CONF-IRM 2022 Proceedings

International Conference on Information Resources Management (CONF-IRM)

10-2022

Present a feature selection technique based on machine learning in order to increase the detection rate of classifiers using CHOA algorithm

Hassan Ghaedi

Sahar Soleimani

20. Present a feature selection technique based on machine learning in order to increase the detection rate of classifiers using CHOA algorithm

Hassan Ghaedi
Department of Computer, Khormuj Branch,
Islamic Azad University, Khormuj, Iran
hassan.ghaedi@iau.ac.ir

Sahar Soleimani
Department of Computer, Khormuj Branch,
Islamic Azad University, Khormuj, Iran
s.soleimani@samabushehr.ac.ir

Abstract

In creating a pattern classifier, feature selection is often used to prune irrelevant and noisy features to producing effective features. Feature selection algorithms; try to classify an instance with lower dimension, instead of huge number of required features, with higher and acceptable accuracy. In fact an instance may contain useless features which might result to misclassification. An appropriate feature selection methods tries to increase the effect of significant features while ignores insignificant subset of features. In this paper, an efficient feature selection algorithm based on Cheetah optimization algorithm and support vector machine (CHOA-SVM) was used. First a population of cheetahs (feature subsets) were randomly generated, and then optimized by CHOA-SVM wrapper algorithms; finally the best fitness feature subset was applied to SVM classification. Experiments over a standard benchmark demonstrate that applying CHOA-SVM in the context of feature selection is a feasible approach and improves the classification results. The simulation experiment results have proved that the feature subset selection algorithm based on CHOA-SVM is very effective.

Keywords: Feature Selection, Machine Learning, Data Mining, Pattern Classification, Cheetah Optimization Algorithm.

1. Introduction

In many pattern classification problems, using a large number of features does not increase the accuracy. Furthermore, sometimes using some features even decreases the prediction accuracy. Therefore, feature selection is used as a valuable pre-processing tool before solving the classification problem. The purpose of the feature selection is to reduce the number of irrelevant characteristics so that the classification accuracy remains acceptable. A good feature selection method can reduce the cost and increases the classification accuracy as well as efficiency (Khalid et al., 2014). The best subset includes the fewest features that have the most of collaboration in the accuracy. The remaining and immaterial features are ignored because many of them are either useless or have little information load. Deleting these features does not cause information problems but raises the computational load for the desired application as well as saving much useless information along with useful information (John et al., 1994). One of the problems related to categorizing is the high dimensionality of the feature space. Some of the features are irrelevant and redundant, and negatively affect the functionality of the classifier. Therefore, feature selection is required to reduce the feature space and increase the efficiency of classifier. A lot of work has been done in the field of feature selection in recent years. (Palanisamy & Kanmani, 2012) used the Artificial Bee Colony (ABC) algorithm for feature selection. The aim is to increase the speed of classification operations and extract useful information among the features. The extracted features have been evaluated by the j48 Decision Tree (DT) algorithm. (Nakamura et al., 2012) have proposed a new feature selection technique inspired by nature. The purpose was to solve the optimization of feature selection. In order to find the features that maximize the accuracy, the power of Bat exploration is combined with the speed of the Optimum-Path Forest (OPF) classifier. To evaluate the results, the algorithm was tested on 5 datasets. (Banati & Bajaj, 2011) have explained the problems related to high-dimensional, noisy and unrelated data, and the objective was to provide a new nature-inspired feature selection algorithm. In this paper, RST combined with the Firefly algorithm is used to find a subset of the features. The proposed hybrid algorithm was applied in the medical fields to find the minimum set of features. Four different medical databases were used as well as different methods for evaluating performance. (Forsati, 2012) proposed a new feature selection method using the ABC algorithm. The purpose was to formulate the feature selection as an optimization problem and provide a new feature selection process to achieve better classification results. In their study, the ABC algorithm was used to solve the feature selection problem. The proposed algorithm has experimented on 6 reference datasets. (Rodrigues et al., 2013) presented the finding a set of features to help the higher detection rates and faster extraction of features. The goal was providing a new feature selection based on binary Cuckoo search. In this paper, a binary version of Cuckoo's search for the purpose of the feature selection is presented. (Pritom et al., 2016) proposed a method to investigate the probability of breast cancer as well as the probability of recurrent breast cancer using various data mining techniques. Cancer patient data was collected from Wisconsin dataset of the UCI machine learning. Tis dataset contains 35 features, which are selected using the feature selection methods and are computed using classification algorithms. According to the results, Naive Bayes algorithm and decision tree provide better and higher accuracy. (Kamel et al., 2019) were used the GWOs and SVM for feature selection and data classification in order to increase the accuracy of breast cancer diagnosis, respectively. The best results were obtained from a hybrid of the SVM algorithm and the GWO to select the subset of the efficient features. (Ghaedi et al., 2022) were used a two-level stacking technique to detect power theft. To increase the correct detection rate of this technique, the CHOA algorithm was used to select the features of the base classifiers. The results of the proposed framework were showed that the efficiency of the proposed framework was higher than other studies. (Saranya & Pravin, 2021) focus on the use of an analysis of feature sensitivity to determine the optimum feature subset using Matlab with improved accuracy and sensitivity for classification. In comparison to the already well-known algorithms for wrapper selection, filter and embedded method, the effectiveness of the proposed algorithms is evaluated. The feature selection methods are more preferred than the feature extraction methods since they preserve the originality of the dataset. Based on this motivation, (Ceylan & Taşkın, 2021) were accordingly modified an evolutionary based optimization algorithm utilizing self-organization map to provide a new feature selection method for the classification of hyper-spectral images. (Baruah et al., 2020) attempt to introduce a PSO based feature selection method using mutual information (MI). Feature-class MI has been used to select a subset of features based on its relevancy. A wrapper-based method is used to find the productiveness of the method by evaluating with different classifiers in different datasets. (Arunadevi & Ganeshamoorthi, 2019) have concentrated on prediction of the breast cancer with few attributes. They have employed feature selection as the preprocessing step for the classification. They used three classifiers and two feature selection strategies for this paper. This work is mainly focused on using the minimal number of attributes for the prediction of cancer in order to reduce the data handling overhead.

