Evaluation Feature Selection Technique on Classification by Using Evolutionary ELM Wrapper Method with Features Priorities

Methaq Kadhum, Saher Manaseer, and Abdel Latif Abu Dalhoum
Computer Science Department, The University of Jordan, Amman, Jordan
Email: methaq_kadhum@yahoo.com, {saher, a.latif}@ju.edu.jo

***Abstract*—Features' selection is a dimension reduction technique that aims to enhance classification accuracy by removing unrelated and redundant features. The Wrapper approach, one features selection strategy, provides an accurate estimation of classification performance. In view of this, we propose a new model of Evolutionary Wrapper Feature selection. This model exploits Extreme Learning Machines (ELM) to evaluate selected subsets, comprising a Genetic Algorithm (GA) as a search algorithm to find a set of feature subsets. A *priority* was assigned to each feature when GA had explored the space of feature combinations. The use of priority avoids replacing one feature with another of higher priority. The goal of this model is to investigate the accuracy rate of using feature selection methods and the impact of using priority with the features. Two machine learning classifiers are considered: the ELM and the Support Vector Machine (SVM). The proposed model is piloted based on a Chronic Kidney Disease dataset (CKD) from UCI. Experimental results indicate that the proposed model can achieve a better accuracy rate with these two classifiers. In addition, it requires much less time to find the best subset of features.**

***Index Terms*—extreme learning machine, evolutionary, wrapper feature selection, genetic algorithm, features priority**

I. Introduction

Classification techniques employ supervised learning algorithms such as Support Vector Machines (SVMs), decision trees and Extreme Learning Machines (ELMs) [1] to classify an instance into classes by studying its features. Supervised learning algorithms learn the classification model by using training data to extract the relationship between each category and its features. Therefore, the classification model can predict the class of new instances by their features [1]. Consequently, the (classification) accuracy of the classification model relies on the ability to recognize the features as relevant features, irrelevant or redundant ones [2], [3]. Therefore, feature selection algorithms have been implemented to select a subset of the relevant features and remove irrelevant and redundant ones [4]-[6].

Manuscript received June 16, 2020; revised November 17, 2020.

Feature selection is a Dimensionality reduction approach, that can be used to obtain relevant features and remove irrelevant and redundant ones based on a defined criteria, enabling learning algorithms to operate faster and more effectively [7], [8]. Feature selection aims to improve classification performance by discarding irrelevant and redundant features, as well as reducing storage resources, improving time efficiency, and reducing the complexity of the prediction model, [7]. Feature selection methods are becoming crucial techniques in many applications, such as medical diagnosis [9] and gene marker recognition [10], etc.

Feature selection algorithms are developed in two basic ways. First of all, one must decide the search strategy as over the feature subset space, the number of possible subsets grows exponentially with the number of features. However, for a dataset with N features, there are (2^N-1) possible non-empty feature subsets. Therefore, many optimization techniques are used such as Sequential Backward selection, Branch-and-Bound, Best-First search, and Genetic Algorithms [4]. Secondly, one must decide on an evaluation strategy, as it is complicated to predict the impact of using a feature on classification performance. Consequently, the two main evaluation strategies in which feature selection is applied are the Filter and Wrapper methods [2], [3]. In comparison with the filter model, the wrapper method provides a more accurate evaluation of classification performance [11].

The impact of using feature selection is usually evaluated by means of the classification accuracy of the machine learning model, such as the use of decision trees, Support Vector Machines (SVM), K-means, and Forward Neural Networks (SLFN) [12], [1]. However, several studies focus on the impact of using feature selection in a different area of research. One of the studies is the SVM-RFE (Support Vector Machine-Recursive Feature Elimination) method, which seeks to apply the Wrapper method that uses Support Vector Machines to measure the performance of features and construct a classifier with high-performance scores [13]. In [9] Evolutionary Extreme Learning Machines (ELM) [14] were combined with the Wrapper feature selection method [7] for Alzheimer's disease on anatomical brain MRI. H. Polat *et al.* presented the SVM classification algorithm with

doi: 10.12720/jait.12.1.21-28

Wrapper and filter strategies that were used to improve the accuracy of classification on the CKD data set and gain a rate of 98%. In this paper, we propose the use of feature selection algorithms with ELM to evaluate Chronic Kidney Disease prediction. [15]

ELM is a Single Hidden Layer Forward Neural network (SLFN) [10] algorithm, that has a fast learning speed and good accuracy compared to existing artificial neural network algorithms [16]. It has been applied as a classifier in many applications, such as: symbols, documents and pattern recognition [17]-[19], image processing [20] and cancer detection [21]. Furthermore, it plays an essential role in reducing dimensionality [22].

