Software Fault Detection using Honey Bee Optimization

D. Asir Antony Gnana Singh
Anna University, BIT Campus,
Tiruchirappalli, India

A. Escalin Fernando
Anna University, BIT Campus,
Tiruchirappalli, India

E. Jebamalar Leavline, PhD
Anna University, BIT Campus,
Tiruchirappalli, India

ABSTRACT
The recent developments in the software technology assist humanities in various fields including engineering, technology, management, medical science, research, education, banking etc. Fault identification is a crucial one to the software testing professionals since a huge number of tests are carried out to identify the level of the defect. Therefore the machine learning algorithms are employed to develop software fault detection model in order to predict the fault in the software. The irrelevant and redundant test data reduces the accuracy of fault detection model. The accuracy of the fault detection model highly depends on the number of significant relevant test data. Therefore feature selection concept is applied to select the accurate features for developing the fault detection model. This paper proposes a method to select appropriate features with honey bee optimization technique for reducing the search space and to improve the accuracy in the software fault detection.

General Terms
Introduction to the feature selection, applications of feature selection, software fault detection using feature selection, data preprocessing for improving accuracy in supervised machine learning algorithm, application of data mining, honey bee optimization in data mining approaches.

Keywords
Feature selection, software fault detection algorithm, honey bee optimization, data mining approaches, machine learning algorithm, improving accuracy in classification algorithms.

1. INTRODUCTION
Nowadays different kinds of software are developed massively to satisfy the needs of the user and to reduce the manual work of the humans. The software is not only used in the field of engineering but also in various other fields. To solve the problem of the user and to assist them in various process of their respective fields the software packages are developed. Therefore, the development of the software becomes very essential. The software is developed with various phases such as analysis, design, implementation, testing, and deployment. The poor quality software and software failures can lead the materialistic and human loss. The quality of the software highly depends on various software metrics such as accuracy, robustness, and error. These software metric values are determined during the testing phase of software development process. Also, the testing phase is very important to determine the quality of software. The software fault detection can be a process of identifying the level of fault in the testing phase. Hence, the software fault detection plays a major role in improving the quality and to determine the possible failure of the software so that various losses can be avoided such as improper function of the machine which runs on the particular software.

Fault detection becomes difficult when the number of the test results is of high-dimension since the test results are used to detect the fault. Predicting the fault manually is impractical therefore the machine learning models are employed to automatically carry out the fault prediction. These machine learning models are known as fault detection model which can be developed using various supervised machine learning algorithms with software error prone data. Developing the fault detection model with high-dimensional data (more number of attributes) takes more computational time and reduces the prediction accuracy.

Since, the high-dimensional data contains redundant and irrelevant attributes. Therefore the redundant and irrelevant attributes have to be removed from the error prone data to improve the predictive accuracy in software fault detection. Attribute selection is a process of removing irrelevant and redundant features of the machine learning models. In order to remove the redundant and irrelevant features from the error prone data and to improve the accuracy in software fault detection, this paper proposes an attribute selection method namely attribute selection for software fault detection. In this proposed method, a honey bee optimization technique is combined with wrapper approach for selecting the significant features from the error prone data.

Machine learning is mainly employed to perform pattern recognition for various computational processes. The main objective of this technique is to build a model for prediction. This technique is mainly employed to manipulate difficult tasks. The machine learning techniques are broadly classified as supervised, unsupervised and reinforcement learning. The predictive model is built to obtain the features that are relevant and non redundant. The Naive Bayes (NB) classifier is used to obtain the result in the form of the probabilistic values which is independent on the relevancy between the selected features. In the Naive Bayes, the process is sustained by first retrieving the data, categorizing it and finally making the decision using support vector machine (SVM). The probability value is obtained by calculating the maximum likelihood of the features. The decision rule is combined with the prediction model to obtain the decisions about the features.

The feature selection is the process of selecting the relevant features which are the subset of features. The selected features are mainly used to create a simplification model, reduce training time and to enhance generalization. The features with non predictive or little information are eliminated. It is also defined as the process of removing the irrelevant and the redundant features from the high dimensional space. The feature selection algorithms are mainly classified into the categories of filter method, wrapper method and embedded method.

The filter method uses the statistical measure to assign a particular score to every feature. The features are given rank based on the score and then based on the priority it is decided whether to accept or remove the features from the dataset. Filter methods are mostly univariate and it considers the features either independently or with regard to the variables that are dependent. Chi squared test, information gain, correlation coefficient scores are some of the examples of the filter methods. The benefit of using the filter method is that the computational cost is low due to the absence of the classifiers. The drawback of this method is that accuracy is low and is mainly applicable for high dimensional data.

The wrapper method selects the features based on the search algorithm where various combinations of the inputs are created, evaluated and compared with others. It is purely a predictive model which assigns a score for each of the feature combination and hence improves the accuracy in the evaluation of the features. Some of the searching techniques such as best-first search, hill-climbing algorithm and forward backward passes are used for adding and removing of features. The major drawback of wrapper method is that accuracy is more only for the specific classifiers. The method is computationally inefficient and hence adaptable only for low dimensional data.

Embedded methods are the method of selecting the features that contribute to the accuracy of the classification model. They are commonly termed as regularization methods. They select the features which are mostly in the predictive form leading to lower complexity of the model. Some of the examples of embedded or regularization methods are Elastic Net and Ridge Regression. The accuracy of this method is high. The part of the learning algorithm is used for the process of feature selection. This method is computationally efficient than the wrapper method but costlier than the filter method.

