Retraction

Retraction: Impact of Feature Selection for Data Classification Using Naive Bayes Classifier (J. Phys.: Conf. Ser. 1879 022088)

Published 13 February 2023

This article has been retracted by IOP Publishing following an allegation that this article may contain questionable references [1].
IOP Publishing has investigated in line with the COPE guidelines and have found the article contains unreliable references that have no significant relevance to the article and agree this article should be retracted [2].
The authors neither agree nor disagree to this retraction.
[1] https://pubpeer.com/publications/2AE026290395620A7918F33163C7B0
[2] Eman Hato 2021 J. Phys.: Conf. Ser. 1879 022088
Retraction published: 13 February 2023
Impact of Feature Selection for Data Classification Using Naive Bayes Classifier

Eman Hato

Computer Science Department, College of Science, Al-Mustansiriyah University

Email: emanhatto@gmail.com

Abstract. In the field of data processing and analysis, the dataset may be a large set of features that restrict data usability and applicability, and thus the dimensions of data sets need to be reduced. Feature selection is the process of removing as much of the redundant and irrelevant features as possible from the original dataset to improve the mining process efficiency. This paper presented a study to evaluate and compare the effect of filter and wrapper methods as feature selection approaches in terms of classification accuracy and time complexity. The Naive Bayes Classifier and three classification datasets from the UCI repository are utilizing in the classification procedure. To investigate the effect of feature selection methods, they are applied to the different characteristics datasets to obtain the selected feature vectors which are then classified according to each dataset category. The datasets used in this paper are the Iris, Ionosphere, and Ovarian Cancer dataset. Experimental results indicate that the filter and wrapper methods provide approximately equal classification accuracy where the average accuracy value of the Ionosphere and Ovarian Cancer dataset is 0.78 and 0.91 for the same selected feature vectors respectively. For Iris dataset, the filter method outperforms the wrapper method by achieving the same accuracy value using only half number of selected features. The results also show that the filter method surpasses when considering the execution time.

Keywords: Feature selection, Naive Bayes classifier, Iris, Ionosphere, Ovarian cancer.

1. Introduction

Data Classification is an analysis technique used to categorize data into different classes. Classification process is carried out in two phases which are training phase and testing phase. In the first phase, also known as learning phase, the classifier model is trained using a classification algorithm with a pre-determined set of data inputs called the training data set. The classifier model, in the second phase, is employed for the classification process with another set of data, called the testing data set. The training and testing data sets are prepared by dividing the general data set by a certain percentage, noting that testing data samples are not included in the training data set [1]. The classification process depends on two main factors which are the classifiers used and the feature vectors extracted [2]. Different advanced classifiers have a powerful learning ability and high performance such as Convolution Neural Networks (CNN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naive Bayes Technique, and Random Forest [3, 4]. Feature extraction is an important step for multimedia processing such as video retrieval, image classification and object recognition. Feature extraction aims to extract relevant information from the data to obtain a robust descriptor and reduce huge data content. Massive data will produce a high dimensionality of the feature vector, which greatly decreases the efficiency of the classification process [5]. Feature selection is always discussed with feature extraction; it is a procedure in machine learning to selects a useful sub-features set that helps in finding the most important features of model...
construction. Feature selection eliminates redundant and irrelevant features as much as possible [6]. The motivation for feature selection is dimensionality reduction of feature vectors to decrease the time complexity, as well as remove irrelevant features that lead to false prediction and thus reduce the performance of classifier [7]. Researchers attempt to extract important features or minimize noise and redundant features from the high dimensions of feature vectors utilizing diverse feature selection techniques.

In [8] Torija and Ruiz proposed prediction model for urban environments based on machine learning regression and feature selection techniques. Three feature selection methods were used which are correlation based feature subset selection, wrapper for feature subset selection and the principal component analysis to reduce time complexity and resources cost for the large number of variables involved in urban environments.