Therefore, the main objective of previous works is to familiarize with the new algorithms used in the area of feature selection and introduce a more effective algorithm that improves the performance by focusing on their drawbacks. In order to obtain better accuracy for classification problems, an efficient feature selection process is required. In this study, the CHOA (Ghaedi et al., 2022) algorithm is applied to implement a feature selection problem, and also the SVM along with the one-versus-rest approach used as classifier.

This paper is organized as follows: Section 2 elucidates the CHOA and proposed algorithm. Section 3 clarifies the design of the proposed system. The results of experiments and their evaluations will be explained and discussed in section 4. Finally, the conclusions of the paper will be presented in section 5.

2. Cheetah optimization algorithm (CHOA)

This algorithm was presented by (Ghaedi et al., 2022). In this algorithm, cheetahs are expert hunters built for speed. They have small heads, long legs, and muscular tails to maintain balance. Their hearts and lungs are larger than usual to supply oxygen to the running muscles. Half of their muscle mass is placed around the spine so that they can be as flexible as a spring. This lengthens their strides, which helps them reach their maximum speed (Wilson et al., 2013). For cheetahs, only three strides are enough to reach a speed from zero to 65 km/h, and it only takes 3 seconds to reach a speed of 110 km/h. Cheetahs need coordinated senses to hunt. Although cheetahs are very fast, they must be close enough to their victim before attacking. When they see victim, they are completely focused (Van der Weyde et al., 2016). Teamwork increases the likelihood of success in hunting. When one gets tired, the other continues to chase. Male cheetahs usually hunt in groups (Farhadinia et al., 2012). The CHOA algorithm is a population-based meta-heuristic algorithm in which the location and velocity of the $object_i$ ($cheetah_i$ or $victim_i$) in the search space are specified as follows:

$$location_{object_i} = [location\_object_1^i, location\_object_2^i, \dots, location\_object_d^i]$$

$$velocity_{object_i} = [velocity\_object_1^i, velocity\_object_2^i, \dots, velocity\_object_d^i]$$

where d is the dimension of the search space.

Figure 1 shows the steps of the CHOA algorithm. All cheetahs choose a direction according to the movement of the victim, and the victim chooses its direction according to the $cheetah\ leader$.

Based on the current velocity of a cheetah and its distance from the victim, a new velocity is calculated for a cheetah. To find the best solution by $cheetah_{leader}$, firstly, all of the N cheetahs and M victims are initialized at random locations in the search space. Because the cheetahs are not far apart in a group, they are located in a range with a radius $R_{Hunting}$(10 meters). At the beginning of the movement (first step), the velocity of all cheetahs and victim is considered zero. In the next steps, the velocities and locations are updated. Figure 2 shows the vectors that determine the movement path of the cheetahs.

As shown in Figure 2, vector 1 represents the coefficient of the previous movement of the $cheetah_i$. Vector 2 represents the coefficient of the vector of the current location to the victim. Finally, based on the results of the two vectors, the new velocity of $cheetah_i$ is determined as vector 3.