Optimization technique is the process of selecting the best set of the features from all the available features. A function is created for the problem of optimization and the selection is done by either maximizing or minimizing the function based on the features that are available. The genetic algorithm is the process of performing the search using the technique of natural selection. The best result of the features is searched among the huge amount of the features. The search is done with the process of the selection, crossover and mutation similar to the natural process. Particle swarm optimization is an iterative method in which the features are selected in an iterative manner. The candidate feature is identified from all the available features and by evaluating all the features. The quality of the subset features is high by selecting the features by generating the intermediate solution as the candidate solution. The ABC (Artificial Bee Colony) algorithm works using the concept of the intelligent bee behavior. It is dependent entirely on the number of the features. This process first finds the feature vector to obtain the minimized objective function. The features are first selected randomly and then by applying the optimization function the neighborhood search is performed to obtain the best subset of the features.

2. RELATED WORD

The features are selected for creating a subset of the features to be given as the input for any classification model. Peng *et al* [1] discussed about the selection of features using dependency criterion on the maximum statistical value and mutual information. A feature subset is selected at a very low cost using the criteria of minimal redundancy maximal relevance (mRMR). The results obtained using the mRMR has quality features selected which in turn leads to improved accuracy in the classification. Jain *et al* [2] conducted a comparative study on various feature selection algorithms and concluded that a sequential forward floating selection algorithm is more efficient in producing accurate results.

Robert T *et al* [3] presented an online mechanism for the selection of features and this method can be used for the evaluation of the multiple set of features. The features in this method are selected based on the computation of the log likelihood ratios among the sample densities and their object background. Huan Liu *et al* [4] developed a new algorithm named as fast correlation based filter (FCBF) which is a novel concept for identifying features that are relevant and to reduce the redundancy among the features that are selected. This concept can be used for the high dimensional data and gives a better efficiency for any large amount of data.

Dan Sommerfield *et al* [5] proposed a method for dynamically searching and evaluating the information of the features in the subset. The performance of this algorithm can be improved by using a combined approach of ID3 and Naive-Bayes induction. This algorithm can be used for overcoming the over fitting problem that are caused during the searching process. Yixin *et al* [6] proposed a new technique known as MILES (multiple-instance learning via embedded instance selection) which correlates the instances. This way of feature selection produces redundancy and selects irrelevant features which are overcome using the SVM to select features for the classifiers henceforth the accuracy and efficiency can be increased.

Michael Jones *et al* [7] described an approach on the machine learning for the detection of visual objects in processing images. Chi2 [8] is an algorithm that is simple general that is used to eliminate the irrelevant attributes. This algorithm makes use of the chi-square statistic value to discretize the attributes and the process is continued till the occurrence of any inconsistency in the data.

Anil K. Jain *et al* [9] introduced the new methodology of selecting the features using the algorithm of Expectation-maximization to cluster the data. The feature selection from the cluster of the data is difficult due to the absence of class labels. This proposed algorithm can also be used when the number of the data is even extended to any number. Chong-Ho Choi *et al* [10] modified the existing algorithm of mutual information feature selector (MIFS) to overcome the limitations by mutual information between the input features and the output classes. They also proposed the Taguchi method to identify the good features. By combining the two algorithms a better features are obtained to deal with the classification problem.

Lei Yu *et al* [11] introduced an intelligent concept of feature selection by building an integrated system. Byung-Ro Moon *et al* [12] proposed a comparatively better hybrid genetic algorithm when compared to the existing normal genetic algorithm and sequential search. This hybridization produces not only a significant improvement in performance but also a controlled size in the subset. To perform the search the general search techniques are embedded in the hybrid genetic algorithm. Chong-Ho Choi *et al* [13] developed a new method for solving the classification problem. The solution is provided by calculating the mutual information which is based on using the Parzen window between the inputs and the variables in the class.

Shai *et al* [14] proposed a methodology for the online tracking using the ensemble classifiers. The combination of the weak classifier with the strong classifier is done using the Adaboost. The updated new weak classifier is able to maintain the temporal coherence. Karl Pfleger *et al* [15] provided a solution to find the subset of features consisting of relevant data. The solutions are provided using the subsequent searching and subset selection algorithms. The cross validation is applied for the subset selection of the features with the use of ID3 and C4.5.

Stefan Dietrich et al [16] presented an approach for selecting the features using the wrapper model which is proved to be more effective than decision tree. A new approach of artificial neural net input gain measurement approximation (ANNIGMA) was proposed to perform the searching of the features. This method is less expensive and accuracy is improved. Luis *et al* [17] applied the concept of feature selection for clustering. The experimental results confirm that that the newly developed procedure of combining the wrapper and filter method are theoretically far superior than the existing methods. They are proved to be more efficient than greedy searching algorithm in terms of computational cost.

Mao, K. Z. [18] developed a new concept of discriminative function pruning analysis (DFPA) which helps in the reduction of the classification speed and the cost. This process is induced by the usage of support vector machine (SVM) which acts as a discriminator function between the training data and the input variables. The advantage of using the DFPA method is that it combines the filter and wrapper methods. The filter method maintains the simplicity and the wrapper method is employed to improve the performance of the proposed algorithm.