Imtiaz et al. in [9] designed brain tumor segmentation method using multi planar superpixel level features extracted from magnetic resonance imaging. A histogram based consistency analysis was applied as a feature selection method to reduce the feature vectors. In [10] Haidar and Verma proposed a method for selecting optimal input features and network parameters based on a hybrid genetic algorithm for artificial neural networks which was applied for rainfall forecasting.

In [11] Bolon-Canedo and Remeseiro give a detailed survey of feature selection techniques using in image analysis process, they considered four fields in the review which are image classification, image segmentation, image annotation and image retrieval. Silva et al. in [12] combined the feature selection methods applied on training data and machine learning algorithms to identify predictors of prediabetes for enhancing predictive performance of prediabetes. Tumar et al. in [13] Binary Moth Flame Optimization (BMFO) was used as feature selection algorithm to reduce data dimensions for improving the performance of Software fault prediction (SFP) model.

Magesh and Swarnalatha employed Decision Tree (TD) for optimal feature selection in [14]. After identifying a subset of the features, classification algorithms such as Random Forest (RF), Support Vector Machine (SVM), Linear Model (LM) and DT were applied for heart disease prediction and their performances are evaluated. Feature dimensions were reduced by approximately 61% without affecting classification accuracy. In [15] Liu et al. presented feature selection method based on independent feature space search to improve the text classification performance.

The summary from previous studies is that feature selection methods aim to reduce the feature dimensions as much as possible while maintaining the predictive accuracy of the classifiers. The objective of this paper is to evaluate and compare the effect of feature selection methods in terms of classification accuracy and time complexity.

2. Feature Selection Approaches

Feature selection methods aim to reduce the data dimensionality by selecting a subset of extracted features to create a classification model. Feature selection methods search for a subset of features that achieve an optimal response from the classification algorithms so that the learning algorithms can focus on the relevant features that are more useful for the prediction process. Improving the performance of prediction and providing more faster and cost effective predictors are the main advantages of feature selection methods. Using too many features can degrade performance of prediction even when all features are relevant and contain information about the response variable [16, 17].

The feature selection methods can be classified into three approaches [18]:

1. Filter approach aims to measure features importance based on the general features characteristics such as features variance and features relevance to the response. The important features are selected as a step in pre-processing the data and then the model is trained using the selected features. Therefore, this approach is not related to the training algorithm.

2. Wrapper approach aims to select better subsets of features to enhance the performance of the learning algorithm. It starts training model using a subset of features and then removes or add a features based on a selection criterion. The selection criterion directly evaluates the model performance that results from removing or adding a feature. Model training and improving is repeated until the stopping criteria are met.
3. Embedded approach learns importance of feature as part of the model learning process. When the model is trained, the importance of the features is obtained in the trained model. Often the embedded approach cannot achieve better learning performance than wrapper approach.

3. Naive Bayes Classifier

Naive Bayesian classifier is a probabilistic classifier that applies the Bayes' theorem along with Naive assumptions about feature independence. Assume that each instance in the learning set is defined by attribute vector \( X = (x_1, x_2, x_3, \ldots, x_n) \), and the m classes \( C = (C_1, C_2, \ldots, C_m) \) exist. The naive Bayes classifier assigns the \( X \) to the class of maximum probability. The maximized probability can be defined as the follows [19, 20]:

\[
P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)}
\]  

(1)

Where

\( P(C_i | X) \): The probability of \( C_i \) in case \( X \) occurs.
\( P(X | C_i) \): The probability of \( X \) in case \( C_i \) occurs.
\( P(X) \): The probability of \( X \).
\( P(C_i) \): The probability of \( C_i \).

The classification problem according to equation (1) is calculating the \( P(C_i | X) \) since it finds the probability that the given \( X \) belongs to class \( C_i \). Because \( P(X) \) is a constant and generally assumed \( P(C_i) \) have the same probability, only the maximum value of \( P(X | C_i) \) must be determined.