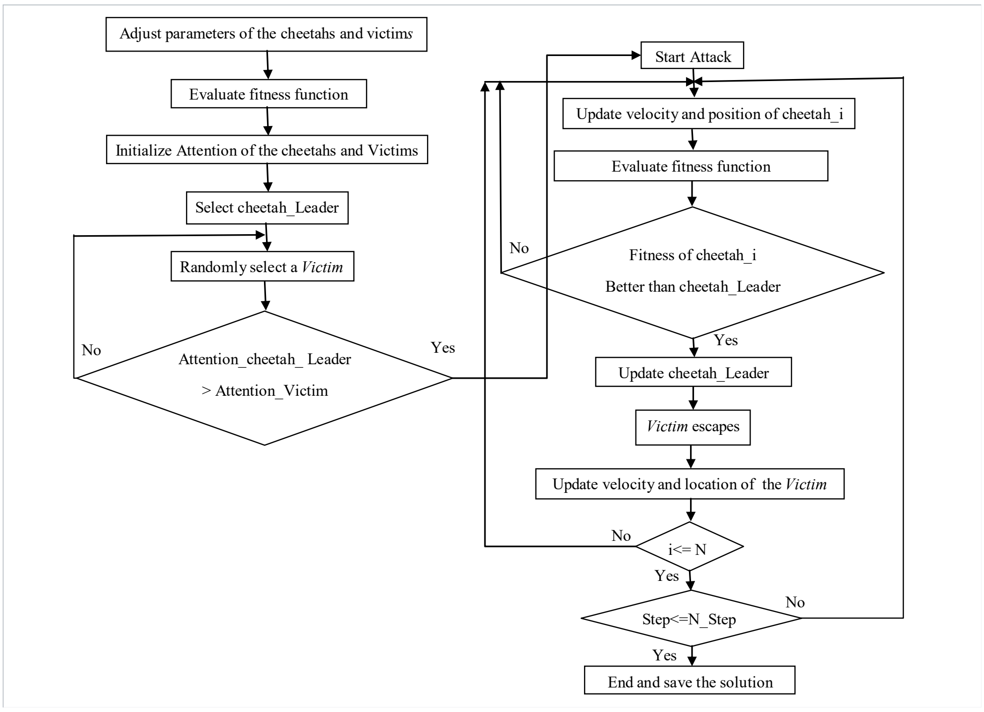

Figure 1: The steps of CHOA algorithm

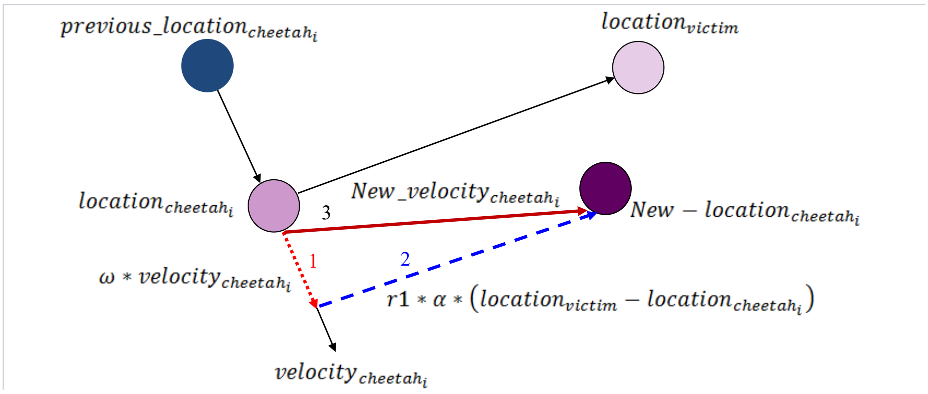

Figure 2: The direction of the cheetah movement

For each cheetah, the fitness of the new location is evaluated, and depending on the obtained value, the cheetah_(Leader) may be updated. If the cheetah_(Leader) improves, the victim will escape and its velocity and location will be updated.

Figure 3 also illustrates the movement pattern of the victim with respect to the changes in the movement of the cheetah_(leader).

As shown in Figure 3, vector 4 represents the coefficient of the previous movement of the victim. Vector 6 represents the coefficient of the vector of the current location to the cheetah_Leader. Finally, based on the result of the two vectors, the new velocity of the victim is determined as vector 6. To attack, the distance of cheetahs to the victim must reach a certain threshold. The best distance is between 30 to 100 meters. Therefore, the locations of the victims are randomly adjusted in the mentioned range. One of the important criteria for starting an attack is Attention.

