optimization and analysis strategy that are common and can be effortlessly applied to accelerate other machine learning techniques. The investigation also improves and explores the deficiency of present SVM tools. At last, the author offer insights on how to place appropriate architectures to appropriate data patterns to attain best performance.

In [18], Arumugam, pre-eminence anticipated algorithm turns to be clearer in huge amount of datasets and micro-array gene expression data, with diminished amount of large training set idea is anticipated. It makes use of two SVM and utilizes data filter based on decision tree that scans complete data needed for small subset of data points. The anticipated approach is conceptually simple, easy to implement for this experiments, and quicker than the conventional SVM training algorithms. It as well captures data patterns and it offers appropriate information to acquire superior performance. The experimental outcomes on micro-array data set, large data set illustrates that anticipated approach is scalable for huge data classification, whilst provoking higher classification accuracy and effectually.

In [19], Md. Hasibur Rahman et al. utilized four conventional classification procedures, boosting ensemble technique, bagging ensemble technique and finally ensemble filtering technique. From the above mentioned techniques, ensemble filtering approach offers better accuracy. In ensemble filtering, it is seems that with characteristics of features accuracy by 25.8 percent accuracy with students' absent days accuracy enhanced by 16.1 percent. This enhancement of accuracy demonstrates strong influence of behavioural features and student absent days in class based on student's performance in academics. Furthermore, this method is constructed to enhance prediction of accuracy based in student's academic performance.

In [20], Auth Pisutaporn et al. described about students' alcohol consumption and recognized the parameters that possess significant influence over student's alcohol consumption. It is determined that male posses the tendency of consuming more alcohol consumption level lesser than female. The superior level of moving out with features leads to the higher consumption of alcohol. Moreover, students who have study time for short weekly, no added educational school support and does not deserve for higher education are extremely likely to have higher alcohol consumption. The author also identified that random forest algorithm carry out better performance than that of decision tree algorithm for classification crisis. At last, it is demonstrated that negative association amongst student grade and alcohol consumption.

III. MATERIALS AND METHODS

a. Dataset

We have presented dataset of 500 students with 16 includes that can be sorted as pursues:

1) Demographic Features, for example, sex, nationality, place of birth, parent.
2) Academic Features, for example, grade, semester, and section id, and subject and students absent days.
3) Behavioural Features, for example, rised hand up in class, viewing declarations, Discussion Groups, visited assets, parent noting review and parent school fulfilment.

These Features demonstrates the cooperation of students with the e-learning framework and furthermore present the parent investment in the learning procedure.

The proposed ELBA-ABCC technique has been analyzed through two data sets, that is, online available EDM dataset and the real time dataset from various knowledge fields [21]. The below given table gives the description about the dataset, which comprises of number of features, number of instances and number of classes for every data set. The available dataset have been extensively used in analysis of data classification, as the data set includes various numbers of classes and features, facilitating the analysis of influence on performance and accuracy when the features were selected.

This investigation utilized students' academic performance data sets that comprises of [] students' record and [] features. The students are categorized into three intervals of numerical values owing to its grade: Lower level: interval includes values ranges from 0-50; Intermediate level: interval includes values ranges from 51-75; and Higher level includes values that range from 75-100.

b. Feature selection in EDM

After the selection of data sets for further processing, based on this investigation, feature selection process has been carried out in the EDM. In general, feature selection is also termed as attribute selection or variable selection. Feature selection is the way of selecting the appropriate subset of relevant features for modelling the system. Feature selection techniques assists in creating accurate prediction models. It will offer better or good accuracy values when using fewer data. It can be utilized to eliminate or identify unnecessary, redundant or irrelevant attributes from data that will not contribute to accuracy of prediction model or it may even diminish model accuracy [22]. Fewer attributes are utilized as they diminish model complexity, and simpler models are easy to understand and to explain.

In recent times, there are two general methods that have been utilized in feature selection process, that are wrapper based and filter based model. The above mentioned model has been sub-divided into Filter based subset evaluation (FSE) and Wrapper based subset evaluation (WSE). At first, FSE has been introduced to eliminate the redundant features and the related issues that arise during filter ranking. It analyses the complete data subset in a multi-vibrate manner, chooses relevant features and explores degree of relationship between them. FSE is heuristic based technique that utilizes probabilities and statistical measures to search and analyse the usefulness of all features that is identified. Consequently, wrapper-based subset evaluation makes use of classifier to evaluate the significance of every feature subset. Typically, WBSE has superior predictive accuracy than FSE. This is due to the fact that selection approach is optimized while evaluating every feature subset with specific classification algorithm.