Assuming that the relationships of attribute in each class are statistically independent of each other, \( P(X | C_i) \) can be estimated as follows [19, 20]:

\[
P(X | C_i) = \prod_{k=1}^{n} P(x_k | C_i)
\]

(2)

This assumption results in efficient data classification process and simpler computation cost.

4. Methodology and Materials

In this section, details of methodology, classification datasets and performance evaluation measures are described.

4.1. The Methodology

The method in this paper is designed for comparison and finding the most appropriate feature selection methods for classification task. Their effects are evaluated based on performance of Naive Bayes classifier according to classification accuracy and time complexity.

The paper method consists of several steps. First, three popular datasets are collected from the UCI repository to conduct the experiments. Each dataset is divided into two subsets including training data and testing data. Second, the features selection methods are applied to select essential features. The next, Naive Bayes classifier is employed for classification process. Further, the comparison is performed to highlight the strengths and effectiveness of features selection methods in data classification.

The feature selection approaches employed in this paper are filter and wrapper method. The filter method rank importance of features using Relief algorithm. Whereas, wrapper method defines a subset of the training data that best predict the testing data by sequentially selecting features until there is no change in defined criterion value. The criterion adopted is mean squared error.

4.2. Classification Datasets

Three classification datasets from the UCI Machine Learning Database Repository were used in the process of evaluating the impact of feature selection methods for data classification. Datasets vary in the number of instances, the number of attributes, and the number of classes, which allows testing the effect of feature selection in different cases.

- **Iris Dataset**: represent data describing Iris flower of three related species. The dataset consists of 150 instances belonging to three classes which are "Iris Setosa", "Iris Virginica" and "Iris Versicolor". Each instance consists of 4 attributes measured in centimeters which are the sepal length, sepal width, petal length and petal width [21].

- **Ionosphere Dataset**: represent data receiving from the classification of radar returns from the ionosphere. The dataset consists of 351 instances belonging to two classes which are "Good" and "bad" and each instance consists of 34 attributes [22].
Ovarian Cancer Dataset: the dataset exemplify the information on cancer of ovarian for different patients. The dataset consists of 261 instances belonging to two classes which are "Cancer" and "Normal" and each instance consists of 4000 attributes [23]. Table 1 gives a summary describing the three datasets.

Table 1. The datasets summarization

| Dataset Name | Attribute Characteristics | Instances Number | Attributes Number | Associated Task |
|--------------|----------------------------|------------------|------------------|-----------------|
| Iris         | Real                       | 150              | 4                | Classification  |
| Ionosphere   | Integer, Real              | 351              | 34               | Classification  |
| Ovarian Cancer | Real                     | 261              | 4000             | Classification  |

4.3. Performance Evaluation Measures
The evaluation measure is based on the classification accuracy that summarizes the classification performance as the percentage of the number of correctly classified instances out of the total number of instances. The measure values are based on the statistical values of True Positive (TP: actual and predicted values are correct positive), True Negative (TN: actual and predicted values are correct negative), False Positive (FP: actual value is negative while predicted value is positive), and False Negative (FN: actual value is positive while predicted value is negative) [24, 25].

Accuracy = \( \frac{TP+TN}{TP+FP+TN+FN} \)  \hspace{1cm} (3)

The error classification can be calculated from classification accuracy because they are complements of each other. The classification error measured as the percentage of the number of incorrectly classified instances out of the total number of instances [1].

Error Rate = Accuracy – 1 \hspace{1cm} (4)

A confusion matrix is a visual summary of the predictions made by a classification model organized as a matrix. It provides a clear picture of which classes are being correctly and incorrectly predicted and what type of errors are being made. Matrix rows represent the actual classes, and matrix columns indicate the predicted classes. The cell values represent the number of predictions made for a class that are actually for a given class. Classification accuracy can be calculated by computing by averaging the values across the main diagonal of confusion matrix. High score refers to high classification accuracy [6, 24].