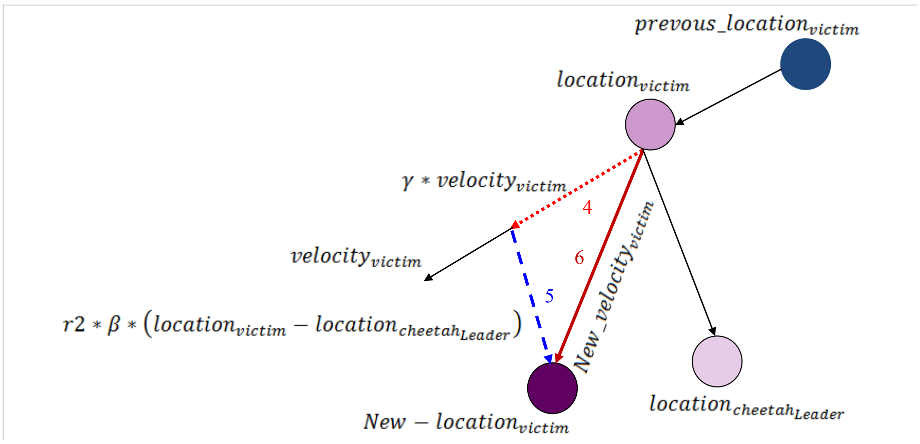

Figure 3: The direction of the victim movement

The Attention of each object (cheetah or victim) is calculated as follows:

$$\text{Attention\_(object\_i)} = \text{(Fitness(object\_i))/(\_i} \\ = \text{1\^{}N)MAX(Fitness(object\_i))} \qquad (1)$$

where $\text{Fitness}(\text{object}_i)$ is the fitness value of object_i and $({}_i = 1^N \text{MAX}\left(\text{Fitness}(\text{object}_i)\right))$

represents the highest amount of fitness among objects. In fact, values of the Attention are between 0 and 1. To attack, the cheetah_Leader must have more Attention than victim. The cheetah with the highest fitness is known as cheetah_Leader. The movement of each cheetah is affected by movement of the victim and is updated as follows:

$$\text{velocity\_(cheetah\_i)} \\ = 〚\omega * \text{velocity}〛 \text{\_(cheetah\_i)} + r1 * \alpha(〚\text{location\_victim} \\ - \text{location}〛 \text{\_(cheetah\_i))} \quad (2)$$

$$\text{New} - \text{location\_(cheetah\_i)} \\ = \text{location\_(cheetah\_i)} + \text{velocity\_(cheetah\_i)} \qquad (3)$$

where ω is the inertia weight. The weight of inertia controls the effect of past velocities on present velocities. In this study, the weight of inertia decreases linearly. The algorithm usually starts moving with a large amount of the weight inertia, which causes a large search space at the beginning of the algorithm, and this weight decreases over time, which causes the search within a small space at the final steps. In fact, inertia indicates how much the cheetah wants to maintain its current state of motion. A lower amount of inertia results in faster convergence of the algorithm, and increasing the amount of inertia increases the number of sudden movements of the cheetahs. velocity_(cheetah_i) represents the velocity of the cheetah_i. α indicates the influence of the location of the location_(victim) on the location of the cheetah_i. location_victim shows the location of the victim. Location_(cheetah_i) represents the current location of the cheetah_i. r1 is a random function with a uniform distribution between 0 and 1 that is used to increase random search and maintain the random nature of the algorithm. During the hunting process, if cheetah_(Leader) improves its fitness, it will not be able to catch the victim and the new location of the victim will be updated as follows:

$$\text{velocity\_victim} = \llbracket \gamma * \text{velocity} \rrbracket\ \_\text{victim} + r2 * \beta * \left(\text{location}_{\text{victim}} - \text{location}_{\text{cheetah\_Leader}}\right) \tag{4}$$

$$\llbracket \text{New} - \text{location} \rrbracket\ \_\text{victim} = \text{location\_victim} + \text{velocity\_victim} \tag{5}$$

where γ is the inertia weight, location_victim is the current location of the victim,. β is the change percentage in the fitness improvement of cheetah_Leader,. and r2 is a random function with a uniform distribution between 0 and 1. The fact that cheetah_Leader is selected based on the distance and attention factors prevents the algorithm from falling into the local optimization. A local optimum of an optimization problem is a solution that is optimal (either maximal or minimal) within a neighboring set of candidate solutions. The two most important factors in meta-heuristic algorithms are diversification and intensification. In the CHOA algorithm, with increasing the attention of the victim, the cheetah has to search a wider space, and diversification increases. Also, if the attention of the victim is low, the cheetah will hunt locally, which means that the intensification will increase.

3. System Design

This study developed a PSO approach, termed CHOA + SVM, for feature selection in the SVM. For the feature selection, if n features are required to decide which features are chosen, then n decision variables must be adopted. The value of n variables ranges between 0 and 1. If the value of a variable is less than or equal to 0.5, then its corresponding feature is not chosen. Conversely, if the value of a variable is greater than 0.5, then its corresponding feature is chosen. Figure 4 illustrates the solution representation.

