ISSN 2347 - 3983
Volume 8. No. 4, April 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter77842020.pdf
https://doi.org/10.30534/ijeter/2020/77842020

An Ensemble Feature Selection Model using Fast Convergence Ant Colony Optimization Algorithm

G. Suresh[1], Dr. R. Balasubramanian[2]
[1]Research Scholar, ManonmaniamSundaranar University, Tirunelveli, Tamilnadu, India, sureshg2233@yahoo.co.in
[2]Professor, Department of Computer Science and Engineering, ManonmaniamSundaranar University, Tirunelveli, India, rbalus662002@yahoo.com

ABSTRACT

Nowadays, several disciplines have to deal with big datasets that additionally comprise a high number of features. It gained more interest in several application domains like bioinformatics, medicine, marketing, or financial businesses, owing to massive collection of raw data that are stored. Feature selection (FS) is a process of choosing a minimal set of features from the actual set of features for optimal reduction in the feature space based on particular validation parameter. Since the dimensionality of a domain gets increased, the feature count N will also get increased. The process of identifying the optimal feature set is generally difficult and several issues relevant to FS have been shown to be a non-polynomial (NP) hard problem. This paper proposes a hybridization of ant colony optimization (ACO) with genetic algorithm (GA) for FS process. Since the ACO algorithm suffers from the drawback of slower convergence and improves the search space exploration process, GA is incorporated into the ACO algorithm. The application of GA in ACO algorithm for FS helps to fasten the convergence rate as well as improves the exploration capability. To assess the effective performance of the projected model, three different benchmark dataset namely chronic kidney disease (CKD), hepatitis and market dataset are utilized. The experimental results show the superior performance of the proposed model over the compared methods.

.

Key words: ACO, Big data, Classification, Feature Selection.

1. INTRODUCTION

At present times, Big data is an important research topic gains significant attention among academicians and research communications. In current digital era, massive quantity of data gets created and being stored at every second from various applications like healthcare, entertainment, telecom, education, and so on. At the same time, it poses several challenges where storing and processing large and varied datasets (known as big data) is not an easier task. So, feature selection (FS) process in Machine Learning (ML) has been treated as a superior technique to select the required features and thereby reduces the learning complexity.

This high-dimensional dataset comprises massive feature sets that do not lead to major complexity; however, there is the degradation in the performance of the trained methods. The main objective of FS is to reduce and improve the dataset by choosing the salient features. Usually, FS avoids indefinite features from actual database with no efficient operation. Practical issue deal with FS is because of these existing aspects namely maximum noise, inaccurate data, insignificant and unnecessary features in raw feature set. Hence, FS is declared as an active region of research work that spreads in entire domain along with pattern examination, data mining, image mining ,text classification and so on [1].

Several techniques have been presented for FS categorization namely wrapper, filter, and hybrid methods. Wrapper model consists of predefined learning technique from which the features are chosen for justifying the learning operations of specific training method. In filter technique, statistical determination is required for feature set in order to apply learning approach. Graphical representations indicated the working procedure of wrapper and filter approaches to identify salient features.

Followed by, hybrid model tends to employ strengths of wrapper as well as filter techniques. Here, subsets are produced and searching operation is performed in many ways. Initially, Sequential Forward Search (SFS) have been applied for initiating the searching process using unfilled feature set and include them efficiently. An alternate option is termed as Sequential Backward Search (SBS) which is utilized with the application of complete set as well as elimination of features effectively. Additionally, third model known as bidirectional selection finds helpful in starting at parallel ends, also performs both adding and removing operations of features at same time. A final technique [2] acquired is to begin the search operation with the random selection of subsets with the application of sequential or bidirectional principles. But, an alternate search algorithm is obtained named as complete search, that offers optimal solution for FS job since entire searching is not possible in case of large features. Besides, the sequential strategy is easy and rapid for implementation; however, it is influenced by nesting effect whereas it could be either added or deleted and vice versa.

To resolve these demerits of sequential search strategy, alternate searching process has been introduced which is termed as floating search model. Many types of predicting schemes attempts in finding solutions for FS which ranges from sub-optimal to adjacent optimal regions due to the application of local search model to the entire process in spite of global search. Alternatively, the search techniques employ a partial search method over other models and suffer from complex computation process. Finally, near-optimal to optimal solutions could not be attained easily with the help of these techniques. Consequently, several research works have been concentrated on global search models.