Huang *et al* [19] used feature selection as an integral part of the subset selection in the concept of data mining. A two phase approach for removing the redundant features is followed in which the process begins by genetic feature selection with inconsistency criterion (GFSIC) which is purely a filter approach for removing the irrelevant features followed by the sensitivity-based feature selection with v-fold Cross-Validation (SBFCV) which is a wrapper approach. The effectiveness of the algorithm is found to be increased when it is used as the pre-processing step. Jyotirmoy Das *et al* [20] presented a feature selection using genetic programming. This technique is mainly adopted to construct a classifier model with the selected subset of features. This algorithm can be used in the ranking of features. The performance of the algorithm in terms of robustness is checked by synthetically adding redundant or irrelevant features.

Reunanen [21] compared various feature selection methods and computed cross-validation by evaluating the performance of different variable subsets. The problem with the evaluation schemes is that they consider biased predictions so they are not reliable. Giles M. Foody *et al* [22] explored that SVM is insensitive to dimensions of data hence it does not require the reduction of the dimensions. The classification accuracy is found to decline as the number of features increase. The accuracy with lesser number of features is more when compared with features of higher dimension.

Feiping *et al* [23] conducted empirical studies for proving effective feature selection for machine learning applications. Steve Goodison *et al* [24] proposed a new feature selection algorithm for converting the non linear complex problems in

the selection to the linearly relevant global features. The complexity of the sample is calculated in a logarithmic way based on the number of the features. The results are verified using supervised learning and the effectiveness of the features is evaluated. The analysis is done by purely distributing the data without making any assumptions of the features. Abbass *et al* [25] used the meta-heuristic and evolutionary techniques for optimization using various tools of the respective domains.

The new algorithm of Honey-bee mating optimization (HBMO) was developed by extending the features of the swarm-based optimization. This can be used for the non linear and continuous data by setting some constraints to obtain the decision. Bahriye Basturk *et al* [26] compared the performance of the constrained and unconstrained optimization algorithms. The result of the comparison is that the unconstrained optimisation results are far better than the constrained optimisation results. The Fisher discriminant [27] is used to rank the features based on the observations of the various samples. Feature extraction for high-dimensional data is also done using the nonparametric [28] methods. Here the rank is given for the entire samples in the dataset. The reduction of outliers is better even for abnormal datasets. This nonparametric selection leads to increased accuracy in selecting features.

Devis *et al* [29] proposed a procedure for selecting the model which is entirely dependent on the alignment of the kernel. Jiawei Han *et al* [30] aimed at finding the maximum of the lower bound than the traditional score. The fisher score is generally applied for obtaining feature subset which is sub-optimal. This method helps in providing solution for the linear programming that is quadratic constrained by gradient descent and ridge regression.

The neural network optimization for detecting wood defect [31] is done using the bees algorithm (BA). The search is performed imitating the searching behavior of the bees. The search is done among the neighbors adopting a general search which is the random search. The optimization is the process of performing a better search using the concept of error propagation. Bahriye Akay *et al* [32] presented a comparative study on optimization algorithms such as ant colony, particle swarm optimization, ant swarm and bird flocks. The bee swarm related algorithms are coined due to the intelligent behavior of the bee swarms.

Sam Kwong *et al* [33] analyzed various optimization algorithms and explored that artificial bee colony (ABC) performs well in exploration but weak in exploitation. So in order to improve the exploitation a new algorithm was coined by integrating the global information. Kalyanmoy *et al* [34] developed a sorting algorithm which is nondominated to overcome the complexity of computation, approach of non-elitism and specification of sharing parameter. The selection operator used here creates a mating pool by the various combinations of the parents and offspring. David E.Goldberg *et al* [35] introduced a compact genetic algorithm which uses the probability value of the distribution for the population. It is also proved that it requires low memory when compared with the ordinary genetic algorithm. The inertia weight is a parameter that is introduced along with the general working of the particle swarm optimisation. The simulated experimental results have proved the effectiveness of the accuracy. The introduction of the new term called inertia weight [36] significantly improved the performance.

Hartmut Schmeck *et al* [37] presented an approach for project scheduling problem which is resource-constrained. On comparing the presented approach with GA, Tabu search and simulated annealing. The proposed algorithm is found to outperform others average. Alex Freitas *et al* [38] proposed a new algorithm known as Ant-Miner for extracting classification rules from the data. They imitated the work of the ant and combined it with mining concepts. The performance of the developed Ant-Miner was compared with other well known data mining classification algorithms.

Nojun Kwak et al [39] discussed about two feature selection algorithms. The limitations of the mutual information feature selector (MIFS) algorithm are studied and the modification is done to overcome those limitations. The new proposed algorithm makes efficient use of the mutual information between the input attributes and the output classes. Another feature selection algorithm makes use of the Taguchi method. Combining the above two algorithms better results for the feature selection can be obtained. The major rule that is followed in the feature selection is the if-then clause but it is time consuming and difficult method.