Figure 5 shows the flowchart for CHOA + SVM. First, the population of particles is initialized, each cheetah having a random position within the D-dimensional space and a random velocity for each dimension. Second, each cheetah’s fitness for the SVM is evaluated. The each cheetah’s fitness in this study is the classification F-measure. If the fitness is better than the cheetah’s best fitness, then the position vector is saved for the cheetah. If the cheetah’s fitness is better than the global best fitness, then the position vector is saved for the global best. Finally the cheetah’s velocity and position are updated until the termination condition is satisfied.

F1	F2	…	Fn
1	0		1

Figure 4. Solution representation.

In order to provide a basic idea of the proposed approach performance, a comprehensive plan is presented on how to select features from the initial datasets, as well as the evaluation of these sub-features, as shown in Figure 5.

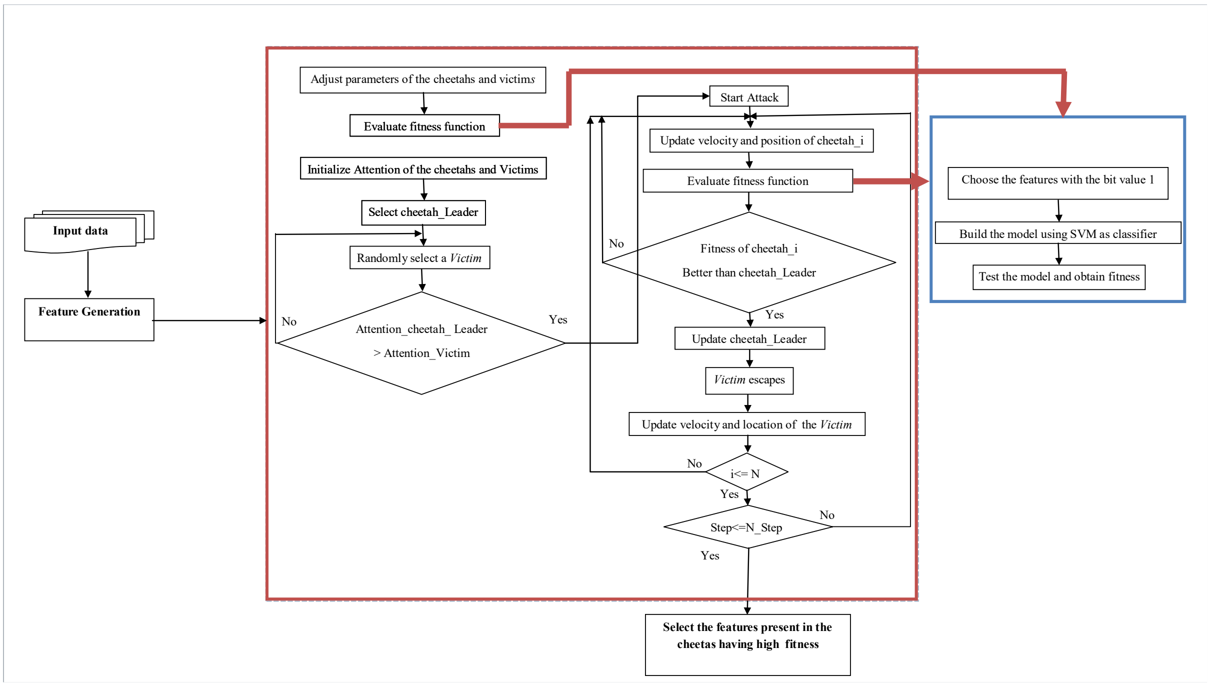

Figure 5: The Proposed Feature Selection based on CHOA

As shown in Figure 5, after feature generation, a subset of the features is selected by using the CHOA algorithm. Then, a model is constructed based on the selected features, and their performance criterions are evaluated by the SVM classifier along with the one-versus-rest approach as well as evaluation methods such as the K-flod Cross-Validation and Booting.

3.1. Feature Selection

The feature selection is used to identify a subset of features that maintain the predictive efficiency at an acceptable level. In order to compare the results of the CHOA algorithm with other algorithms, the common and well-known ACO and PSO algorithms are used in this study. All three algorithms are in the category of evolutionary algorithms inspired by the natural selection process.

3.2. Evaluation of Classifier

Two methods of K-Flod Cross-Validation and Bootsting (Kohavi, 1995) are used to evaluate the performance of the classifiers. The K-Flod method is used for problems with a low number of features, and the Bootsting method is also used in cases where the number of features is moderate or high. In the K-Flod method, the dataset is divided into k equal parts. The k-1 section is used as a training dataset to build the model and evaluation operation is carried out with one remainder part. The process will be repeated to k times. The evaluation of the final accuracy of the classifier is equal to K average of calculated accuracy. Unlike the K-Flod method, in the Bootsting method, a record has previously used in training phase, can be reused for training. That is, the training records are selected by sampling with inserting from the original datasets.