In addition, the Genetic Algorithm (GA) has gained a significant amount of attention from researchers, and is used with feature selection algorithms as a search algorithm, to find a set of feature subsets [23]. The genetic algorithm is a random search strategy that is often applied in wrapper models [24], [25].

In this paper, we propose a new approach to the Evolutionary-Wrapper Feature selection method; it is based on GA as a search algorithm (GWFP), but differs insofar as it does not select a feature subset randomly. GA selects and replaces features of the populations based on priorities that were assigned to the features according to their importance; (the concept of) Extreme Learning Machines (ELM) was used as an evaluation algorithm. The motivation for this work is as follows: first, we want to reduce feature space. Second, we wish to investigate the increase the accuracy rate by using feature selection methods. Third, we seek to explore the effects of using the feature priority value on accuracy classification

The model that is developed in this work will be referred to as ELM-GWPF. It will be evaluated based on the Chronic Kidney Disease dataset.

In addition, the performance of ELM-GWPF will be evaluated according to the ELM classifier and SVM classifier. Moreover, the priority of attributes for this dataset, as shown in Table I, were fed into a genetic algorithm to improve its performance, these priorities were extracted from research in [26], [27] and [28].

TABLE I. PRIORITY OF ATTRIBUTES IN CKD DATASETS

Attribute	Priority
Red Blood Cells	1
Serum Creatinine	2
Hemoglobin	3
Packed Cell Volume	4
Pus Cell	5
Diabetes Mellitus	6
Specific Gravity	7
Albumin	8
Hypertension	9

The outline of the paper is as follows: Section II describes the main concept and definitions. Section III presents the evolutionary wrapper feature selection method. Section IV presents the experimental study. Finally, Section V gives our conclusions.

II. MAIN CONCEPT AND DEFINITIONS

In this section, basic concepts and definitions regarding the feature selection and classification algorithm are provided, and the data set is then presented.

A. Feature Selections

Feature Selection methods are one topic of representation learning with machine learning, and are a type of dimension reduction technique [8], [29]. Feature Selection is capable of choosing a small subset of relevant features from the original ones by removing irrelevant and redundant features, [3] without significant loss of information or negative degradation of the learning performance [30], [31]. The removal of these irrelevant and redundant features aims to improve classification performance, as well as reducing computational and storage costs, improving time efficiency, and simplifying the prediction model itself [7]. Thus, it is essential to many applications such as text recognition [32] image processing [33], security application [34], and genomic anatomy [35]. However, the main challenge of features selection methods is the need to recognize the best subset of features to obtain the best classifications [36].

Typically, feature selection methods include four steps as shown in Fig. 1: candidate feature subset, evaluation of subset, stopping conditions and performance [37], [38].

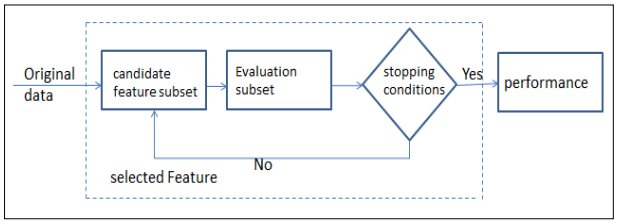

Figure 1. Feature selection process.

Candidate subset generation is the process of selecting feature subsets. The obtained subset using a selection algorithm (e.g. complete, heuristic, random), will be used as input for the evaluation function. Moreover, the feature selection algorithm can start at three points L it may start with an empty set of features, start with the complete set of features, or start with a random subset of features [39], [40]. We also note that Stopping conditions are used to avoid the state of a comprehensive or infinitive loop in feature search method [38].

The evaluation function evaluation the value of the candidate subset obtained by the search, compares the current evaluation value with the best optimal value stored before, and then chooses the best or higher value [38]. With regard to evaluation measurement, feature selection methods are usually divided into three strategies: Filters, Wrappers, and Embedded methods [7]. The difference, however, comes from whether or not implementing an evaluation measure exploits a classification method (e.g. wrappers and embedded methods) (e.g. filters) [36].

Performance is not a part of the feature selection algorithm but is necessary to examine feature subset

accuracy in the classifier and compare the results with the original results, or other candidate features results for the dataset [41].

- Wrapper methods

Wrapper methods are presented by John in 1994, as shown in Fig. 2 [42]. Wrapper methods exploit learning algorithms to evaluate selected subsets, without any limitation to the learning algorithm [43]; consequently Decision trees, ELM, KNN, SVM and etc. may be used for the wrapper method [11]. The evaluation measure of a new subset will terminate when the required accuracy is obtained. Because of this, wrapper methods can obtain better subsets than filter methods, but they are more computationally demanding than filter methods. However, they perform better for feature selection. [7], [36].