The popular method that is used for tackling of the feature selection is the principal component analysis (PCA). In this method the existing attributes are transformed into new ones by considering the significant attributes for classification. The main drawback is the scaling of the attributes which causes changes in the results. The process of selecting the attributes proposed by Javed *et al* [40] follows a procedure which is of three steps. The first is the generation of all combinations of the attributes. The second is the selection of candidate attribute list followed by attaining the best set of attributes. The process of selecting the candidate attribute is done by creating a dataset and giving it as an input to the classifier to classify the data. The process is repeated until the value of the position of the attributes starts repeating. The candidate attribute set is prepared by sorting the paired attributes in the decreasing order. The attributes that heads the list are selected as the best candidate attributes. Shahla *et al* [41] worked on the combination of two major algorithms namely genetic algorithm and ant colony optimization. This hybrid algorithm yielded better result in terms of searching speed and selection of essential features. Huan Liu *et al* [42] discussed a solution for large datasets.

The combination of the forward and the wrapper methods are inbuilt in artificial bee colony (ABC) [43] which is used as the optimization technique to reduce the redundant data. The search for the features is done by two different methodologies namely full search and random search. The major methods that are followed for the feature selection is either filter method or wrapper method. In the filter approach the filtering of the features is done before the classification. In the wrapper approach the candidate feature is generated to produce the subset of the features. The Artificial Bee Colony (ABC) is used for the reduction of the dimensions for the specific feature selection in the optimization search of the dataset. DAAG Sing et al discussed the data mining and different approaches for feature selection [44-48].

Generally, classification model is built which is used as the prediction model for detecting the error present in the software. The goal of building the model is to minimize the number of the error in the developing software. The wrapper technique is applied evaluate the features selected and compared with the others to evaluate the performance of the other employed algorithms. This follows a strategy of searching for the features and obtains subset with optimal features, then applying any of the learning algorithms and finally the model is evaluated by testing the result with the selected subset of features.

3. METHODOLOGIES
3.1 Gain ratio
Information gain ratio is the metric to obtain information from various features that have multi values for the attributes based on the number of features and the size of the branch in the tree structure.

3.2 Info gain
The information gain measures the information to predict the class attribute from the information obtained from all the features in the dataset. It mainly concentrates on the reduction of entropy value for the features in the dataset.

3.3 Symmetric uncertainty
The symmetric uncertainty is a metric that measures the validity of the classification based on entropy. The accuracy is improved without affecting the relative functions of the classes.

3.4 ReliefF
The ReliefF is mainly applied for the selection of features using the concept of binary classification. It is independent of the information from other related features. It is used to improve the reliability of the approximation of the probabilistic function.

4. PROPOSED METHOD
4.1 Dataset details
The datasets are taken from the PROMISE repository where the features are the defects that are found in the testing phase of the project during the development stage. The number of features is varying from 9 to 40 with the class attribute. The features consist of number of operators, operands, branch count, numeric data, and variables that are present in the code. There are some of the static measures that are used to detect whether the software is error prone or not.

4.2 Experimental setup
The WEKA data mining tool is used in this experiment. The experiment is carried out by defining the number of selected features (NSF) for the threshold value. The threshold value strictly depends on the number of features on the dataset. The number of selected features in every iteration is denoted by β. The calculation for the fitness value is given by Equation (1).

$$\beta = N/R \tag{1}$$

where N is the total number of attributes in the dataset and R is a positive value to equally divide the features of the dataset.

The threshold value with different fitness value is given by Equation (2).

$$TV_x = x\beta \tag{2}$$

where $x \in \{1,2,.........N\}$ N is the total number of the features in the dataset.

4.3 Experimental procedure

The classification accuracy is obtained for various feature selection methods such as oneR (OR), gain ratio (GR), information gain (IG), reliefF (RF) and symmetric uncertainty (SU). The proposed method is carried out by giving the input dataset to the ranker, OneR attribute evaluation and the features are prioritized based on the merit value obtained from oneR. The values are selected based on the fitness values that are calculated using the Equation (1). The number of cycles is based on the attainment of the repetitive accuracy or it can be set manually. The number of features to be used in the each iteration is calculated using the Equation (2).

4.4 Proposed algorithm

Input: The dataset with entire features

Output: The selected set of the features

Assumptions:

1. Flowers are considered as nodes (features).

2. Honey content is considered as the calculated merit value of each flower (feature).

3. Bee is considered as an optimizing searching agent.

4. Bee is attracted by the TV number of flowers which has more honey (merit value).

5. The same quantity honey will be sucked from each flower (sucking rate).

6. The flower is visited till the maximum sucking capacity of the bee is reached.

Optimal search strategy

Step 1: begin

Step 2: Find the merit value for all the features.

Step 3: Select TV number of features based on the merit value of the features.

Go to evaluation strategy.

Step 4: Reduce the merit value of the features visited using the sucking rate.

Step 5: Order the features based on the modified merit value and go to step 3.

Evaluation strategy

Step 1: Find the fitness value (classification accuracy).

Step 2: Check the fitness value whether it is optimal or not.

If yes selected features are considered as most significant features.

If no go to step 4 in optimal search strategy.