Global search technique plays a significant role in discovering solutions for complete search space based on the function of multiple agents which applies global search employing local search method, hence, the capability of finding high qualified solutions are obtained with small interval of duration. In order to obtain global search researchers have simulated annealing, genetic algorithm, Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) algorithms to resolve FS operations. Though various optimization algorithms were present, in this paper, ACO algorithm is utilized for FS.

Previous FS techniques have been grouped into 3 modules: wrapper, filter, and hybrid models. The wrapper technique is a familiar one, which offers features with maximum salient property when compared to filter methods. It is due to the fact that the features are used collectively and it is expensive in terms of processing. Additionally, various studies regarding FS have been mentioned [3]. Based on the solution obtained from FS, filter model is assumed to be rapid for execution because it evaluates the operation of features in the absence of original technique among inputs as well as output of information. A feature could be chosen or removed in accordance with predefined procedure like mutual data [4], authorized unit analysis, separate component determination, independent class estimation, or parameter range. Filter techniques [5]consist of few merits such as performance efficiency; therefore, saliency of chosen features is not enough due to non-biasing of classification methods. Wrapper models comprise of various number of techniques [6]projected a set of sequential search strategies to discover subset of salient features. In [7], features have been included to Neural Network (NN) on the basis of SFS while learning operation is carried out.

In recent times, many techniques have been presented [8], which tend to concentrate in SFS-based FS. These models correlate features from 2 groups, such as, identical and non-identical, that is induced to the NN training method consecutively. The final stage of operation is NN classification model which holds essential data from provided databases, subset of salient features is produced along with decreased unwanted data. [9] presented an ACO based FS and classification process on the application of bankruptcy. In [10], the significance of the Master River Multiple Creeks Intelligent Water Drops (MRMC-IWD) is presented by the use of real-world optimization problems relevant to FS and classification process is validated. In [11], two wrapper FS models utilizing Salp Swarm Algorithm are presented. The crossover operator is applied along with the transfer functions for improvising the technique. In [12], a novel meta-heuristic optimization model called as chaotic crow search algorithm (CCSA) is presented for resolving these issues. The presented model is employed for optimizing FS process on 20 benchmark datasets. In [13], whale optimization method based on wrapper feature selection.

From alternate studies [14], SBS is added into FS with the application of NN where the lower salient features are removed sequentially during the training process. Diverse techniques have been applied for various heuristic models for calculating the saliency from features. [15], evaluated the saliency of features with the help of NN training approach where single feature is deployed in input layer simultaneously. In order to determine the saliency of a feature, two diverse weights analysis-based heuristic methods have been applied [7]. Therefore, [14]introduced a complete feature set for NN training scheme which is employed for all feature that is removed for particular time using a cross-check of NN operation.

This research proposes a hybrid ACO with genetic algorithm for FS process. Since the ACO algorithm suffers from the drawback of slower convergence and improves the search space exploration process, GA is incorporated into the ACO algorithm. The application of GA in ACO algorithm for FS helps to fasten the convergence rate as well as improves the exploration capability. To assess the effective performance of the projected model, three different benchmark dataset namely chronic kidney disease (CKD), hepatitis and market dataset are utilized. The experimental results show the superior performance of the proposed model over the compared methods.

2. PROPOSED ACOGA-FS MODEL

The main objective of FS is to eliminate the problem of computing sizes of subsets; ACOGA-FS utilizes a model to determine the subset size. This system leads the ants for making subsets in a decreased type. This technique employed ants for constructing separate subset as depicted in Figure 1.