Typically, in wrapper methods, the search space for subsets is a known issue, where the number of subsets arises exponentially with the number of features [44]. Hence, the suboptimal subset is found heuristically by employing various search strategies to find subsets heuristically, such as: Sequential search, Evolutionary search [45], [46], Branch-and-bound [47] and Best-first [48]. Moreover, in the genetic algorithm [45], [49], one evolutionary search strategy is often applied in wrapper models; it can always find the best suboptimal subset that achieves good results in the. classification rate [24], [25], [50].

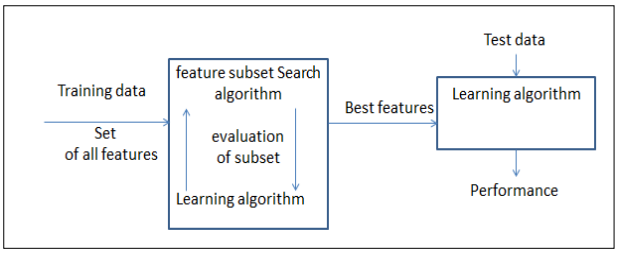

Figure 2. The procedure of wrapper selection.

B. Genetic Algorithm (GA)

The GA technique is an evolutionary algorithm that is implemented to gain a near-optimal solution in large space problems [51]. The development of a GA technique in regards to natural selection and Mendel's laws of inheritance is that it simulates the process of natural evolution which involves the encoding of chromosomes; the selection of Genetic manipulation and evolution; crossover and mutation operation; and, finally, the generation and evaluation of new generations. GA takes a large amount of time during optimization because it include many parameters that need to be processed and encoded. [52].

C. Extreme Learning Machine

An Extreme Learning Machine (ELM) [14] is a learning algorithm for a Single hidden Layer Feedforward Neural network (SLFN), as shown in Fig. 3. ELM is very efficient and effective compared to standard SLFN, since ELM does not need to tune the hidden weights in implementation, [16] but it chooses hidden nodes randomly and determines the output weights using analytical calculation [14], [53]. Hence, an algorithm of ELM can be explained as follows: Given N distinct training samples

$$N = (\boldsymbol{x}_i, \boldsymbol{t}_i) \mid \boldsymbol{x}_i \in \mathrm{R}^m, \boldsymbol{t}_i \mathrm{R}^m \in, i = 1, \ldots, N,$$

The output function of ELM based on SLFNs for a training with L hidden nodes and g(x) as an activation function can be represented by

$$F_L(X) = \sum_{i=1}^{L} \beta i\, G(a_i, b_i, x_j), j = 1, \ldots, N \qquad (1)$$

where $F_L(X)$ is the output vector of the SLFN with respect to the input sample x_j. a_i, b_i are learning parameters generated randomly for the hidden node, a_i is the input weight vector connecting the i^{th} hidden node and input node and b_i is the bias of the i^{th} hidden node. While, the $G(a_i, b_i, x_j)$ is the activation function of the original ELM, and βi is the link of output weight connecting the i^{th} hidden node and the output nodes.

Equation (1) can be written compactly as

$$H\beta = \mathrm{T} \qquad (2)$$

where, H is the output matrix of the hidden layer, T is the target output and β is the output weights.

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix} \mathrm{NXL}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \cdot \\ \cdot \\ \beta_1^T \end{bmatrix} \mathrm{LXN}\ T = \begin{bmatrix} t_1^T \\ \cdot \\ \cdot \\ t_N^T \end{bmatrix} \mathrm{NXM}$$

However, ELM chooses hidden nodes parameters (e.g., a_i, b_i) randomly, and to minimizes the cost function $(F_L(X) - T)$. So, from a linear algebraic viewpoint, Eq. (2) becomes a linear equation, and the output weights β can be analytically determined by finding a least-square solution as follows

$$\beta = H^{\dagger} T \qquad (3)$$

where $H^{\dagger}$ is the Moore-Penrose generalized inverse of matrix H and $T = [t_1, \ldots, t_N]^T$. So, the output weights can be set mathematically, which means less training time by avoiding the adjustment of the network's parameters iteratively with some appropriate learning parameters (e.g. learning rate and iterations)

Hence, the ELM algorithm steps can be summarized as follows:

1) Input a training dataset $N = (x_i, t_i) \mid x_i \in R^m$, $t_i \in R^m$, i=1, …, N.
2) Randomly assign the parameters of hidden nodes, input weight and bias (a_i, b_i) for finding hidden layer output matrix H.
3) Finding the output weight

$$\beta = H^{\dagger} T \qquad (4)$$

III. Extreme Learning Machines - Wrapper Feature Selection Classifier

Classifier aims to classify an instance based on its features. However, if features comprise the irrelevant or the redundancy features within, these features may have a negative impact on the classification performance of the classifier; to overcome this problem feature selection strategy seeks to terminate irrelevant and redundant features to enhance classification performance [2], [3].