3.3. One-Versus-Rest method

In cases where the number of clusters is more than two clusters, finding the boundaries of classification is more complicated. One of the methods that makes binary classification possible for classifying multi-class problems is the one-versus-rest approach (Rocha & Goldenstein, 2014). The use of this method in multiclass classes such as decision trees and neural networks often reinforces classifier. In this paper, a SVM classifier is used along with the one-versus-rest approach to better evaluate the model.

3.4. Evaluation criteria

In this section, the most important criteria for evaluating the performance of classification are introduced. The concept of confusion matrix is described before examining a variety of classification criterions. This matrix classifies the algorithm as for the input dataset according to the type of problem clusters, and Table 1 presents a confusion matrix for the two-classes classification problem.

		Predicted Label	
		Positive	Negative
Actual label	Positive	True Positive (TP)	False negative (FN)
	Negative	False Positive (FP)	True negative (TN)

Table 1: Confusion Matrix for the two classes classification

In many articles, the accuracy criterion is used as a performance criterion for classification. However, this criterion solely is not an appropriate criterion for evaluating the performance. The reason for this is that in the accuracy relationship, the value of the records of different clusters is considered the same. Therefore, if in a specific application, the value of a cluster differs from that of another, then the correct or wrong prediction of the label of that cluster will have a different benefit or disadvantage relative to the records of the other clusters. Therefore, Precision, Recall, F-Measure and ROC criterions along with Accuracy (Hossin & Sulaiman, 2015) are used to measure performance, as shown in Table 2.

Criterion	Formula
Accuracy	(TP+TN)/(TP+TN+FN+FP)
Precision	TP/(TP+FP)
Recall/Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)
F-Measure	(2×Pre×Recall)/(Pre+Recall)
ROC Area	Cumulative distribution function between Sensitivity and 1- Sensitivity

Table 2: Performance criteria

Accuracy indicates the number of samples that are properly classified. Precision checks whether the number of correct examples is categorized, how many to the positive category. Sensitivity and Recall indicate that the number of positive samples is properly classified. The F-Measure is, in fact, an average between the accuracy and Recall parameters, and is used in cases where the importance of each Recall and Pre cannot be significant. The Specificity states how many negative samples are correctly classified. The ROC curve is derived from the cumulative distribution function between sensitivity and 1-specificity and in the modeling, the surface under the ROC curve is used to measure the accuracy of the model. If the ROC sub-curve is closer to 1 the model's accuracy at a good situation and closer to 0.5 is indicative of the low accuracy of the model and the inappropriate prediction of the model.

Therefore, by studying the CHOA algorithm, its use in feature selection and clustering customers, as well as the use of feature correlation, local search and hybrid structure of collective learning, a framework, is proposed that aims to increase the recognition rate of customers who theft electricity.

4. Adjustment of Experiments

In order to experiment the proposed method and compare the results with the results of other methods, adjusting the parameters and using reference datasets are required. In this section, these experiments are performed using reference datasets.

4.1. Dataset

To evaluate the efficiency of the CHOA + SVM combination method, Vowel, Wine, WDBC, Ionosphere, Sonar and Glass datasets are used. Table 3 shows the characteristics of these datasets. These datasets are available from the UCI data repository.

Dataset	Number of instances	Number of classes	Number of features
Vowel	528	11	10
Wine	178	3	13
WDBC	569	2	32
Ionosphere	351	2	34
Sonar	208	2	60
Glass	214	6	9

Table 3: The characteristics datasets

Attributes represent the number of table columns related to data. Instances represent the number of records in each table. The number of classes also indicates the number of classes and categories of each record in the table. If the number of features is between 10 and 19, the sample groups are considered as small, if the number of features is between 20 and 49, the sample groups are regarded as medium, and if the number of features is more than 50, the sample groups are as large. In cases where the number of features is small, the K-Flod Cross Validation method is used to evaluate the classifier, and if the number of features is moderate and large, the Bootstrap method is used.

4.2. Experimental Results and Evaluation

To select a classifier in combination with the CHOA algorithm, the initial experiment of the reference datasets without a selectable feature is firstly evaluated by the SVM, KNN and DT algorithms. Algorithm with better performance is selected as fitness function in the CHOA algorithm. First, without the feature selection, on the primary datasets, the result of the classification algorithms is calculated, as listed in Table 4.