Computational complexity is an important aspect of an effective classification model. Since features selection helps reduce the time consumed, execution time is also used to evaluate performance; it measures the response speed of the classification model.

5. Experimental Results
Initially, from each dataset, 70% of samples are selected randomly for training data and 30% of samples are also selected randomly for testing data. The method implementation and experiments are performed on Hp PC with an Intel Core i7–5500 4.40GHz CPU and 12GB RAM running Matlab 2018a.

The filter method is applied to three datasets, where the importance of each feature attribute is calculated. To get an accurate assessment, Naïve Bayes performs the classification process using the most important feature attribute, and then gradually increases the feature attributes used in the classification process according to their importance. The classification accuracy, error rate and execution time are recorded with each addition.

Table 2 shows the classification performance of Iris dataset for different cases, starting with the first case that uses the most important attribute feature, which is the fourth attribute, and ending with the last case that uses all the features attribute. Table 3 shows confusion matrixes of the predictions made by a classification model to the four cases.
Table 2. Effect of the filter method for Iris dataset classification

| Cases  | Selected Attributes | Accuracy | Error Rate | Execution Time in sec. |
|--------|---------------------|----------|------------|------------------------|
| Case 1 | 4th                 | 0.944    | 0.055      | 0.1162                 |
| Case 2 | 4th, 3rd            | 0.972    | 0.027      | 0.1361                 |
| Case 3 | 4th, 3rd, 1st       | 0.972    | 0.027      | 0.1512                 |
| Case 4 | 4th, 3rd, 1st, 2nd  | 0.972    | 0.027      | 0.1812                 |

Table 3. Confusion matrices of filter method for Iris dataset.

### Case 1

| Classes Name | Setosa | Versicolor | Virginica |
|--------------|--------|------------|-----------|
| Setosa       | 1      | 0          | 0         |
| Versicolor   | 0      | 1          | 0         |
| Virginica    | 0      | 0.083      | 0.916     |

### Case 2

| Classes Name | Setosa | Versicolor | Virginica |
|--------------|--------|------------|-----------|
| Setosa       | 1      | 0          | 0         |
| Versicolor   | 0      | 1          | 0         |
| Virginica    | 0      | 0.0833     | 0.916     |

### Case 3

| Classes Name | Setosa | Versicolor | Virginica |
|--------------|--------|------------|-----------|
| Setosa       | 1      | 0          | 0         |
| Versicolor   | 0      | 1          | 0         |
| Virginica    | 0      | 0.083      | 0.916     |

### Case 4

| Classes Name | Setosa | Versicolor | Virginica |
|--------------|--------|------------|-----------|
| Setosa       | 1      | 0          | 0         |
| Versicolor   | 0      | 1          | 0         |
| Virginica    | 0      | 0.166      | 0.833     |

The effects of filter method are clear; it can correctly determine the most important feature attribute. It can be noted that the classification accuracy based on only one feature attribute is very close to the accuracy using more than one feature attribute and with less execution time.

With the same procedure, the filter method is applied to the Ionosphere and Ovarian Cancer dataset, taking into account that the number of feature attributes is large in these datasets, so larger numbers of feature attributes are used by the classifier in each time for brevity.

The performance results are presented in Table 3 and Table 4 for Ionosphere and Ovarian Cancer dataset respectively, while Table 5 and Table 7 show the corresponding confusion matrices.