In an initiative to display the effectiveness of the aforementioned strategy, we propose a new approach of feature selection that combines priorities of features with ELM-GA wrapper feature selection (ELM-GWPF). The aim of ELM-GWPF is to improve classification performance, enhance time efficiency, and simplify the classifier model itself.

The computational work schema for this approach is depicted in Fig. 3. This can be described in the following detailed steps:

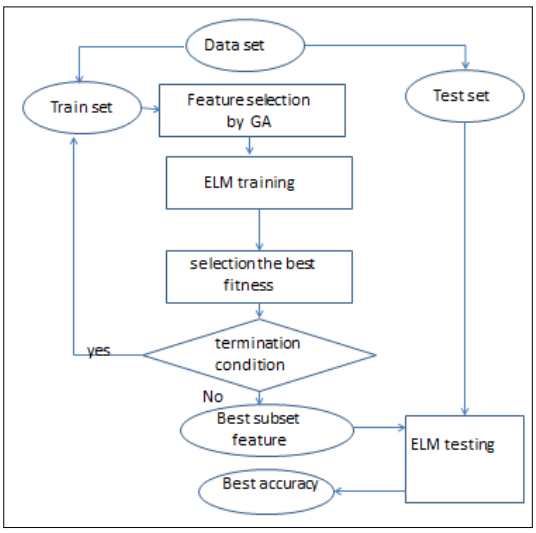

Figure 3. Extreme learning machines - wrapper feature selection computational scheme.

First of all, the original data is uploaded and then the GA runs to search for the best feature subset.

Algorithm 1 describes the main steps involved in the modified GA, which defines a set of solutions that collectively represent a population. The algorithm starts by setting priority pi to each attribute (i.e. i attribute number) in (Line 1). Then, the population p is created by randomly generating a collection of solutions in (Line 2), the population p is a set of N solutions at iteration t and each solution is a subset of features that can be noted in equation (5).

$$P(t) = \{p(t)_{j=1}^{n} = 1\}_{i=1}^{N} \quad (5)$$

$$s.t.p(t)j = \begin{cases} 1\ candedate\ feature \\ 0\ not\ candedate\ feature \end{cases}$$

where $P(t)$ is the population at iteration t, N is the number of population size, n is the full solution, j is the feature number and $p(t)i$ is the solution at the population at iteration t.

The solutions are evaluated by computing the fitness function (Line 4), after which the result is compared with the current solution. In Line 6 features selection priority was examined to avoid removing high priority attribute then the new solution is subsequently formed by a crossover function in Line 7. The next step is a mutation function, which is invoked to replace the worst solution with a new one (Line 8). Finally, these steps are repeated until the stop condition is met (Lines 3).

Consequently the classification performance of a subset of features $P(t)$, is known as the fitness $f\ (S)$, where the objective of the feature selection is to maximize $f(s)$ which can be formulated as follows:

$$max\ f\ (S)\ s.t.S = P(t) \quad (6)$$

Algorithm 1: The pseudo code of the modified Genetic Algorithm

Procedure Genetic ( )
1. Set pi for attributes.
2. Initialize (P, C.), //initialize the population
3. while (the termination condition is satisfied) do
4. Evaluate (P)
5. Best =Select (P) //select best fitness.
6. Examine the priority of attribute to remove the less one.

7. Crossover (P, C) //to produce new solution using priority

8. Mutation (P, C) //replace the worst solution with the best one using priority

9. end while

10. return Best // return best solution

End Procedure

A. Performance Evaluation Measures

The classification performance of a subset of features $f(s)$ can be calculated in different indicators by comparing the predicted and real classification of the instance's classes in the testing dataset. In [54] Different performance indicators were presented from these, we used the following indicators:

- Precision (P) is defined as the proportion of the true positives against all the positive results

$$Precision\ (P) = TP/(TP + FP)$$

- Total Accuracy, is the proportion of true results (both true positives and true negatives) in the population.

$$Total\ Accuracy\ (TA) = TP + TN/(TP + TN + FP + FN)$$

- Recall, is defined as the percentage of positive labeled instances that were predicted as positive.

$$Recall\ (R) = TP/(TP + FN)$$

- F-Measure is the harmonic mean of precision and recall.

$$F - measure = 2PR/(P + R)$$

where,

- True Positive (TP), number of instance predicted positive that are actually positive
- False Positive (FP), number of instance predicted positive that are actually negative
- True Negative (TN), number of instance predicted negative that are actually negative
- False Negative (FN), number of instance predicted negative that are actually positive.