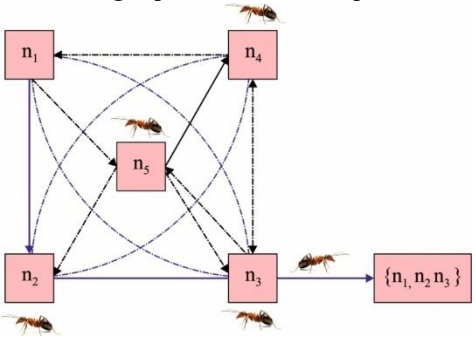

Figure 1: ACO-FS based subset generation

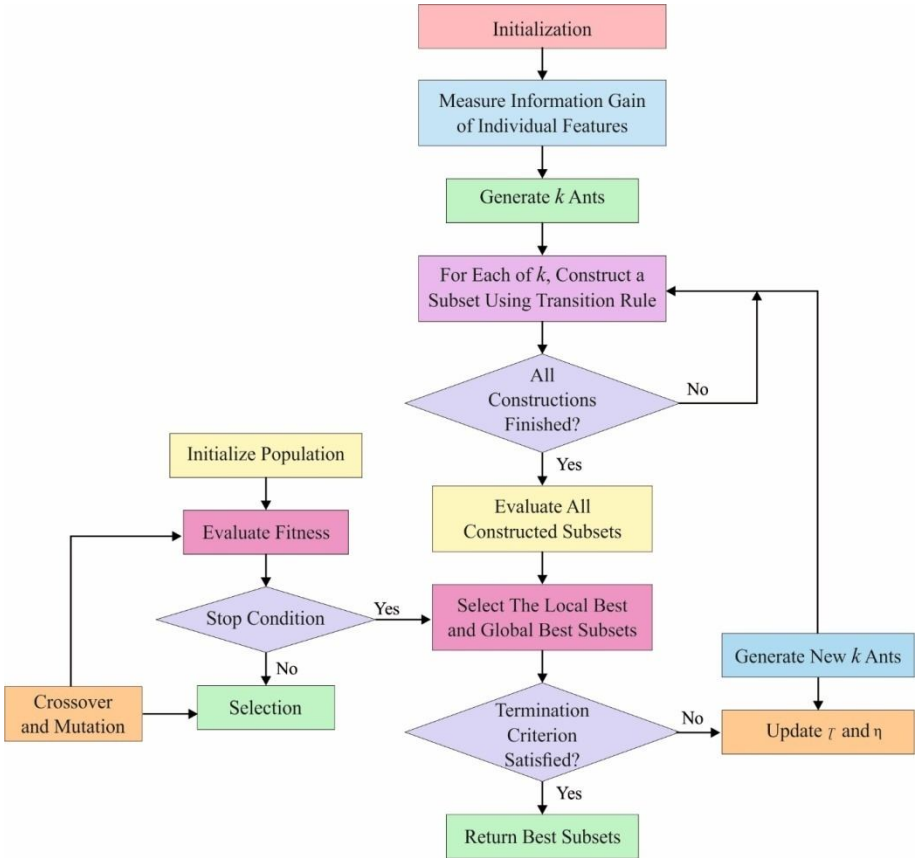

Figure 2: Flowchart of ACOGA-FS

Conversely, the limitation of computing subset size should not be considered as a major issue for determining its size, since, it could be monitored after a definite radius, extended boundary for bounding system leads to poor results of FS. Since the ACO algorithm suffers from the drawback of slower convergence and improves the search space exploration process, GA is incorporated into the ACO algorithm. The application of GA in ACO algorithm for FS helps to fasten the convergence rate as well as improves the exploration capability. The proposed model includes a hybrid exploration approach (i.e., a group of the wrapper as well as extract manners) with planning various principles for the global exploration capability of the ants. The integration of 2 algorithms in ACOGA-FS obtains maximum value of FS from a provided dataset. Here, $\mathrm{t}_1, \mathrm{t}_2, \cdots, t_5$ signifies the separate features. As an instance, one ant located in t_1 created one subset $\{\mathrm{t}_1, \mathrm{t}_2, t_3\}$. The phases of ACOGA-FS can be defined using a flowchart as revealed in Figure 2 and are explained here.

- Let T be a feature group of a provided data set D containing d separate classes, C_c $(forc = 1, 2, \dots, d)$. Assume t is the sum number of features of T. First the pheromone paths τ while heuristic data η of all t features with transfer equivalent rates of τ and η.

- Compute the data, form separate n features utilizing data gain measurement system. In this effort, we utilize these data gain assets of features as an extracting device in planning principles to give hybrid exploration during FS method.

- Create a group of artificial k ants equal to t, namely, $\mathrm{k} = t$.