Recently, the wrapper based approach utilizes classification procedure to examine every set of features. This has turns to be excessively expensive to execute. However, while dealing with huge database that comprises of numerous features, the wrapper turns to be uncontrollable. Wrappers are connected with classifier algorithm and that makes it complex to shift from one classifier to another as selection process has to be total re-initiation. Dissimilar to wrapper, selection criteria of filters uses correlation functions and distance measures: as it does not necessitate re-execution for diverse learning classifiers. The outcome of this execution is extremely faster than that of wrapper based approaches. Filters are finely suited to large database environments that comprise numerous features. Researchers have frequently uses the filter as an alternative to wrapper, as the latter one is expensive and time consuming to run.

c. LogitBoost algorithm

The initiation of LogitBoost algorithm was started to overcome the drawbacks that are encountered using the Adaboost algorithm in handling outliers and noise. Logitboost algorithm utilizes binomial log likelihood that modifies the loss function linearly. As an alternative, Adaboost makes use of exponential loss function that modifies exponentially with classification error [23]. This is the cause of utilizing LogitBoost that tends to be less sensitive to noise and outliers. Based on the knowledge, no further investigations are carried out in analyzing the performance of Logitboost algorithm in the field of EDM.

IV. METHODOLOGY

In this research, anomaly detection approach comprises of two sub-divisions: pre-processing (hybrid feature selection) and data mining (classification algorithm). The figure 1 shows the flow diagram of the proposed work:

a. Pre-Processing

In pre-processing step, hybrid feature selection method is applied, so as to leverage the strength of both wrappers based and filter based approach. As well, anticipated filter based subset evaluation was utilized to resolve filter ranking crisis while encountering the occurrence of redundant features.

In stage 1, feature selection process starts with filter subset evaluation. It initially processes the original features 'F' and produces a new set 'S' of reduced features, where S ⊆ F. Here, correlation feature selection approach is adopted owing to its robustness in eliminating irrelevant and redundant features. This method overcomes the problem of redundant features in correlation feature selection as provided in Equation [1]. In addition, feature reductions in feature ranking are generally defined devoid of the need to carry out further processing (for instance, cover percentage, prediction strategy). Correlation feature selection is an intelligent filter algorithm that examines features subset owing to the heuristic evaluation functions.

$$F_s = \frac{\overline{krcf}}{\sqrt{k+k(k-1)\overline{rff}}} \qquad (1)$$

The above given equation specifies the merit function, F, which is utilized to choose a subset 's' comprising 'k' number of features. Both irrelevant and redundant features are depicted using 'rcf' which specifies the mean relationship of every feature to its class 'rff' which is the mean of relationship amongst the feature. An exhaustive search is not viable in huge datasets owing to its high complexity. As well, heuristic search techniques like artificial bee colony algorithm is used as a search function. This is due to the experimentation that reveals that ABC provides global optimum solution and is extremely robust than that of Particle swarm optimization (PSO), best first and greedy techniques. Moreover, with this stage, it is complex to provide assistance for truncation that computational effort utilizing wrapper approach as it deals only with reduced feature set in contrast to original feature sets.

In stage 2, reduced feature set 'S' accumulated from FSE was merged with the WSE approach to generate final set of optimal features K, where K ⊆ S ⊆ M. The anticipated wrapper and filter based hybridisation approach leverages strength of each to provide much better result based on false alarm rate, accuracy and fewer irrelevant and redundant features. This is owing to the fact that filter method will not fins best available subset, as it is less dependent on classifier. Alternatively, wrapper approach is seems to be extremely effectual and offer better accuracy. Nonetheless, it is computationally complex while dealing with huge amount of dataset. Therefore, it leverages strength of both the techniques while merging them with hybrid feature selection method. Logitboost algorithm and artificial bee colony algorithm is used in WSE to examine selected features with genetic search and to describe final 'K' feature subset. This search will be continued to train newer model for every subset and stop as the final optimum subset is determined.