IV. Experimental Study

A. Dataset Description

The Chronic Kidney Disease dataset (CKD) was uploaded by the UCI Machine Learning Repository in 2015 from the Apollo hospital for a period 2 of months, and has 400 instances; 250 instances with CKD and 150 instances without CKD (NotCKD). In addition, it has 25 attributes; 14 nominal, and 11 numeric as mentioned in Table II [55]. This data includes some missing values. To estimate the missing values, we use k-nearest neighbor model.

TABLE II. Attributes of CKD Dataset

No.	Attribute	No.	Attribute
1	Specific Gravity	**13**	Pus Cell clumps
2	Albumin	**14**	Age
3	Sugar	**15**	Blood
4	Red Blood Cells	**16**	Blood Glucose Random
5	Pus Cell	**17**	Blood Urea
6	Bacteria	**18**	Serum Creatinine
7	Hypertension	**19**	Sodium
8	Diabetes Mellitus	**20**	Potassium
9	Coronary Artery Disease	**21**	Hemoglobin
10	Appetite	**22**	Packed Cell Volume
11	Pedal Edema	**23**	White Blood Cell Count
12	Anemia	**24**	Red Blood Cell Count

B. Experimental Conditions

The parameters of the Genetic algorithm used in this article are detailed in Table III. Mutation rate is set to 1/n where n is the number of attributes.

For ELM we tested 50 hidden units [56], and used the sigmoid activation function, because the dataset used is binary, where ELM usually uses a nonlinear Sigmoid activation function with this type of data [57], [58]. while out of from the data, the value of the ratio of the amount of training data and test data is 70:30 with 70 percent being employed for the training phase, and the rest (30 percent) being used for the testing phase.

TABLE III. GA Algorithm Parameters

1	Minimum number of attributes
5	Population size
100	Maximum number of generations
Tournament	Selection
0.25	Tournament size
1/n	Mutation rate
0.5	Crossover rate

C. Performance Evaluation

Feature selection methods are used to reduce the features of the dataset to improve classification performance. In this paper, the classification conducted has been made for two classifiers ELM and SVM using three different methods: first, the result of the classification with all features (Base classifier); secondly, the result of the classification with the best subset of features but without the priority of the features (ELM-GWF); thirdly the result of the classification for our approach ELM-GWFP.

The value of the classification result can be seen in Table IV. This table shows the changes in classification results from the base classifier. There is an increase in classification results. On the third method, there are changes from the first and second methods. The changes can be seen graphically in Fig. 4.

From Table IV, the accuracy of the first method using a/the base classifier is quite high, and the accuracy rate in both ELM and SVM is, 0.960 and 0.958 respectively. The second method where the features are selected by ELM-GWF was able to increase the accuracy of the base classifier. Accuracy rate in SVM and ELM increased by 0.005 and 0.020, respectively. Meanwhile, in the third method, the features were selected using the ELM-GWFP method, the accuracy rate of the base classifiers was increased, and the accuracy rate in ELM was 0.981 and in SVM is 0.975. Consequently, the highest accuracy rate was achieved in the third method, with ELM as a base classifier.

So our proposed Feature selection method namely (ELM-GWFP) was able to improve the classification results on the base classifiers. The improvement of classification results can be seen in the four test parameters i.e. accuracy; precision; recall and f-measure as shown in Fig. 4 and Table IV. This improvement results from the use of a feature priority when selecting a subset of features used to classify data. The utility of the feature priority comes from avoiding high-priority feature interchange with a low-priority feature.

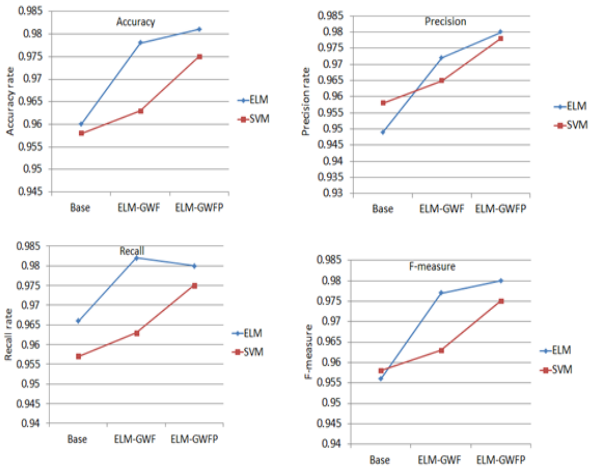

Figure 4. Evaluation parameter.