- Choose the subsets size r to every k ant based on the subsets size calculation system. Behind, pursue the convention probabilistic changeover principle to choose features for creating the subsets as pursues:

$$P_x^{\mathrm{k}}(\mathrm{z}) = \begin{cases} \dfrac{[\tau_{\mathrm{x}}(\mathrm{z})]^{\alpha}[\eta_{\mathrm{x}}(\mathrm{z})]^{\beta}}{\sum_{\mathrm{u}\in y^{\mathrm{k}}}[\tau_{\mathrm{u}}(\mathrm{z})]^{\alpha}[\eta_{\mathrm{u}}(\mathrm{z})]^{\beta}} & \text{if } x \in y^{k} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where y^{k} is the group of possible features which could be extra for the restricted result, τ_{x} and η_{x} are the pheromone as well as heuristic values related by features $x(x = 1, 2, \dots, t)$, while α and β are 2 attributes which resolve the corresponding value of pheromone as well as heuristic data. Noticeable, because the first value of τ and $\eta$to every separate feature is equal, Eq. (1) illustrates arbitrary performance in SC initially.

- When the creations of subset have been done with every ant, after that maintain; or else, continue to Step (4).

- Calculate the subsets $SS^{k}(z)$ according to the subset estimation system as well as compute the classifier action $\left(SS^{k}(z)\right)$. Now, $SS^{k}(z)$ mentioned the subsets created with k ants at iteration z.

- Choose the local optimal subsets, $SS^{loc}(z)$ with every $SS^{k}(z)$ and the global optimal subset, SS^{glb} with every $SS^{loc}(z)$ in accordance by chosen system. Now, $z = 1, 2, 3, \cdots, ITR$ where ITR are amount of iterations.

- Verify whether $SS^{loc}(z)$ attains existing accuracy, or technique implements an iteration threshold,ITR_{th}, then conclude the FS method. Exactly, ITR_{th} mentioned definite number of iterations that the technique could not detect some further alteration of SS^{glb}. But, if the execution criterions are fulfilled, next SS^{glb} is, consequently maintained as a result of optimal subset. Accumulate the actions of every local optimal best subset$\gamma\left(SS^{loc}(z)\right)$ to additionally utilize it.

- Inform the values of τ and η to every feature based on principles of pheromone information as well as heuristic data quantity, correspondingly.

- Create a novel group of artificial k ants and progresses it.

Currently understandable that the scheme following ACOGA-FS are directly, i.e., showing the ants utilizing an enclosed system while giving hybrid method to the ants explore. For attaining the efficient exploration process, a data gain calculation method has been included that does not need classy calculation and executed only once during the FS procedure. For optimal consideration, every feature of ACOGA-FS is presented provided with the subsequent segments.

2.1 Determination of subset size

In an ACO technique, the action of ant's gains importance to solve various combinatorial optimized problems. Consequently, in resolving the FS problem, leading ants in the extracted paths offers more benefits. The ACOGA-FS technique utilizes a directly mechanism for resolving the subsets size$r$. It utilizes an easier probabilistic procedure by a restriction as well as arbitrary operation. The plan of utilizing such a probabilistic method is to give data for the arbitrary operation in such methods which reduce the subset sizes which holds a maximum possibility of chosen individuals. It is essential in the sense that ACOGA-FS could led to an exacting way with the option of decreased size subsets of leading features. The subsets size calculation system can be utilized could be defined in 2 directions as pursues. Initially, ACOGA-FS utilizes a probabilistic method altered for choosing the size of a subsets $r(\leq t)$ as pursues:

$$P_r = \frac{t - r}{\sum_{x=1}^{l}(t - x)} \quad (2)$$

where, P_r denotes increased linearly as r is decreased, r is limited with a constraint, such as, $2 \leq r \leq \delta$. So, $r = 2, 3, \cdots, \delta$, where $\delta = \mu \times t$ and$loc = t - r$. Now, μ is a user defined attribute that managesδ. Its rate is based upon t to provide datasets. If $\delta$is nearby$t$, next the exploration space describes the relevant features develops as a maximum calculation price, while increasing danger of inefficient feature subsets might be created. Because the goal of the presented ACOGA-FS is to choose subsets of leading features in a lesser series, we desire the length of the chosen subsets to be among 3 and 12 based upon the provided dataset. Thus, μ is set as$\mu \in [0.1, 0.6]$. It is standardized for all values of P_r in such a direction that outlines every probable rate in P_r is equal to 1.Followed by, ACOGA-FS uses every value of P_r to arbitrary selection model for calculating the size of the subsets, r finally. This selection technique is identical to roulette wheel method.