Stage 3 is termed as classification stage. In this stage, final optimum subset 'K' generated by WSE was examined using ABC classifier with 10-fold cross validation. It comprises of numerous strategies for computing fitness function. Every steps of classification was constructed from diverse original dataset samples. Outcomes were determined based on votes attained from every tree that specifies the tree decision concerning class object. Most votes are based on the fitness computation.

Feature selection processes were generated with the training data that comprises of mixture of normal or accuracy prediction strategy. The feature significance were computed based on the correlation function in filter process, whilst in the wrapper process, classification algorithm is utilized. Moreover, features that are highly connected with other features describe the redundancy, and this feature has to be eliminated in the stage 1 and stage 2 processes itself. Moreover analysis on feature selection by proposed method is explained in the next stage.

b. Ensemble Classification-AABCC

In this section, the anticipated ensemble classifier technique based on boosting algorithm is explained in detail. Here, LogitBoost algorithm which is an enhanced version of boosting algorithm is used which serves as a meta classifier for boosting classification. The preliminary examination and experiments of investigations, found that this algorithm is extremely suited for dealing with outlier and noisy data over the extensively used Adaboost algorithm. Assume a training data set with 'N' samples and partitioned with two classes. Two classes are defined as B → {−1, +1}, that is, sample in class B = are instances of normal student performance, whereas B = -1 are instances of abnormal performance. Set the training data as$\{(A_1, B_1), \ldots, (A_i, B_i) \ldots (A_n, B_n)\}$, where A_i is feature vector and B_i is target class. LogitBoost algorithm comprises of subsequent steps:

1) Input dataset I = $\{(A_1, B_1), \ldots, (A_i, B_i) \ldots (A_n, B_n)\}$ where $A_i \in A$ and $B_i \in B = \{-1, +1\}$. Input number of 'K' iterations.
2) Initialized weights $W_i = {}^{1}/_{N}$, where $i = 1, 2, 3, \ldots, N$; Initiate committee function F(X) and probability estimates $P(X_i) = \frac{1}{2}$
3) Repeat $K = 1, 2, \ldots, K$

   i) Compute response and weights as in Equation [2] & [3]:

$$Z_i = \frac{B_i - p(A_i)}{P(A_i)(1-P(A_i)} \qquad (2)$$

$$W_i = P\,(A_i)(1 - P(A_i)) \qquad (3)$$

   ii) Fit function $f_k(A)$ by weighted least squares regression of $Z_i\ to\ X_i$ using weightsW_i. In this investigation, utilize artificial bee colony idea to fit data using weights W_i.

   iii) Update as in Equation [4] & [5]:

$$F\,(A) \to F\,(A) + \frac{1}{2}\,fk(x) \qquad (4)$$

$$P\,(X) = \frac{e^{F(A)}}{e^{F(A)} + e^{-F(A)}} \qquad (5)$$

   iv) Classifier output as in Equation 6:

$$sign\,[F(A) = sign\left[\sum_{k=1}^{k} fk(x)\right] \qquad (6)$$

Here, sign [F (A)] is function that has two probable output classes as in Eq.7:

$$sign\,[F(A)] = \begin{Bmatrix} 1, & if\ F(A) < 0 \\ -1, & if\ F(A) \geq 0 \end{Bmatrix} \qquad (7)$$

Fig 1: Flow diagram of proposed work

Algorithm:

1. Select the appropriate subset using the Filter based and wrapper based method which attains higher classification accuracy.
2. ABCC parameter setting comprises of size of bee colony, maximum cycles and restriction for trail.
3. Evaluate quality of food source using fitness function, which is the classification accuracy of predicting the performance.
4. Cycle 1
5. While $cycle < maximumcycles >$ Do
6. Produce new employed bees (new candidate solutions)
7. Generating new solution using fitness calculation.
8. Adopt LogitBoost approach
9. Probability value determination using fitness values.
10. Produce new onlooker bees (construct candidate solutions) with probability estimation
11. Generating new solution using fitness calculation.
12. Recognize the abandoned solutions and generate new solutions using scot bee randomly.
13. Save best solution
14. Cycle 2
15. End while
16. Generate prediction rate
17. Train ABCC based Logitboost approach
18. Classify the performance of bee(students)
19. Calculate classification accuracy
20. End