Table 4. Effect of the filter method for Ionosphere dataset classification

| Cases  | Selected Attributes | Accuracy | Error Rate | Execution Time in sec. |
|--------|---------------------|----------|------------|------------------------|
| Case 1 | The first 4th attributes | 0.784    | 0.215      | 0.3030                 |
| Case 2 | The first 8th attributes | 0.874    | 0.125      | 0.3273                 |
| Case 3 | The first 16th attributes | 0.899    | 0.100      | 0.3696                 |
| Case 4 | The first 24th attributes | 0.923    | 0.076      | 0.4340                 |
| Case 5 | The first 32th attributes | 0.927    | 0.072      | 0.4831                 |
| Case 6 | All attributes       | 0.912    | 0.087      | 0.5159                 |
### Table 5. Confusion matrixes of filter method for Ionosphere dataset

| Case 1 Classes Name | Good | Bad |
|---------------------|------|-----|
| Good                | 0.585| 0.414|
| Bad                 | 0.015| 0.984|

| Case 2 Classes Name | Good | Bad |
|---------------------|------|-----|
| Good                | 0.780| 0.219|
| Bad                 | 0.031| 0.968|

| Case 3 Classes Name | Good | Bad |
|---------------------|------|-----|
| Good                | 0.829| 0.170|
| Bad                 | 0.031| 0.968|

| Case 4 Classes Name | Good | Bad |
|---------------------|------|-----|
| Good                | 0.878| 0.121|
| Bad                 | 0.031| 0.968|

| Case 5 Classes Name | Good | Bad |
|---------------------|------|-----|
| Good                | 0.902| 0.097|
| Bad                 | 0.0468| 0.953|

| Case 6 Classes Name | Good | Bad |
|---------------------|------|-----|
| Good                | 0.902| 0.097|
| Bad                 | 0.078| 0.921|

### Table 6. Effect of the filter method for Ovarian Cancer dataset classification

| Cases | Selected Attributes | Accuracy | Error Rate | Execution Time in sec. |
|-------|---------------------|----------|------------|------------------------|
| Case 1 | The first 5\textsuperscript{th} attribute | 0.907 | 0.092 | 8.7709 |
| Case 2 | The first 500\textsuperscript{th} attribute | 0.858 | 0.141 | 11.4085 |
| Case 3 | The first 1000\textsuperscript{th} attribute | 0.839 | 0.160 | 14.2684 |
| Case 4 | The first 1500\textsuperscript{th} attribute | 0.839 | 0.160 | 17.8197 |
| Case 5 | The first 2000\textsuperscript{th} attribute | 0.858 | 0.141 | 19.4085 |
| Case 6 | The first 2500\textsuperscript{th} attribute | 0.858 | 0.141 | 21.9515 |
| Case 7 | The first 3000\textsuperscript{th} attribute | 0.858 | 0.141 | 25.1981 |
| Case 8 | The first 3500\textsuperscript{th} attribute | 0.839 | 0.160 | 27.3748 |
| Case 9 | All attributes | 0.826 | 0.173 | 31.1855 |

### Table 7. Confusion matrixes of filter method for Ovarian Cancer dataset.

| Case 1 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 1      | 0      |
| Normal              | 0.185  | 0.814  |

| Case 2 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.148  | 0.851  |

| Case 3 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.185  | 0.814  |

| Case 4 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.185  | 0.814  |

| Case 5 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.148  | 0.851  |

| Case 6 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.148  | 0.851  |

| Case 7 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.185  | 0.814  |

| Case 8 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.148  | 0.851  |

| Case 9 Classes Name | Cancer | Normal |
|---------------------|--------|--------|
| Cancer              | 0.864  | 0.135  |
| Normal              | 0.148  | 0.851  |
For Ionosphere dataset, the accuracy is improving every time a new feature attribute is added and correspondingly execution time increases as shown in Table 3. Referring to Table 4, fluctuations in the accuracy values can be observed. Where the accuracy values are expected to have a uniform increase when the feature attribute used are also increased, as in previous datasets. This is due to the variousness and heterogeneity of the Ovarian Cancer dataset, which influenced the process of determining the importance of each feature attribute. The wrapper method is applied on the training data only from each dataset to determine the feature attributes that can be used in the classification process as explained earlier. The results of criterion values and the final features attributes selected are listed in Table 5, Table 6 and Table 7 for Iris, Ionosphere and Ovarian Cancer dataset respectively.