5. RESULTS AND DISCUSSION

Table 1. Classification accuracy of NB with feature selection methods against TV on AR1 Dataset

NSF	Feature selection methods					
	OR	**GR**	**IG**	**RF**	**SU**	**PRO**
TV1	85.95	85.95	85.95	85.95	85.95	92.56
TV2	85.95	85.95	85.95	85.95	85.95	90.91
TV3	85.95	85.95	85.95	85.95	85.95	90.08
TV4	85.95	85.95	85.95	85.95	85.95	89.26
TV5	85.95	85.95	85.95	85.95	85.95	90.08

Table 2. Classification accuracy of NB with feature selection methods against TV on AR3 Dataset

NSF	Feature selection methods					
	OR	**GR**	**IG**	**RF**	**SU**	**PRO**
TV1	88.89	88.89	88.89	88.89	88.89	92.06
TV2	88.89	88.89	88.89	88.89	88.89	93.65
TV3	88.89	88.89	88.89	88.89	88.89	90.48
TV4	88.89	88.89	88.89	88.89	88.89	92.06
TV5	88.89	88.89	88.89	88.89	88.89	90.48

Table 3. Classification accuracy of NB with feature selection methods against TV on AR4 Dataset

NSF	Feature selection methods					
	OR	**GR**	**IG**	**RF**	**SU**	**PRO**
TV1	85.98	85.98	85.98	85.98	85.98	86.92
TV2	85.98	85.98	85.98	85.98	85.98	88.79
TV3	85.98	85.98	85.98	85.98	85.98	86.92
TV4	85.98	85.98	85.98	85.98	85.98	87.85
TV5	85.98	85.98	85.98	85.98	85.98	88.79

Table 4. Classification accuracy of NB with feature selection methods against TV on AR5 Dataset

NSF	Feature selection methods					
	OR	**GR**	**IG**	**RF**	**SU**	**PRO**
TV1	86.11	86.11	86.11	86.11	86.11	91.67
TV2	86.11	86.11	86.11	86.11	86.11	91.67
TV3	86.11	86.11	86.11	86.11	86.11	88.89
TV4	86.11	86.11	86.11	86.11	86.11	88.89
TV5	86.11	86.11	86.11	86.11	86.11	88.89

Table 5. Classification accuracy of NB with feature selection methods against TV on AR6 Dataset

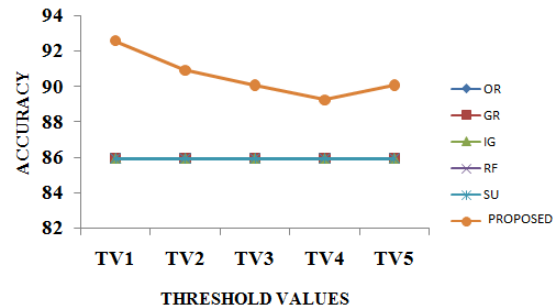| NSF	Feature selection methods					
	OR	**GR**	**IG**	**RF**	**SU**	**PRO**
TV1	85.15	85.15	85.15	85.15	85.15	88.12
TV2	85.15	85.15	85.15	85.15	85.15	88.12
TV3	85.15	85.15	85.15	85.15	85.15	88.12
TV4	85.15	85.15	85.15	85.15	85.15	87.13
TV5	85.15	85.15	85.15	85.15	85.15	88.12

Table 6. Classification accuracy of NB with feature selection methods against TV on CM1 Dataset

NSF	Feature selection methods					
	OR	GR	IG	RF	SU	PRO
TV1	84.94	84.94	84.94	84.94	84.94	88.96
TV2	84.94	84.94	84.94	84.94	84.94	88.35
TV3	84.94	84.94	84.94	84.94	84.94	87.95
TV4	84.94	84.94	84.94	84.94	84.94	87.35
TV5	84.94	84.94	84.94	84.94	84.94	87.15

Table 7. Classification accuracy of NB with feature selection methods against TV on DATATRIEVE Dataset

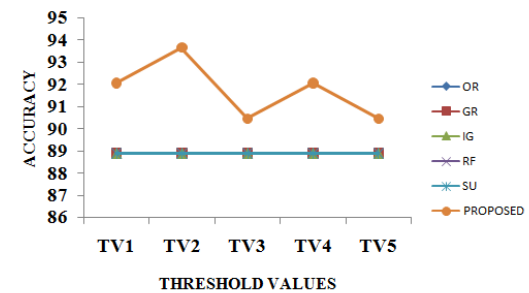| NSF	Feature selection methods					
	OR	GR	IG	RF	SU	PRO
TV1	83.08	83.08	83.08	83.08	83.08	92.31
TV2	83.08	83.08	83.08	83.08	83.08	87.69

Table 8. Classification accuracy of NB with feature selection methods against TV on KC1 Dataset

NSF	Feature selection methods					
	OR	GR	IG	RF	SU	PRO
TV1	82.55	82.55	82.55	82.55	82.55	85.44
TV2	82.55	82.55	82.55	82.55	82.55	84.54
TV3	82.55	82.55	82.55	82.55	82.55	84.16
TV4	82.55	82.55	82.55	82.55	82.55	83.55
TV5	82.55	82.55	82.55	82.55	82.55	83.17

Table 9. Classification accuracy of NB with feature selection methods against TV on KC2 Dataset

NSF	Feature selection methods					
	OR	GR	IG	RF	SU	PRO
TV1	83.72	83.72	83.72	83.72	83.72	85.06
TV2	83.72	83.72	83.72	83.72	83.72	84.67
TV3	83.72	83.72	83.72	83.72	83.72	84.10
TV4	83.72	83.72	83.72	83.72	83.72	84.48
TV5	83.72	83.72	83.72	83.72	83.72	84.48

Table 10. Classification accuracy of NB with feature selection methods against TV on KC3 Dataset