A summary of classifiers results value with and without using feature selection methods is given in Table

V. These classifiers, without using feature selection, have the least accuracy rate in the diagnosis of CKD, while they have the highest accuracy rate with the ELM-GWFP method.

From Table V, it can be observed that the dimension was reduced by using the ELM-GWF method (15 attributes selected from 25 attributes) and SVM and ELM classifiers have an accuracy rate of higher than the accuracy rate of using all attributes of the dataset.

However, this is not the best feature selection method because our method has higher accuracy rates with the lowest reduced dimension (10 attributes selected from 25 attributes).

TABLE IV. CLASSIFICATION RESULT

Parameter (%)	***Classifiers***		
	EXLM		**SVM**
	First Method		
Accuracy	0.960		0.958
Precision	0.949		0.958
Recall	0.966		0.957
F-measure	0.956		0.958
	Second Method		
Accuracy	0.978		0.963
Precision	0.972		0.965
Recall	0.982		0.963
F-measure	0.977		0.963
	Third Method		
Accuracy	0.981		0.975
Precision	0.980		0.978
Recall	0.980		0.975
F-measure	0.980		0.975

TABLE V. SUMMARY OF CLASSIFIERS RESULTS VALUE WITH AND WITHOUT USING FEATURE SELECTION METHODS

Parameter	**Classifiers**	
	EXLM	**SVM**
	First Method	
Accuracy	0.96	0.958
Number of Features	25	25
	Second Method	
Accuracy	0.978	0.963
Number of Features	15	15
	Third Method	
Accuracy	0.981	0.975
Number of Features	10	10

To obtain a better understanding and illustration, GA was applied with the population per generation through 100 generations. Fig. 5 shows the plot of both the best accuracy per generation, when the GA was employed without feature priorities, and the best accuracy per generation when the GA was employed with feature priority. The best accuracy was reached during 35 generations by using feature priority, while during 50 generations; the best accuracy was reached for GA that is employed without the priorities constraint. Hence the feature's priority reduces performance time.

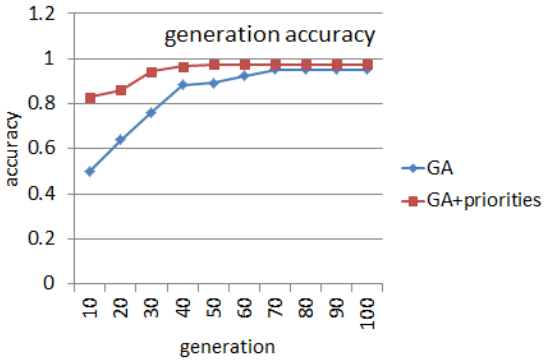

Figure 5. Best accuracy per generation along GA, without the priorities of the features and with the priorities of the features.

V. CONCLUSION

This study aims to evaluate the impact of employing features selection technique for machine learning classifier algorithms.

The initial features were selected using evolutionary wrapper method feature selection using ELM as the base classifier training algorithm; GA was used as a search algorithm. Moreover, there were different priorities for features considered for enhancing GA performance. We referred to our model in this work by the abbreviation ELM-GWFP (ELM, Genetic algorithm, Wrapper feature selection, Priority of features).

At the classification stage, ELM and SVM were used to evaluate the effect of our model in improving classifier performance. Classification performance was measured by using four parameters, namely accuracy; f-measure recall and precision.

The experimental results have demonstrated that classifiers with features selection have produced superior prediction performance in terms of classification accuracy for our considered dataset. Furthermore, using the priority of the features has an essential impact on results accuracy.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Methaq Kadhum: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Drafting the manuscript. Saher Manaseer and Abdel Latif Abu Dalhoum: drafting and revise the manuscript. All authors had approved the final manuscript.

REFERENCES

[1] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, pp. 3-24, 2007.

[2] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1-2, pp. 273-324, Dec. 1997.

[3] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205-1224, 2004.