c. Rule format

In classification technique, classification rule for every attribute comprises of two parts: antecedent and consequent. Structure of rule generation is given below. Feature 1 to feature N are the attributes of dataset. Every attribute possess it lower bound value which is the lowest value for this rule (lowest performance of students) and upper bound is the highest value of this rule (highest performance). There exist other values that are connected with classification rule: predictive class (Class A), fitness value and cover percentage of rule [24]. These three values possess closer relationship amongst fitness function and prediction strategy.

```
Struct rulegeneration
{
Double *lower bound;
Double *upper bound;
Char cName[50];
Double prec;
Double *fitvalue;
};
```

d. Fitness Function

To examine fitness value, the fitness function will be utilized for classification purpose indeed of calculating nectar amount. The equation is given below 8:

$$Fitness\ value = \frac{TP}{TP+FN} * \frac{TN}{TN+FP} \quad (8)$$

Where TP, FN, FP and TN are sum of various record types and represents True positive, false positive, false negative and true negatives related to rule respectively [25]. Before providing these four values, two significant concepts have to be introduced:

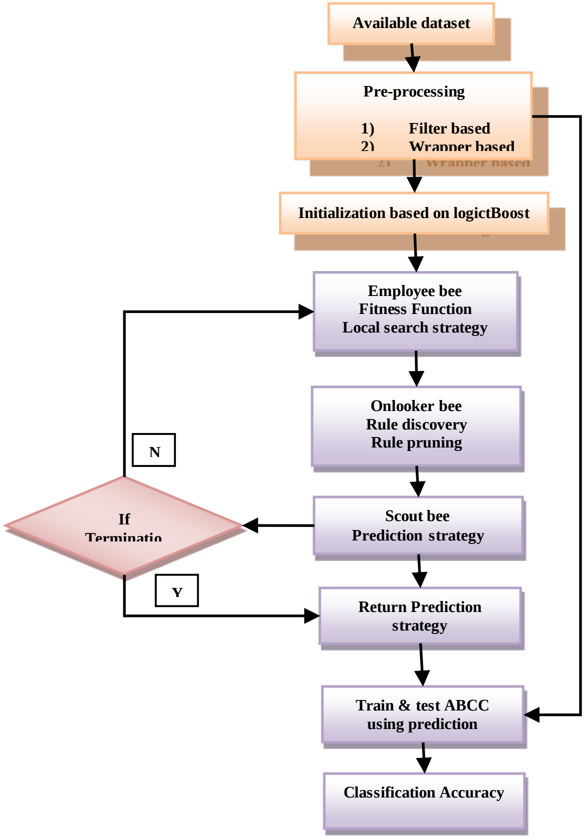

Fig 1: Flow diagram of the proposed ELBA-ABCC method

When the algorithm determines the type of record, it will compute each feature in the record. If value of feature lies between lower bound and upper bound, it is represented that the feature can be covered using the generated rules. If all features of record are covered under the generated rule; it specifies that the record will be covered by rules.

If class of evaluated record is equivalent to prediction class by rule, this specifies that record has class predicted by rule.

i. True positives (TP): number of records comes under generated rule that have class predicted by rule;
ii. False negatives (FN): number of records which does not generated with the rule but they have class predicted by rule;

iii. False positives (FP): number of records which comes under the rule generated but the class do not predicted by rule;

iv. iv. True negatives (TN): number of records not covered by generated rule and that do not possess class predicted by rule.

e. Exchanged Local Search Strategy

In ABCC, when an employed bee does not fulfil the requirements and does not reach maximum cycle number, it has to search for new food source followed by local search strategy. The standard ABC algorithm with the below mentioned algorithm has to execute local search technique. Moreover, it will take huge amount of time for classification of dataset that comprises huge amount of data and it is not appropriate for classification. In accordance to the ABCC, significant factors for classification and for enhancing the accuracy, a simpler local search strategy termed as 'Exchanged" is used to replace local search strategy as in Equation 9.

$$P_{ij} = NP_{kj} \qquad (9)$$

Where, P_{ij} specifies position of new food source and NP_{kj} specifies the neighbourhood of previous food source. The value of 'i' and 'k' lies between 1 and S, but 'k' has different value from 'i'. In addition, 'j' is number of dimension. In classification process, dimension of dataset is equivalent to number of features in dataset. $k \in \{1,2,\ldots,S\}$ and $j \in \{1,2,\ldots,D\}$ are chosen random parameters. When these parameters are given under the classification region as Equation [10] & [11] as given below:

$$P_{ij} lower\ bound = NP_{kj}\ lower\ bound \qquad (10)$$

$$P_{ij} upper\ bound = NP_{kj} upper\ bound \qquad (11)$$

Here, the value of 'k' is not equal to 'i'. To compute the performance of these two local search approaches, the online available dataset is utilized. The details of the dataset are given above and the related configuration control parameters have to be considered.