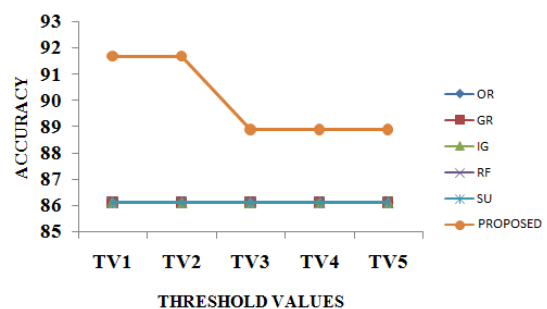| NSF	Feature selection methods					
	OR	GR	IG	RF	SU	PRO
TV1	84.93	84.93	84.93	84.93	84.93	89.74
TV2	84.93	84.93	84.93	84.93	84.93	88.21
TV3	84.93	84.93	84.93	84.93	84.93	87.77
TV4	84.93	84.93	84.93	84.93	84.93	86.90
TV5	84.93	84.93	84.93	84.93	84.93	86.68

Fig 1: Classification accuracy of NB with feature selection methods against TV on AR1 Dataset

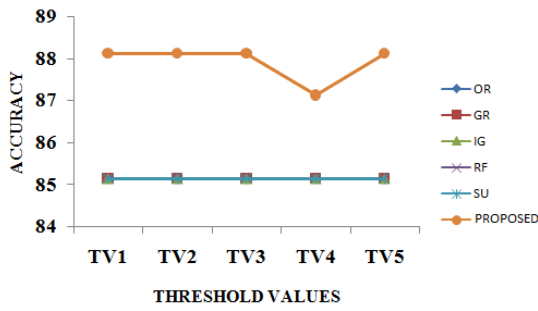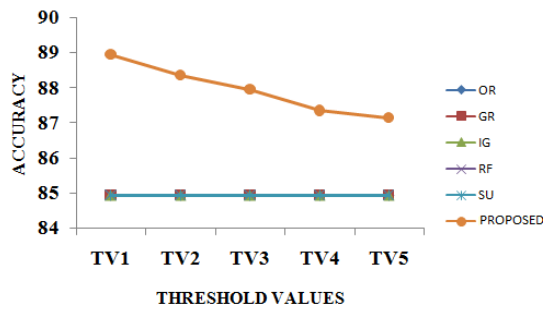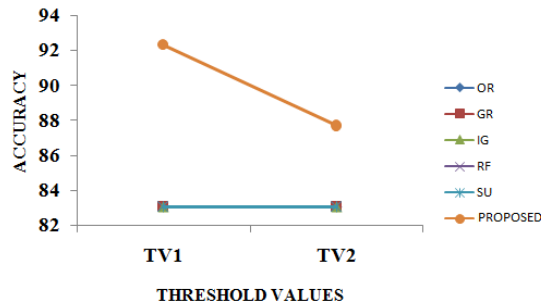**Fig 2: Classification accuracy of NB with feature selection methods against TV on AR3 Dataset**
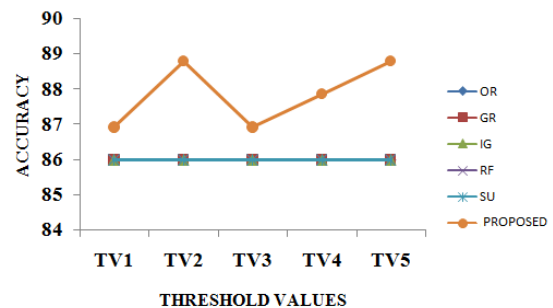
Fig 3: Classification accuracy of NB with feature selection methods against TV on AR4 Dataset

Fig 4: Classification accuracy of NB with feature selection methods against TV on AR5 Dataset

Fig 5: Classification accuracy of NB with feature selection methods against TV on AR6 Dataset

Fig 6: Classification accuracy of NB with feature selection methods against TV on CM1 Dataset

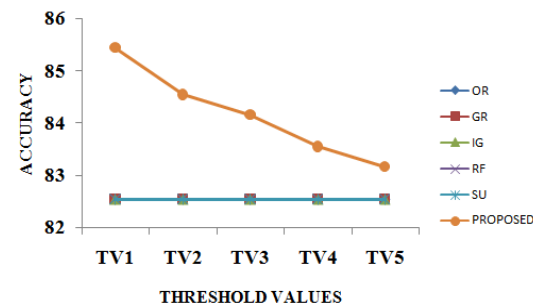**Fig 7: Classification accuracy of NB with feature selection methods against TV on DATATRIEVE Dataset**

Fig 8: Classification accuracy of NB with feature selection methods against TV on KC1 Dataset

Table 1 to Table 10 and Fig 1 to Fig 10 illustrate the classification accuracy of NB classifier with various feature selection methods with respective threshold values TV that denotes the number of selected features on the datasets. From Table 1 to Table 10 and Fig 1 to Fig 10, it is observed that the existing methods such as OR, GR, IG, RF and SU produced same accuracy for all the datasets. The proposed method

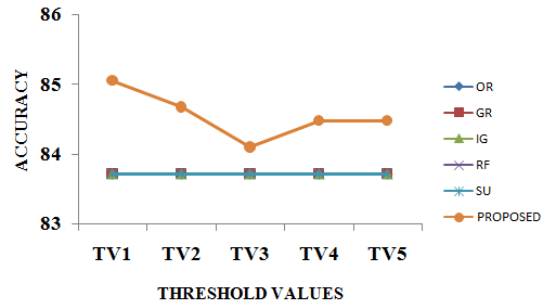produces higher accuracy than the existing feature selection methods for all the datasets.