[4] L. C. Molina, L. Belanche, and À. Nebot, "Feature selection algorithms: A survey and experimental evaluation," in *Proc. IEEE International Conference on Data Mining*, 2002, pp. 306-313.
[5] B. Raman and T. R. Ioerger, "Enhancing learning using feature and example selection," *Texas A&M Univ. Coll. Stn.*, 2003.
[6] S. Wang, J. Wei, and Z. Yang, "Discrimination structure complementarity-based feature selection," *Comput. Intell.*, vol. 33, no. 4, pp. 863-898, 2017.
[7] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157-1182, 2003.
[8] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications*, Springer, 2008, vol. 207.
[9] D. Chyzhyk, A. Savio, and M. Graña, "Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI," *Neurocomputing*, vol. 128, pp. 73-80, 2014.
[10] M. Banerjee, S. Mitra, and H. Banka, "Evolutionary rough feature selection in gene expression data," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.)*, vol. 37, no. 4, pp. 622-632, 2007.
[11] M. A. Hall and L. A. Smith, "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper," in *Proc. FLAIRS Conference*, 1999, pp. 235-239.
[12] B. Kumari and T. Swarnkar, "Filter versus wrapper feature subset selection in large dimensionality micro array: A review," *International Journal of Computer Science and Information Technologies*, vol. 2, no. 3, pp. 1048-1053, 2011.
[13] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, no. 1-3, pp. 389-422, 2002.
[14] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 2006.
[15] H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *J. Med. Syst.*, vol. 41, no. 4, 2017.
[16] L. Chen, G. B. Huang, and H. K. Pung, "Systemical convergence rate analysis of convex incremental feedforward neural networks," *Neurocomputing*, vol. 72, no. 10-12, pp. 2627-2635, 2009.
[17] B. P. Chacko, V. R. V. Krishnan, G. Raju, and P. B. Anto, "Handwritten character recognition using wavelet energy and extreme learning machine," *Int. J. Mach. Learn. Cybern.*, vol. 3, no. 2, pp. 149-161, 2012.
[18] W. Zheng, Y. Qian, and H. Lu, "Text categorization based on regularization extreme learning machine," *Neural Comput. Appl.*, vol. 22, no. 3-4, pp. 447-456, 2013.
[19] K. Anam and A. Al-Jumaily, "Evaluation of extreme learning machine for classification of individual and combined finger movements using electromyography on amputees and non-amputees," *Neural Networks*, vol. 85, pp. 51-68, 2017.
[20] A. A. Mohammed, R. Minhas, Q. M. J. Wu, and M. A. Sid-Ahmed, "Human face recognition based on multidimensional PCA and extreme learning machine," *Pattern Recognit.*, vol. 44, no. 10-11, pp. 2588-2597, 2011.
[21] S. Saraswathi, S. Sundaram, N. Sundararajan, M. Zimmermann, and M. Nilsen-Hamilton, "ICGA-PSO-ELM approach for accurate multiclass cancer classification resulting in reduced gene sets in which genes encoding secreted proteins are highly represented," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 8, no. 2, pp. 452-463, 2011.
[22] K. Anam and A. Al-Jumaily, "A novel extreme learning machine for dimensionality reduction on finger movement classification using sEMG," in *Proc. 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2015, pp. 824-827.
[23] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature Extraction, Construction and Selection*, Springer, 1998, pp. 117-136.
[24] W. H. Hsu, "Genetic wrappers for feature selection in decision tree induction and variable ordering in Bayesian network structure learning," *Inf. Sci. (Ny).*, vol. 163, no. 1-3, pp. 103-122, 2004.
[25] J. Lu, T. Zhao, and Y. Zhang, "Feature selection based-on genetic algorithm for image annotation," *Knowledge-Based Syst.*, vol. 21, no. 8, pp. 887-891, 2008.
[26] K. M. Z. Hasan and M. Z. Hasan, "Performance evaluation of ensemble-based machine learning techniques for prediction of chronic kidney disease," *Emerging Research in Computing, Information, Communication and Applications*, pp. 415-426, 2019.
[27] M. S. Wibawa, I. M. D. Maysanjaya, and I. M. A. W. Putra, "Boosted classifier and features selection for enhancing chronic kidney disease diagnose," in *Proc. 5th International Conference on Cyber and IT Service Management*, 2017.
[28] R. Misir, M. Mitra, and R. K. Samanta, "A reduced set of features for chronic kidney disease prediction," *J. Pathol. Inform.*, vol. 8, 2017.
[29] H. Liu and H. Motoda, *Feature Extraction, Construction and Selection: A Data Mining Perspective*, Springer Science & Business Media, 1998, vol. 453.
[30] J. Li and H. Liu, "Challenges of feature selection for big data analytics," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 9-15, 2017.
[31] R. P. S. Manikandan and A. M. Kalpana, "A study on feature selection in big data," in *Proc. International Conference on Computer Communication and Informatics*, 2017, pp. 1-5.
[32] S. Wang, D. Li, X. Song, Y. Wei, and H. Li, "A feature selection method based on improved fisher's discriminant ratio for text sentiment classification," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8696-8702, 2011.
[33] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1178-1192, 2012.
[34] W. Lee, S. J. Stolfo, and K. W. Mok, "Adaptive intrusion detection: A data mining approach," *Artif. Intell. Rev.*, vol. 14, no. 6, pp. 533-567, 2000.
[35] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinform. Comput. Biol.*, vol. 3, no. 2, pp. 185-205, 2005.
[36] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70-79, 2018.
[37] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 1-4, pp. 131-156, 1997.
[38] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, no. 4, pp. 491-502, 2005.
[39] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artif. Intell.*, vol. 97, no. 1-2, pp. 245-271, 1997.
[40] W. Siedlecki and J. Sklansky, "On automatic feature selection," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 2, no. 2, pp. 197-220, Jun. 2004.
[41] Y. L. Wu, C. Y. Tang, M. K. Hor, and P. F. Wu, "Feature selection using genetic algorithm and cluster validation," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2727-2732, 2011.
[42] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Machine Learning Proceedings*, Elsevier, 1994, pp. 121-129.
[43] S. Maldonado and R. Weber, "A wrapper method for feature selection using support vector machines," *Inf. Sci. (Ny).*, vol. 179, no. 13, pp. 2208-2217, 2009.
[44] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artif. Intell.*, vol. 151, no. 1-2, pp. 155-176, 2003.
[45] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.
[46] E. Alkayal, N. Jennings, and M. Abulkhair, "Optimizing resource allocation using multi-objective particle swarm optimization in cloud computing systems," doctoral dissertation, University of Southampton, 2018.
[47] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Comput.*, no. 9, pp. 917-922, 1977.
[48] H. Arai, C. Maung, K. Xu, and H. Schweitzer, "Unsupervised feature selection by heuristic search with provable bounds on suboptimality," in *Proc. Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
[49] A. Alexandridis, P. Patrinos, H. Sarimveis, and G. Tsekouras, "A two-stage evolutionary algorithm for variable selection in the development of RBF neural network models," *Chemom. Intell. Lab. Syst.*, vol. 75, no. 2, pp. 149-162, 2005.