2.2 Subset Evaluation

Subset estimating subsets play a vital role together with further basic functions of ACO to choose relevant features in FS operations. Generally, wrapper methods are applied in the estimation of the tasks, whereas wrapper model is more efficient than filter approach. Usually, discovered relevant subsets with minimal size are usually desirable because of the small rate in hardware execution. Different other predefining techniques the optimal relevant feature subsets are identified finally as a group of the local optimal as well as global optimal features are chosen.

A. Local best selection

Resolve the local optimal subset, $SS^{loc}(z)$ to appropriate $z(z \in 1,2,3,\ldots)$ iteration based on $Max\left(\gamma\left(SS^{k}(z)\right)\right)$where $SS^{k}(z)$ is the number of subsets created with k ants, while $k = 1, 2, \cdots, t$.

B. Global best selection

Solve the global optimal subset (SS^{glb}), the optimal subset of relevant features from every local optimal results in such a way that SS^{glb} is evaluated by presently determined local optimal subset, $SS^{loc}(z)$ at each z iteration with their classifier actions. When $SS^{loc}(z)$ is created as optimal, next $SS^{loc}(z)$ is returned withSS^{glb}. While any action is established related to several time, afterward choose the one along with the 2, i.e., SS^{glb} while $SS^{loc}(z)$ as an optimal subset that consist of decreased size. To be pointed that, the initial iteration $SS^{loc}(z)$ is measured as SS^{glb} to determine the global best ones, GA is executed.

C. GA based global best selection

GA is considered as a most popular model applied to resolve optimization issues. Chromosomes are the basic population as well as candidate solution. There are 3 processes present in GA namely mutation, crossover and selection operators. Meta-heuristic techniques are arbitrary in nature; thus, it is complex for solving optimal solution in search space. When ACO technique is merged with GA, hybrid model leads to update the GA local search process. Hence, GA declares the search space as global search whereas ACO technique is named as local search. Therefore, the population of GA is generalized on the basis of ACO method. This pattern makes the ACO model of faster convergence for obtaining optimized result. Few steps followed in algorithm are:

i. Choose the primary population, fix the counter value as equal to zero (g=0).

ii. Try to manage the counter value. If condition is attained, then follow step 8.

iii. The objective functions are to be computed for all solution candidates according to the fitness value achieved.

iv. An optimal population is accomplished in present generation which is termed as elegant solution (P_e indicates population). The elite solutions are duplicated to apply for upcoming generation.

v. From this stage, many solution candidates are extracted from residual population P_c that is selected. Hence crossover operator is referred as integration or combination pair.

vi. Remaining population is assumed, in order to apply the mutation operator to solution candidates.

vii. Extra unit is added to the counter and follow step 2.

viii. Acquire optimal attributes to ACO technique as well as choose number of solutions in this model.

ix. Counter number is managed when condition is satisfies and the algorithm is completed.

x. Motion computation: Determine the movement of population induced by alternate individuals; similar motion; external diffusing process.

xi. Upgrade the unique location of ACO in search space.

xii. Fix g = g + 1; and get back to step 9.

3. EXPERIMENTAL VALIDATION

To ensure the effective characteristics of the presented model, a set of three-benchmark dataset is utilized and the details are provided below.

3.1. Dataset used

For assessing the efficient performance of the ACOGA-FS technique, CKD [16] dataset is utilized. It is shown that the test CKD dataset has a totality of 400 instances with the presence of 24 attributes. Additionally, the instances are classified based on two classes namely absent and present or CKD/non-CKD. Moreover, the test hepatitis dataset has a total of 155 instances with the presence of 19 attributes. Besides, the instances are classified with respect to two classes absent and present. Afterwards, the applied marketing dataset has a total of 8933 instances with the presence of 9 attributes.

3.2. Results Analysis

Table 1 shows the FS results of the ACOGA-FS and GA models on the CKD dataset. The table values indicated that a total of 16 features were chosen by ACOGA-FS algorithm and a total of 17 features were chosen by GA. The presented ACOGA-FS algorithm offers FS results with the least good cost of 0.00049 whereas the GA model offers poor outcome by attaining higher best cost of 0.003042. This higher value indicates the ineffective characteristics of the presented model on the applied CKD dataset.