**Table 8. Wrapper feature selection for Iris dataset**

| Steps | Added Attributes | Criterion Value |
|-------|------------------|-----------------|
| 1     | 4<sup>th</sup>    | 0.00457491      |
| 2     | 2<sup>nd</sup>    | 0.00456496      |
| 3     | 3<sup>rd</sup>    | 0.00449546      |
| 4     | 1<sup>st</sup>    | 0.00308608      |

**Final Attributes**: 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>

**Table 9. Wrapper feature selection for Ionosphere dataset**

| Steps | Added Attributes | Criterion Value |
|-------|------------------|-----------------|
| 1     | 5<sup>th</sup>   | 0.00704117      |
| 2     | 1<sup>st</sup>   | 0.00434487      |
| 3     | 10<sup>th</sup>  | 0.00416756      |
| 4     | 24<sup>th</sup>  | 0.00401223      |

**Final Attributes**: 1<sup>st</sup>, 5<sup>th</sup>, 10<sup>th</sup>, 24<sup>th</sup>

**Table 10. Wrapper feature selection for Ovarian Cancer dataset**

| Steps | Added Attributes | Criterion Value |
|-------|------------------|-----------------|
| 1     | 3032<sup>th</sup>| 0.00748559      |
| 2     | 2337<sup>th</sup>| 0.00432453      |
| 3     | 2680<sup>th</sup>| 0.00218727      |
| 4     | 2735<sup>th</sup>| 0.00085638      |
| 5     | 2236<sup>th</sup>| 0.00045620      |

**Final Attributes**: 2236<sup>th</sup>, 2337<sup>th</sup>, 2680<sup>th</sup>, 2735<sup>th</sup>, 3032<sup>th</sup>

It is noticeable that the feature dimensions of Ionosphere and Ovarian Cancer dataset were reduced by wrapper method, while the Iris dataset was not affected. The classification was performed by Naive Bayes algorithm based on the selected attributes feature only. The results are presented in Table 8 details the classification accuracy, error rate and execution
time of the three datasets, while Table 11 illustrate the classification performance as confusion matrices.

**Table 11. Effect of the wrapper method for datasets classification**

| Dataset Name       | Attributes Included | Accuracy | Error Rate | Execution Time in sec. |
|--------------------|---------------------|----------|------------|------------------------|
| Iris               | 4th, 2th, 3th, 1th  | 0.972    | 0.027      | 4.5212                 |
| Ionosphere         | 1th, 5th, 10th, 24th| 0.793    | 0.206      | 37.654                 |
| Ovarian Cancer     | 2236th, 2337th, 2680th, 2735th, 3032th | 0.917    | 0.082      | 2400.106               |

**Table 12. Confusion matrices of wrapper method for datasets classification**

**Iris Dataset**

| Classes Name | Setosa | Versicol | Virgina |
|--------------|--------|----------|---------|
| Setosa       | 1      | 0        | 0       |
| Versicol     | 0      | 1        | 0       |
| Virginica    | 0      | 0.083    | 0.916   |

**Ionosphere Dataset**

| Classes Name | Good | Bad |
|--------------|------|-----|
| Good         | 0.63 | 0.36|
| Bad          | 0.04 | 0.95|

**Ovarian Cancer Dataset**

| Classes Name | Cancer | Normal |
|--------------|--------|--------|
| Cancer       | 0.945  | 0.054  |
| Normal       | 0.111  | 0.888  |

Although the accuracy classification achieved by applying the wrapper method or filter method appears to be similar to the same number of selected feature attributes as shown in Table 8. But the execution time increased exponentially when using a large dataset, which is consistent with the expected results.

It should be noted that the wrapper method was able to identify the most related feature attributes from Ovarian Cancer dataset. Whereas the five selected feature attributes employed in the classification process can achieve 0.917 accuracy classification. This indicates the efficiency and ability of the wrapper method to deal with heterogeneous and varied datasets.