Fig 9: Classification accuracy of NB with feature selection methods against TV on KC2 Dataset
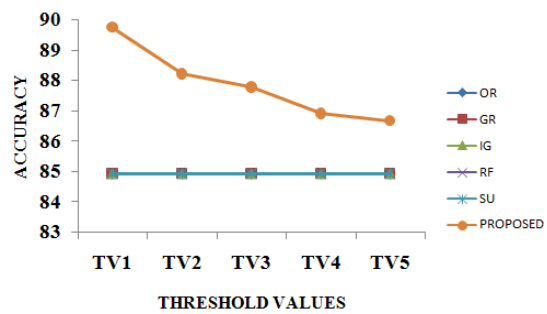

Fig 10: Classification accuracy of NB with feature selection methods against TV on KC3 Dataset

6. CONCLUSION

This paper proposed a honey bee optimization based algorithm for selecting features for software fault detection. The performances of the various feature selection methods such as OR, GR, IG, RF and SU along with the proposed honey bee optimization are compared and the proposed method is found to produce higher classification accuracy. The number of features to be selected is based on the threshold value that depends on the number of features in the dataset. Therefore, using the proposed honey bee optimization algorithm, the features that are more effective for the detection of the fault in the software can be predicted using the naive Bayes classifier.

7. REFERENCES

[1] Peng, Hanchuan, Fuhui Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." Pattern Analysis and Machine Intelligence, IEEE Transactions on 27.8 (2005): 1226-1238.

[2] Jain, Anil, and Douglas Zongker. "Feature selection: Evaluation, application, and small sample performance." Pattern Analysis and Machine Intelligence, IEEE Transactions on 19.2 (1997): 153-158.

[3] Collins, Robert T., Yanxi Liu, and Marius Leordeanu. "Online selection of discriminative tracking features." Pattern Analysis and Machine Intelligence, IEEE Transactions on 27.10 (2005): 1631-1643.

[4] Yu, Lei, and Huan Liu. "Feature selection for high-dimensional data: A fast correlation-based filter solution." ICML. Vol. 3. 2003.

[5] Kohavi, Ron, and Dan Sommerfield. "Feature Subset Selection Using the Wrapper Method: Overfitting and Dynamic Search Space Topology." KDD. 1995.

[6] Chen, Yixin, Jinbo Bi, and James Z. Wang. "MILES: Multiple-instance learning via embedded instance selection." Pattern Analysis and Machine Intelligence, IEEE Transactions on 28.12 (2006): 1931-1947.

[7] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[8] Liu, Huan, and Rudy Setiono. "Chi2: Feature selection and discretization of numeric attributes." tai. IEEE, 1995.

[9] Law, Martin HC, Mario AT Figueiredo, and Anil K. Jain. "Simultaneous feature selection and clustering using mixture models." Pattern Analysis and Machine Intelligence, IEEE Transactions on 26.9 (2004): 1154-1166.

[10] Kwak, Nojun, and Chong-Ho Choi. "Input feature selection for classification problems." Neural Networks, IEEE Transactions on 13.1 (2002): 143-159.

[11] Liu, Huan, and Lei Yu. "Toward integrating feature selection algorithms for classification and clustering." Knowledge and Data Engineering, IEEE Transactions on 17.4 (2005): 491-502.

[12] Oh, Il-Seok, Jin-Seon Lee, and Byung-Ro Moon. "Hybrid genetic algorithms for feature selection." Pattern Analysis and Machine Intelligence, IEEE Transactions on 26.11 (2004): 1424-1437

[13] Kwak, Nojun, and Chong-Ho Choi. "Input feature selection by mutual information based on Parzen window." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.12 (2002): 1667-1671.

[14] Avidan, Shai. "Ensemble tracking." Pattern Analysis and Machine Intelligence, IEEE Transactions on 29.2 (2007): 261-271.

[15] John, George H., Ron Kohavi, and Karl Pfleger. "Irrelevant features and the subset selection problem." Machine Learning: Proceedings of the Eleventh International Conference. 1994.

[16] Hsu, Chun-Nan, Hung-Ju Huang, and Stefan Dietrich. "The ANNIGMA-wrapper approach to fast feature selection for neural nets." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 32.2 (2002): 207-212.

[17] Talavera, Luis. "An evaluation of filter and wrapper methods for feature selection in categorical clustering." Advances in Intelligent Data Analysis VI. Springer Berlin Heidelberg, 2005. 440-451.

[18] Mao, K. Z. "Feature subset selection for support vector machines through discriminative function pruning analysis." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 34.1 (2004): 60-67.

[19] Yuan, Huang, et al. "A two-phase feature selection method using both filter and wrapper." Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference

Proceedings. 1999 IEEE International Conference on. Vol. 2. IEEE, 1999.

[20] Muni, Durga Prasad, Nikhil R. Pal, and Jyotirmoy Das. "Genetic programming for simultaneous feature selection and classifier design." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 36.1 (2006): 106-117.

[21] Reunanen, Juha. "Overfitting in making comparisons between variable selection methods." The Journal of Machine Learning Research 3 (2003): 1371-1382.

[22] Pal, Mahesh, and Giles M. Foody. "Feature selection for classification of hyperspectral data by SVM." Geoscience and Remote Sensing, IEEE Transactions on 48.5 (2010): 2297-2307.