Table 1: FS Results on CKD dataset

Iteration count	Total No. of Features	Proposed-ACOGA-FS		
		Selected Features	Cost	% of Features Reduced
1	24	16	0.00092	33
2	24	16	0.00092	33
3	24	16	0.00092	33
4	24	16	0.00062	33
5	24	16	0.00049	33

Table 2: Comparative analysis of existing FS with proposed method for CKD Dataset

Methods	Best Cost	Selected Features
ACOGA-FS	0.00049	14,15,16,19, 5,20,13,24,23, 6,22, 7,17,12, 3, 2
GA-FS	0.003042	20,24,23, 6,15, 3, 5,12,18, 8,17, 4,10,19, 7,22,16
PCA	0.706100	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17, 18,19,20
CFS	0.748000	1,2,4,6,9,10,11,12,14,16,17,18,19,20,21,22,23

Table 2 shows the experimental outcome of different FS models on the tested CKD dataset. It is shown that the CFS method offered poor outcome with the highest best cost of

0.748000 whereas slightly lower best cost of 0.748000 is exhibited by PCA algorithm. At the same time, the presented ACOGA-FS model offers superior FS results by attaining the feature subset with the minimal cost of 0.00049.

Table 3: Accuracy Comparisons of Before and After FS Methods for CKD Dataset

Classifiers	Before FS	After FS
FNC	98.34	98.87
DT	98.96	99.15
LR	99.32	99.56
MLP	98.97	99.78
RBFNetwork	98.37	99.34
OlexGA	99.12	99.54
RMSProp+MLP	**99.46**	**99.93**

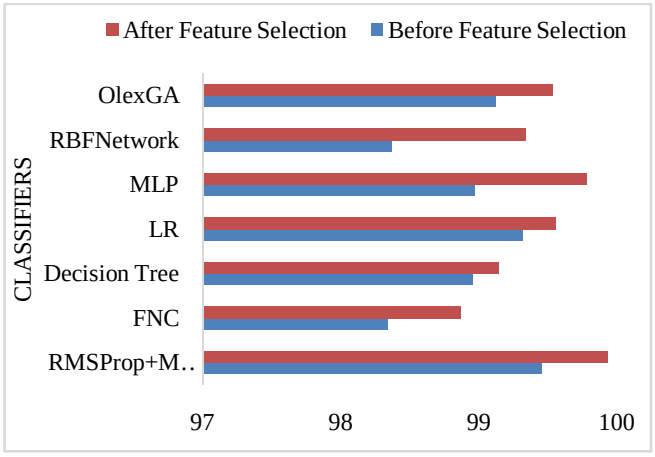

Figure 3: Accuracy of Before and After FS on Various Classifiers for CKD Dataset

Table 3 and Figure3 made a comparative study of different classifiers with the use of presented ACOGA-FS technique. It is shown that the RMSProp with MLP achieves a minimum of 99.64% before FS and it is raised to 99.93% after FS. At the same time, it is noted that Olex GA shows a minimum of 99.12% before FS and it is raised to 99.54% after FS. In the same way, the FCN classifier obtains an accuracy of 98.34 before FS and it is increased upto 98.87% after FS. Besides, the DT model offers a minimum of 98.96% accuracy before FS and it is improved to 99.15% after FS.

Additionally, the LR model offers a minimum of 99.32% before FS and it got raised to 99.56% after FS. Moreover, the MLP model shows significant results by the use of FS where the accuracy attained before FS is 98.97% and it is increased to 99.78% by the use of FS. Furthermore, the OlexGA model offers an accuracy of 98.37% before FS and is increased to 99.34% after FS on the applied CKD dataset. These values pointed out that the accuracy is significantly improved by the use of FS process.

4. CONCLUSION

This paper has presented a hybrid ACOGA-FS algorithm FS process. Here, GA is incorporated to the ACO algorithm to resolve the limitations of slower convergence and improves the search space exploration process, GA is incorporated into the ACO algorithm. The application of GA in ACO algorithm for FS helps to fasten the convergence rate as well as improves the exploration capability. An extensive experimental analysis takes place on three-benchmark dataset. The simulation outcome strongly pointed out the superior classification results by the use of FS process. In future, the performance of the presented model can be enhanced by the use of missing value replacement techniques.