No.	Dataset	Number of class	Number of features	Number of instances	Performance evaluation	Classification algorithm (%)		
						KNN	Decision Tree	SVM
1	Vowel	11	10	528	F-Measure(%)	77.40	81.90	83.25
					ROC-Area(%)	63.80	74.63	76.62
2	Wine	3	13	178	F-Measure(%)	88.15	88.69	97.87
					ROC-Area(%)	60.11	65.23	98.54
3	WDBC	2	32	569	F-Measure(%)	89.65	87.26	86.65
					ROC-Area(%)	76.69	78.36	71.13
4	Ionosphere	2	34	351	F-Measure(%)	84.45	92.23	91.12
					ROC-Area(%)	76.66	89.68	88.36
5	Sonar	2	60	208	F-Measure(%)	91.58	90.36	94.36
					ROC-Area(%)	80.06	86.41	88.87

Table 4: The values of evaluation criteria for algorithms without the feature selection

The method of evaluating the performance of the classifiers for the Wine and Vowel datasets is K-Flod method, and that of the WDBC, Ionosphere, and Sonar datasets is Boosting. The brown color indicates the best result of the F-Measure criterion for classification algorithms. Gray color represents the best result of the ROC criterion. According to Table 4, among the classification algorithms, in four datasets, the SVM algorithm shows the best result, so that the values of F-Measure and ROC curves for Vowel dataset are 83.25 and 76.62, respectively, for Wine dataset, respectively equivalent to 97.87 and 98.54, and for Sonar data are 94.36 and 88.87, respectively.

In the second experiment, all three ACO, PSO, and CHOA algorithms use the SVM classifier together with the one-versus-rest approach, which performed better at the previous stage compared to the two classifiers of KNN and Decision Tree. Table 5 presents the comparison results of three algorithms such as ACO, PSO and CHOA. As can be seen from Table 5, CHOA algorithm had the highest value for ROC curve in all datasets and the best value of F-Measure in five datasets. The results show that the CHOA algorithm reduces the number of selected features compared to other algorithms and, at the same time, increases the values of the F-Measure and ROC curves.

No.	Dataset	Number of features	Feature selection								
			ACO			PSO			CHOA		
			Number of features	F-Measure(%)	ROC area(%)	Number of features	F-Measure(%)	ROC area(%)	Number of features	F-Measure(%)	ROC area(%)
1	Vowel	10	8	78.63	80.06	7	80.11	84.32	4	98.23	97.36
2	Wine	13	10	90.18	89.97	8	98.17	98.65	2	100	100
3	WDBC	32	25	86.11	84.44	20	88.69	87.36	10	98.25	98.04
4	Ionosphere	34	30	89.32	90.45	20	91.69	94.74	14	98.87	97.38
5	Sonar	60	48	90.11	90.29	38	89.36	88.74	11	99.32	98.89

Table 5: Comparison of evaluation criteria for ACO, PSO and CHOA algorithms

In another experiment, the proposed method was compared with (Shojaee et al., 2021) research. Table 6 shows the results for Vowel, Glass and Wine datasets. According to Table 6, the CHOA-SVM has higher accuracy.

Dataset	Proposed method(CHOA)	(Shojaee et al., 2021)
Vowel	98.23%	91%
Glass	99.46%	98%
Wine	100%	93%

Table 6: Comparison of Prediction Accuracies of three datasets

In the last experiment, the performance of proposed method has been compared with (Zhang et al., 2020) work. The evaluation criterion is Accuracy. According to Table 7, the Accuracy of proposed method for all three datasets is higher than the (Zhang et al., 2020) work.

No.	Dataset	(Zhang et al., 2020)	Proposed method (CHOA)
1	Sonar	98.08%	99.34%
2	Wine	100%	100%
3	WDBC	97.92%	98.21%

Table 7: Comparison of the Accuracy of the proposed method with (Zhang et al., 2020)

According to the obtained results, it was proved that the proposed method classified the data with high efficiency.

5. Conclusion

This study used the CHOA algorithm to select the features so that accuracy was maintained at its highest level. Initially, without the feature selection, the F-Measure and ROC criterions were computed using the SVM, KNN, and DT classification patterns. The results revealed that The SVM classification had the best performance. Then, F-Measure and ROC criterions were measured using the feature selection algorithms of CHOA, ACO, and PSO. In order to evaluate the classification algorithm, the K-Flod method was used for small sample group testing problems and for the problems of testing medium and large groups, the Bootstrap method is used. The comparison of three ACO, PSO, and CHOA algorithms showed that the CHOA algorithm improved the F-Measure and ROC curves by decreasing the number of selected features. Therefore, the CHOA algorithm is suitable for the feature selection problem and can handle datasets with a high number of noisy and unrelated features. Nevertheless, the CHOA-based approach is not restricted to SVMs and can be easily extended to work in combination with other kernel methods, such as Gaussian process regression and classification. Future works include improving the convergence properties of CHOA on the feature selection problem, perhaps investigating an alternative selection strategy.