[23] Nie, Feiping, et al. "Efficient and robust feature selection via joint ℓ_2, 1-norms minimization." Advances in neural information processing systems. 2010.

[24] Sun, Yijun, Sinisa Todorovic, and Steve Goodison. "Local-learning-based feature selection for high-dimensional data analysis." Pattern Analysis and Machine Intelligence, IEEE Transactions on 32.9 (2010): 1610-1626.

[25] Afshar, Abbass, et al. "Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation." Journal of the Franklin Institute 344.5 (2007): 452-462.

[26] Karaboga, Dervis, and Bahriye Basturk. "Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems." Foundations of Fuzzy Logic and Soft Computing. Springer Berlin Heidelberg, 2007. 789-798.

[27] Wang, Jianyu, Xilin Chen, and Wen Gao. "Online selecting discriminative tracking features using particle filter." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2. IEEE, 2005.

[28] Kuo, Bor-Chen, and David Landgrebe. "Nonparametric weighted feature extraction for classification." Geoscience and Remote Sensing, IEEE Transactions on 42.5 (2004): 1096-1105.

[29] Tuia, Devis, et al. "Learning relevant image features with multiple-kernel classification." Geoscience and Remote Sensing, IEEE Transactions on 48.10 (2010): 3780-3791.

[30] Gu, Quanquan, Zhenhui Li, and Jiawei Han. "Generalized fisher score for feature selection." arXiv preprint arXiv:1202.3725 (2012).

[31] Pham, Duc Truong, et al. "Optimising neural networks for identification of wood defects using the bees algorithm." Industrial Informatics, 2006 IEEE International Conference on. IEEE, 2006.

[32] Karaboga, Dervis, and Bahriye Akay. "A survey: algorithms simulating bee swarm intelligence." Artificial Intelligence Review 31.1-4 (2009): 61-85.

[33] Zhu, Guopu, and Sam Kwong. "Gbest-guided artificial bee colony algorithm for numerical function optimization." Applied Mathematics and Computation 217.7 (2010): 3166-3173.

[34] Deb, Kalyanmoy, et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II." Evolutionary Computation, IEEE Transactions on 6.2 (2002): 182-197.

[35] Harik, Georges R., Fernando G. Lobo, and David E. Goldberg. "The compact genetic algorithm." Evolutionary Computation, IEEE Transactions on 3.4 (1999): 287-297.

[36] Shi, Yuhui, and Russell Eberhart. "A modified particle swarm optimizer." Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on. IEEE, 1998.

[37] Merkle, Daniel, Martin Middendorf, and Hartmut Schmeck. "Ant colony optimization for resource-constrained project scheduling." Evolutionary Computation, IEEE Transactions on 6.4 (2002): 333-346.

[38] Parpinelli, Rafael S., Heitor S. Lopes, and Alex Freitas. "Data mining with an ant colony optimization algorithm." Evolutionary Computation, IEEE Transactions on 6.4 (2002): 321-332.

[39] Kwak, Nojun, and Chong-Ho Choi. "Input feature selection for classification problems." Neural Networks, IEEE Transactions on 13.1 (2002): 143-159.

[40] Khan, Javed, et al. "An attribute selection process for software defect prediction." Informatics, Electronics & Vision (ICIEV), 2014 International Conference on. IEEE, 2014.

[41] Nemati, Shahla, et al. "A novel ACO–GA hybrid algorithm for feature selection in protein function prediction." Expert systems with applications 36.10 (2009): 12086-12094.

[42] Dash, Manoranjan, and Huan Liu. "Feature selection for classification." Intelligent data analysis 1.1 (1997): 131-156.

[43] Schiezaro, Mauricio, and Helio Pedrini. "Data feature selection based on Artificial Bee Colony algorithm." EURASIP Journal on Image and Video Processing 2013.1 (2013): 1-8.

[44] D. Asir Antony Gnana Singh, S.Appavu alias Balamurugan, E. Jebamalar Leavline, "A novel feature selection method for image classification", Optoelectronics and Advanced Materials – Rapid Communications, vol. 9, no. 11-12, 2015. ISSN: Print: 1842-6573.

[45] D. Asir Antony Ganana Singh and E. J. Leavline, "Data Mining In Network Security - Techniques & Tools: A Research Perspective", Journal of Theoretical and Applied Information Technology, vol. 57, (2013).

[46] D.Asir Antony Gnana Singh ,S.Appavu alias Balamurugan, E.Jebamalar Leavline, "Improving performance of Supervised Learners using Unsupervised Variable Selection Algorithm: A novel approach", International journal of soft computing,vol. 9, No.5, pp.303-307, 2014, ISSN : 1816-9503

[47] D.Asir Antony Gnana Singh ,S.Appavu alias Balamurugan, E.Jebamalar Leavline, "Improving the Accuracy of the Supervised Learners using Unsupervised based Variable Selection", Asian Journal of Information Technology, vol. 30, No.9, pp. 530 – 537, 2014, ISSN : 1682-3915 (Print), ISSN : 1993-5994 (Online)

[48] D. Asir Antony Gnana Singh, E. Jebamalar Leavline, Decision making in enterprise computing: A data mining approach, International journal of core engineering and management, vol. 1, issue 11, Feb 2015, pp. 103 - 113, ISSN: 2348-9510