It is found that the proposed 'Exchanged' local search can outperform original local search in ABCC algorithm for data classification applications.

f. Rule Discovery

The ultimate objective of classification rule mining is to determine the set of rules which can determine specific class from various groups. Therefore, rule discovery phase is the most significant phase in classification algorithm, as rule set is the outcome of this phase. The ABCC based rule discovery shows the lower bound and the upper bound value of predicting the accuracy based on every attribute as in Equation [12] & [13]:

$Lower\ bound = f - k_1 * (F\_\max - F\_\min)$ (12)

$Upper\ bound = f + k_2 * (F\_\max - F\_\min)$(13)

In the above mentioned equation, F_{max} and $F_{min}$are maximum value and minimum value of features correspondingly. The difference amongst the feature value shows the mean range of features. 'f' is considered as the original feature value. K_1 and K_2 are two random values which lie between 0 and 1.

The classification rule mining algorithm can generate the rules automatically for every class. For selected class, it will discover rules iteratively till rule set can merge all instances belong to that class. Every rule abide by rule structure and rule set comprises of numerous rules.

g. Rule Pruning

After processing all the classes and entire rule have been generated, each rule will be given to rule pruning process. The ultimate goal of rule pruning is to eliminate redundant feature limitation that has been included unnecessarily in the rule set. As irrelative attributes with negative influence is considered for classification outcome, rule pruning will increase the accuracy. Process of rule pruning will be repeated till the set of effectual rules are generated.

h. Prediction Strategy

Pruned rule set will be utilized for prediction of new data in which their classes are unknown. Sometimes, testing data record will be produced with more than one rule for diverse classes. Prediction strategy will also describe which class has to be predicted. There are three significant steps for prediction strategy; it is specified as given below:

1) Compute prediction value for all rules which cover test data record;
2) Gather prediction value in accordance to various possible classes;
3) Choose class which possess highest prediction value as final class.

After the procedure of prediction strategy, the core is the prediction function which is used to compute the prediction value for each rule. It is defined in Eq.2.7 as below:

After the process of prediction, the ultimate target is to compute the prediction function that is utilized for evaluating prediction value for every rule as in Equation [14]:

$$Prediction_{value} = (\alpha * rule\ fitness\ value) + (\beta * rule\ cover\ percentage) \qquad (14)$$

Where $\alpha \in [0,1]$ and $\beta \in [1-\alpha]$ are two weighted parameters related to rule fitness value and rule cover percentage respectively. The above given equation can calculate fitness value for each rule. Rules cover percentage is depicted as proportion of records which is covered by rule that have class predicted by rule (TP). It is calculated by Equation given below [15]:

$$Cover\ \% = \frac{TP}{N} \qquad (15)$$

Where, 'N' is total amount of records which comes under the predicted class by rules. Prediction strategy balances the effect of cover percentage and fitness value for final predicted class. The values of α and β has to be selected more carefully, as they affect the classification accuracy. This new approach has provides a novel method for classification rule mining based ABCC-ELBA optimization algorithm. Finally, to evaluate proposed algorithm, compute accuracy for every validation data set and evaluate average of 'K' time validations as final accuracy.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present the datasets used for evaluation, the number of samples used in our experiments, experimental tools employed in our proposed approach, and the evaluation metrics adopted to measure the performance of the proposed approach.

Table I: Feature selection

Feature selection	No. of features	Selected features
Original feature	30	$f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17, f18, f19, f20, f21, f22, f23, f24, f25, f26, f27, f28, f29, f30$
Selected feature based on (WSE) & (FSE)	10	$f2, f4, f6, f8, f10, f12, f14, f16, f18, f20$

Once classification replica has been trained by 10-folds cross validation, the validation process starts. Validation is a significant phase in generating predictive model, it examines, how realistic predictive is. The validation model is utilized for 25 students.