References

Arunadevi, J., & Ganeshamoorthi, K. (2019). Feature Selection Facilitated Classification For Breast Cancer Prediction. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 560–563. https://doi.org/10.1109/ICCMC.2019.8819752

Banati, H., & Bajaj, M. (2011). Fire Fly Based Feature Selection Approach. International Journal of Computer Science Issues, 8(4), 473–480.

Baruah, H. S., Thakur, J., Sarmah, S., & Hoque, N. (2020). A Feature Selection Method using PSO-MI. 2020 International Conference on Computational Performance Evaluation (ComPE), 280–284. https://doi.org/10.1109/ComPE49325.2020.9200034

Ceylan, O., & Taşkın, G. (2021). Feature Selection Using Self Organizing Map Oriented Evolutionary Approach. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 4003–4006. https://doi.org/10.1109/IGARSS47720.2021.9553491

Farhadinia, M. S., Hosseini-Zavarei, F., Nezami, B., Harati, H., Absalan, H., Fabiano, E., & Marker, L. (2012). Feeding ecology of the Asiatic cheetah Acinonyx jubatus venaticus in low prey habitats in northeastern Iran: Implications for effective conservation. Journal of Arid Environments, 87, 206–211. https://doi.org/https://doi.org/10.1016/j.jaridenv.2012.05.002

Forsati, R. (2012). A Novel Approach for Feature Selection based on the Bee Colony Optimization. International Journal of Computer Applications, 43(8), 30–34.

Ghaedi, H., Kamel Tabbakh Farizani, S. R., & Gaemi, R. (2022). A Novel Meta-heuristic Framework for Solving Power Theft Detection Problem: Cheetah Optimization Algorithm. International Journal of Industrial Electronics Control and Optimization, 5(1), 63–76. https://doi.org/10.22111/ieco.2022.39528.1370

Hossin, M., & Sulaiman, M. . (2015). A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant Features and the Subset Selection Problem (W. W. Cohen & H. B. T.-M. L. P. 1994 Hirsh (eds.); pp. 121–129). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-1-55860-335-6.50023-4

Kamel, S. R., YaghoubZadeh, R., & Kheirabadi, M. (2019). Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0247-7

Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. 2014 Science and Information Conference, 372–378. https://doi.org/10.1109/SAI.2014.6918213

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, 1137–1143.

Nakamura, R. Y. M., Pereira, L. A. M., Costa, K. A., Rodrigues, D., Papa, J. P., & Yang, X. S. (2012). BBA: A binary bat algorithm for feature selection. Brazilian Symposium of Computer Graphic and Image Processing, 291–297. https://doi.org/10.1109/SIBGRAPI.2012.47

Palanisamy, S., & Kanmani, S. (2012). Artificial Bee Colony Approach for Optimizing Feature Selection. International Journal of Computer Science Issues, 9(3), 432–438.

Pritom, A. I., Munshi, M. A. R., Sabab, S. A., & Shihab, S. (2016). Predicting breast cancer recurrence using effective classification and feature selection technique. 2016 19th International Conference on Computer and Information Technology (ICCIT), 310–314. https://doi.org/10.1109/ICCITECHN.2016.7860215

Rocha, A., & Goldenstein, S. K. (2014). Multiclass From Binary: Expanding One-Versus-All, One-Versus-One and ECOC-Based Approaches. IEEE Transactions on Neural Networks and Learning Systems, 25(2), 289–302. https://doi.org/10.1109/TNNLS.2013.2274735

Rodrigues, D., Souza, L. A. M. P. A. N., & Ramos, C. C. O. (2013). BCS : A Binary Cuckoo Search Algorithm for Feature Selection. 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013), 465–468.

Saranya, G., & Pravin, A. (2021). An Efficient Feature Selection Approach using Sensitivity Analysis for Machine Learning based Heart Desease Classification. 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), 539–542. https://doi.org/10.1109/CSNT51715.2021.9509673

Van der Weyde, L. K., Hubel, T. Y., Horgan, J., Shotton, J., McKenna, R., & Wilson, A. M. (2016). Movement patterns of cheetahs (Acinonyx jubatus) in farmlands in Botswana. Biology Open, 6(1), 118–124. https://doi.org/10.1242/bio.021055

Wilson, J. W., Mills, M. G. L., Wilson, R. P., Peters, G., Mills, M. E. J., Speakman, J. R., Durant, S. M., Bennett, N. C., Marks, N. J., & Scantlebury, M. (2013). Cheetahs, Acinonyx jubatus, balance turn capacity with pace when chasing prey. Biology Letters, 9(5), 20130620. https://doi.org/10.1098/rsbl.2013.0620

Zhang, Y., Zhao, Z., Zhao, S., Liu, Y., & He, K. (2020). A New Embedded Feature Selection Method using IBALO mixed with MRMR criteria. Journal of Physics: Conference Series, 1453(1). https://doi.org/10.1088/1742-6596/1453/1/012027