REFERENCES

1. Chandra Sekhar Reddy, N. Purna Chandra Rao Vemuri, Govardhan, A. **An Emperical Study on Support Vector Machines for Intrusion Detection,** *International Journal of Emerging Trends in Engineering Research*, Vol. 7, No. 10, pp. 383-387, October 2019. https://doi.org/10.30534/ijeter/2019/037102019
2. Rufo I. Marasigan Jr, Alvin Sarraga Alon, Mon Arjay F. Malbog, Joshua S. Gulmatico, **Copra Meat Classification using Convolutional Neural Network,***International Journal of Emerging Trends in Engineering Research*, Vol. 8. No. 2, February 2020 https://doi.org/10.30534/ijeter/2020/30822020
3. Subramaniyam, S,**Performance Analysis on Diesel Engine using Neem and Soya Bean Oil,***International Journal of Emerging Technologies in Engineering Research (IJETER)*, Vol. 5, Issue 8, August 2017.
4. F. Ros, and S. Guillaume. **From Supervised Instance and Feature Selection Algorithms to Dual Selection: A Review.** In Sampling Techniques for Supervised or Unsupervised Tasks, Springer, Cham, pp. 83-128, 2020. https://doi.org/10.1007/978-3-030-29349-9_4
5. A.K. Das, S. Das, and A. Ghosh. **Ensemble feature selection using bi-objective genetic algorithm,** *Knowledge-Based Systems*, Vol. *123*, pp. 116-127, 2017.
6. S. R.Ahmad, A. A. Bakar, and M. R. Yaakub. **A review of feature selection techniques in sentiment analysis,** *Intelligent Data Analysis*, Vol. 23, No. 1, pp. 159-189, 2019. https://doi.org/10.3233/IDA-173763
7. M. M.Sakr, M. A. Tawfeeq, and A. B. El-Sisi. **Filter Versus Wrapper Feature Selection for Network Intrusion Detection System,** In *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 209-214, 2019
8. S.Guan, J. Liu, and Y. Qi. **An incremental approach to contribution-based feature selection.***Journal of Intelligence Systems*, Vol. 13, No. 1, 2004.
9. Y.B.Wah, N. Ibrahim, H.A. Hamid, S. Abdul-Rahman, and S.Fong. **Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximising Classification Accuracy,** *Pertanika Journal of Science & Technology*, Vol. 26, No. 1, 2018.

10. J. Uthayakumar, N. Metawa, K. Shankar, and S.K. Lakshmanaprabu, **Financial crisis prediction model using ant colony optimization,** *International Journal of Information Management,* 2018.
11. B.O.Alijla, C.P. Lim, L.P.Wong, A.T. Khader, and M.A.Al-Betar. **An ensemble of intelligent water drop algorithm for feature selection optimization problem.** *Applied Soft Computing*, No. 65, pp.531-541, 2018. https://doi.org/10.1016/j.asoc.2018.02.003
12. H. Faris, M.M. Mafarja, A.A. Heidari, I. Aljarah, A.Z. Ala'M, S. Mirjalili, and H. Fujita. **An efficient binary salp swarm algorithm with crossover scheme for feature selection problems**, *Knowledge-Based Systems*, Vol. 154, pp.43-67, 2018. https://doi.org/10.1016/j.knosys.2018.05.009
13. G.I.Sayed, A.E. Hassanien, and A.T.Azar. **Feature selection via a novel chaotic crow search algorithm,** *Neural Computing and Applications*, Vol. 31, No. 1, pp.171-188, 2019. https://doi.org/10.1007/s00521-017-2988-6
14. M. Mafarja, and S.Mirjalili. **Whale optimization approaches for wrapper feature selection,** *Applied Soft Computing*, Vol. 62, pp.441-453, 2018. https://doi.org/10.1016/j.asoc.2017.11.006
15. E.Gasca, J. S.Sanchez, and R. Alonso. **Eliminating redundancy and irrelevance using a new MLP-based feature selection method,***Pattern Recognition,* Vol. 39, pp. 313–315, 2006. https://doi.org/10.1016/j.patcog.2005.09.002
16. S. Abe. **Modified backward feature selection by cross validation.***In Proceedings of the European symposium on artificial neural networks,* 2015, pp. 163– 168.
17. **Chronic Kidney Disease Dataset,** available at https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_ Disease