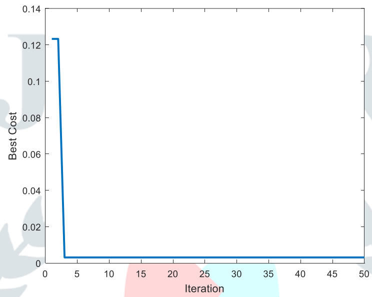

Fig 2: Simulated outcome of best cost vs number of iterations

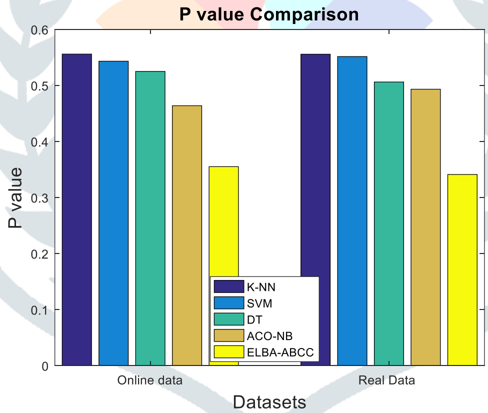

Fig 3: Graphical representation of P-value computation of ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

Figure 3 shows the graphical representation of P-value of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using ELBA-ABCC is 0.355, 0.341 respectively.

Table II: Tabular representation for P-value outcomes of ELBA-ABCC and existing k-NN, SVM, DT

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC
1	0.5558	0.5431	0.5249	0.4637	0.355
2	0.5556	0.5514	0.5060	0.4931	0.341

Table II depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT. The P-Value attained for prevailing methods are 0.5558, 0.5556, 0.5431, 0.5514, 0.5249, and 0.5060 correspondingly.

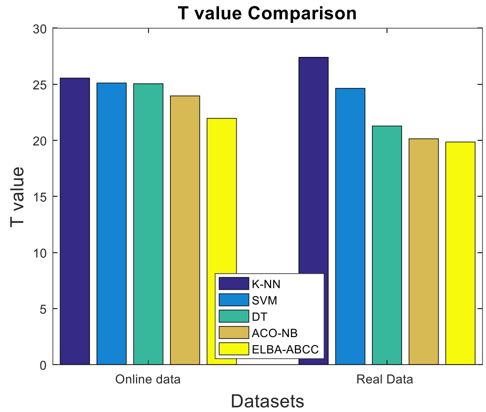

Fig 4: Graphical representation of T-measure for performing ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

Figure 4 shows the graphical representation of T-value of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using ELBA-ABCC is 21.96, 19.85 respectively.

Table III: Tabular representation for T-value outcomes of ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC
1	25.5546	25.1147	25.0510	23.9635	21.96
2	27.4008	24.6391	21.2804	20.1384	19.85

Table III depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT. The T-Value attained for prevailing methods are 25.5546, 27.4008, 25.1147, 24.6391, 25.0510, and 21.2804 correspondingly.

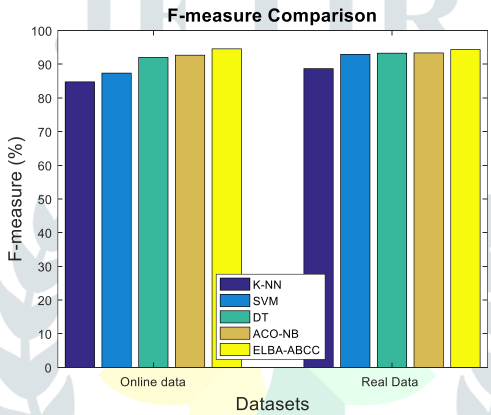

Fig 5: Graphical representation of F-measure for performing ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

Figure 5 shows the graphical representation of F-measure of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using ELBA-ABCC is 94.56 and 94.35 respectively.

Table IV: Tabular representation for F-measure outcomes of ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC
1	84.7702	87.3614	92.0183	92.6802	94.56
2	88.6957	92.9263	93.2766	93.3519	94.35

Table IV depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT. The F-measure attained for prevailing methods are 84.7702, 88.6957, 87.3614, 92.9263, 92.0183, 93.2766correspondingly.