[50] L. H. Chiang and R. J. Pell, "Genetic algorithms combined with discriminant analysis for key variable identification," *J. Process Control*, vol. 14, no. 2, pp. 143-155, 2004.
[51] D. Whitley, "An executable model of a simple genetic algorithm," in *Foundations of Genetic Algorithms*, Elsevier, 1993, vol. 2, pp. 45-62.
[52] M. Gendreau and J. Y. Potvin, "Metaheuristics in combinatorial optimization," *Ann. Oper. Res.*, vol. 140, no. 1, pp. 189-213, 2005.
[53] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 42, no. 2, pp. 513-529, 2011.
[54] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427-437, 2009.
[55] M. Kumar, "Prediction of chronic kidney disease using random forest machine learning algorithm," *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, no. 2, pp. 24-33, 2016.
[56] D. Xiao, B. Li, and Y. Mao, "A multiple hidden layers extreme learning machine method and its application," *Mathematical Problems in Engineering*, 2017.
[57] J. Peng, L. Li, and Y. Y. Tang, "Combination of activation functions in extreme learning machines for multivariate calibration," *Chemom. Intell. Lab. Syst.*, vol. 120, pp. 53-58, 2013.
[58] Z. X. Chen, H. Y. Zhu, and Y. G. Wang, "A modified extreme learning machine with sigmoidal activation functions," *Neural Comput. Appl.*, vol. 22, no. 3-4, pp. 541-550, 2013.

Methaq Kadhum is a PhD Candidate in the Department of Computer Science at the University of Jordan. She received her B.Sc. in electric engineering from Al-Mustansiriya University in Iraq and M.Sc. in Computer information system from Middle East University (MEU) University in Jordan.

Saher Manaseer is an associate professor at the University of Jordan. He received his PhD in Computer Science from the Department of Computing Science at the University of Glasgow.
His Scientific interests include Computer Networks and Embedded Systems. Currently, Dr. Manaseer is an active researcher in the field of Mobile Ad Hoc Networks.

Abdel Latif Abu Dalhoum received his PhD degree in computer science from the University Autonoma De Madrid, Spain in 2004. He is currently a full Professor of evolutionary algorithms and complex systems at the University of Jordan. He has published about 35 papers in evolutionary algorithms, cellular automata, Fractals and DNA computing. He is a member of GHIA research group at the University Autonoma De Madrid.