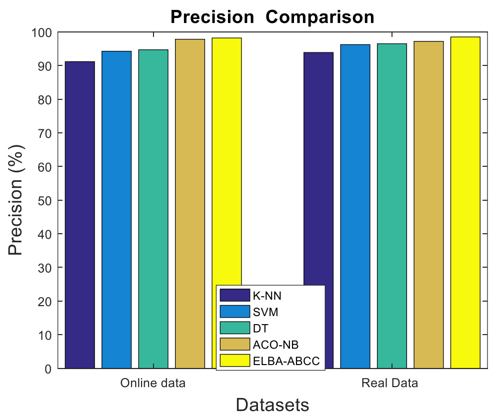

Fig 6: Graphical representation of precisio for performing ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

Figure 6 shows the graphical representation of precisio of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using ELBA-ABCC is 98.2 and 98.5 respectively.

Table V: Tabular representation for precision outcomes of ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC
1	91.1721	94.2630	94.7021	97.8204	98.2
2	93.8838	96.2255	96.4893	97.1901	98.5

Table V depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT. The precision attained for prevailing methods are 91.1721, 93.8838, 94.2630, 96.2255, 94.7021, and 96.4893 correspondingly.

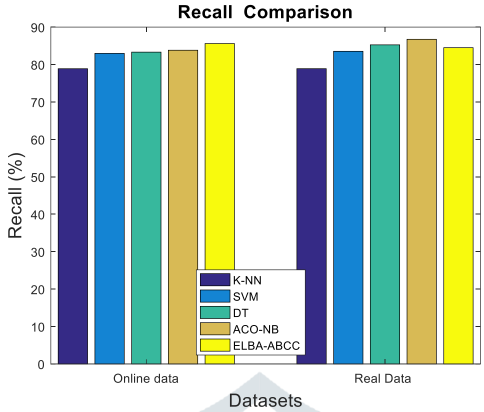

Fig 7: Graphical representation of Recall for performing ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

Figure 7 shows the graphical representation of Recall % of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using ELBA-ABCC is 85.61 and 84.5 respectively.

Table VI: Tabular representation for Recall outcomes of ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC
1	78.8785	82.9769	83.3104	83.8203	85.6
2	78.8910	83.5338	85.2581	86.7424	84.5

Table VI depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT. The Recall attained for prevailing methods are 78.8785, 78.8910, 82.9769, 83.5338, 83.3104, and 85.2581correspondingly.

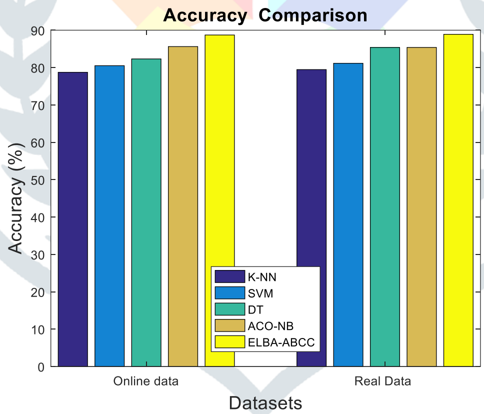

Fig 8: Graphical representation of Accuracy for performing ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

Figure 8 shows the graphical representation of Accuracy % of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using ELBA-ABCC is 88.7 and 88.9 respectively.

Table VII: Tabular representation for accuracy outcomes of ELBA-ABCC and existing k-NN, SVM, DT, ACO-NB

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC
1	78.7342	80.5064	82.3165	85.6190	88.72
2	79.4602	81.1438	85.3756	85.3888	88.9

Table VII depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT. The accuracy attained for prevailing methods are 78.7342, 79.4602, 80.5064, 81.1438, 82.3165, and 85.3756 correspondingly. As per the investigation, accuracy of 85.6190% is attained in our anticipated replica through validation phase. The outcome of validation phase shows reliability of anticipated replica.

VI. CONCLUSION

This work considers online available dataset and real time data set, manually annotated with Blooms' cognitive levels, and transform into Rule discovery, Rule Pruning and prediction strategy. With the available dataset, numerous experiments have been conducted, and the outcomes depict better performance of the proposed ELBA-ABCC owing to its prediction accuracy. Moreover, when anticipated ELBA-ABCC based initialization method is utilized, a significant enhancement in the performance of students is attained. As well, the outcomes specifies feature selection like wrapper based and filter based approaches plays significant role in the performance of ELBA-ABCC. The simulation was performed in MATLAB environment, and the results were compared with the existing methods, which confirm the efficiency of the anticipated ELBA-ABCC approach in terms of accuracy. The attained accuracy is about 88.9%. The future research direction is that the accuracy prediction of student's academic performance can be computed using any machine learning based hybrid classifiers.

REFERENCES

[1] Abaidullah, A. M., N. Ahmed, et al. (2015). "Identifying Hidden Patterns in Students' Feedback through Cluster Analysis." International Journal of Computer Theory and Engineering 7(1): 16-20.

[2] Ballou, D. and M. G. Springer (2015). "Using Student Test Scores to Measure Teacher Performance Some Problems in the Design and Implementation of Evaluation Systems." Educational Researcher 44(2): 77-86.

[3] Bhardwaj, B. K. P., Saurabh (2012). "Data Mining: A prediction for performance improvement using classification." (IJCSIS) International Journal of Computer Science and Information Security, pparXiv preprint arXiv:1201.3418 Vol. 9(4): 136-140.

[4] Fortenbacher, A., L. Beuster, et al. (2013). LeMo: A learning analytics application focussing on user path analysis and interactive visualization. Proceedings of the IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems.

[5] Gobert, J. D., Y. J. Kim, et al. (2015). "Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld." Thinking Skills and Creativity.

[6] Huang, S. and N. Fang (2013). "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models." Computers & Education 61:133-145.

[7] Mohd Arshad, P., N. Buniyamin, Ab Manan J. (2014). Neural Network and Linear Regression Methods for Prediction of Students' Academic Achievement. 2014 IEEE Global Engineering Education Conference (EDUCON), Istanbul, Trukey, IEEE.

[8] Rawat, S. S., S. S. Reddy, et al. (2015). "Ascertaining the Factors Influencing Students' Performance for Engineering Pedagogy." Journal of Engineering Education Transformations: 30-33.

[9] Sanzana, M. B., S. S. Garrido, et al. (2015). "Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests." Studies in Educational Evaluation 44: 50-59.

[10] Zhang Y, e. a. (2010). Using data mining to improve student retention in HE: a case study. Proc.12th Int. Conf. on Enterprise Information Systems, Volume 1: Databases and Information Systems Integration., Portugal.

[11] Norlida Buniyamin, "Educational Data Mining for Prediction and Classification of Engineering Students Achievement", IEEE 2015

[12] Ashish Dutt, "A Systematic Review on Educational Data Mining", IEEE 2016.

[13] Khokhoni Innocentia Mpho Ramaphosa, "Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools", IEEE 2018

[14] Vinayak Hegde, "Higher Education Student Dropout Prediction and Analysis through Educational Data Mining", ICISC 2018.

[15] Maryam Zaffar, "Performance Analysis of Feature Selection Algorithm for Educational Data Mining", IEEE Conference on Big Data and Analytics (ICBDA), 2017

[16] Xiaofeng Ma, "Student Pass Rates Prediction Using Optimized Support Vector Machine and Decision Tree", IEEE 2018.

[17] Yang You, "MIC-SVM: Designing A Highly Efficient Support Vector Machine For Advanced Modern Multi-Core and Many-Core Architectures", International Parallel & Distributed Processing Symposium, IEEE 2014.

[18] Arumugam.P, "Efficient Decision Tree Based Data Selection and Support Vector Machine Classification", Elsevier, 2017.

[19] Md. Hasibur Rahman, "Predict Student's Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques", 2nd International Conference on Electrical & Electronic Engineering (ICEEE), 27-28 December 2017, RUET, Rajshahi, Bangladesh.

[20] Auth Pisutaporn, "Relevant Factors and Classification of Student Alcohol Consumption", IEEE International Conference on Innovative Research and Development (ICIRD) 11-12 May 2018,Bangkok Thailand.

[21] Anwar Ali Yahya, "Swarm intelligence-based approach for educational data classification", Journal of King Saud University –Computer and Information Sciences

[22] Luthfia Rahman, "Feature Selection Methods in Improving Accuracy of Classifying Students' Academic Performance", International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017.

[23] Muhammad Hilmi Kamarudin, "A LogitBoost-Based Algorithm for Detecting Known and Unknown Web Attacks", IEEE Access.

[24] Mohd Afizi Mohd Shukran, "Artificial Bee Colony based Data Mining Algorithms for Classification Tasks", Modern Applied Science Vol. 5, No. 4; August 2011.

[25] Fabio Pagnotta and Hossain Amran, "Using Data Mining To Predict Secondary School Student Alcohol Consumption", 2016.