A clear picture was presented by visualizing the comparison of the filter and wrapper method according to the classification accuracy and execution time in a graphical representation illustrated in Figure 1 and Figure 2.

**Figure 1. The Accuracy comparison**
The comparison was made and it can be inferred from the accuracy values that applying the filter and wrapper method has almost the same accuracy value, but the wrapper method consumed more time.

6. Conclusions
Feature selection methods aim to find the most relevant features to the problem domain resulting in improved prediction accuracy and computational speed. The effect of two feature selection methods was assessed in this paper depended on the time consuming for determining the selected feature vectors and the classification accuracy of Naive Bayes algorithm for different datasets.

It is noted upon examining the results obtained that the wrapper method achieved almost the same classification accuracy as the filter method, but at the cost of time consumption. The exact time required to determine the selected feature vectors increases dramatically with the volume of the dataset. Regardless of the extraction time, the wrapper method specified the most relevant features than the filter method when using heterogeneity dataset.

On the other hand, the filter method gave flexibility in choosing the appropriate features that helped to achieve high classification accuracy while maintaining a reasonable execution time. This makes the filter method more suitable for datasets of various sizes.

Acknowledgments
The work presented in this paper was supported by Al-Mustansiriyah University (www.uomustansiriyah.edu.iq), which is gratefully acknowledged.

References
1. Raychaudhuri, K., Kumar, M. and Bhanu, S. "A Comparative Study and Performance Analysis of Classification Techniques: Support Vector Machine, Neural Networks and Decision Trees." Singh M., Gupta P., Tyagi V., Sharma A., Oren T., Grosky W. (Eds.), Advances in Computing and Data Sciences, Communications in Computer and Information Science. In proceedings of International Conference on Advances in Computing and Data Sciences, Springer, Singapore, vol. 721, pp. 13-21, 2016.
2. Wu, C., Li, Y., Zhao, Z., and Liu, B. "Image Classification Method Rationally Utilizing Spatial Information of the Image." Multimedia Tools and Applications, vol. 78, no. 14, pp.19181-19199, 2019.
3. Xin, M., and Wang, Y. "Research on Image Classification Model Based on Deep Convolution Neural Network." EURASIP Journal on Image and Video Processing 2019, vol. 1, no. 40, pp. 1-11, 2019.
4. Noi, P. T., Kappas, M. "Comparison of Random Forest, K-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery." Sensors, vol. 18, no. 1, pp. 1-20, 2018.

5. Zhang, L., and Lu, X. "Feature Extraction Based on Support Vector Data Description." Neural Processing Letters, vol. 49, no. 2, pp. 643-659, 2019.

6. Chen, R., Dewi, C., Huang, S. "Selecting critical features for data classification based on machine learning methods." Journal of Big Data, vol. 7, no.52, pp. 1-26, 2020.

7. Azhar, M. A., and Thomas, P. A. "Comparative Review of Feature Selection and Classification Modeling." In 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), IEEE, pp. 1-9, 2019.

8. Torija, A. J., and Ruiz, D. P. "A General Procedure to Generate Models for Urban Environmental-Noise Pollution Using Feature Selection and Machine Learning Methods." Science of the Total Environment, vol.505, pp.680-693, 2015.

9. Intiaz, T., Rifat, S., Fattah, S. A. and Wahid, K. A. "Automated Brain Tumor Segmentation Based on Multi-Planar Superpixel Level Features Extracted From 3D MR Images." IEEE Access, vol. 4, pp. 25335-25349, 2019.

10. Haidar, A., and Verma, B. "A Novel Approach for Optimizing Climate Features and Network Parameters in Rainfall Forecasting." Soft Computing, vol. 22, no. 24, pp. 8119-8130, 2018.

11. Bolon-Canedo, V., Remeseiro, B. "Feature Selection in Image Analysis: a Survey." Artificial Intelligence Review, vol.53, pp. 2905-2931, 2020, https://doi.org/10.1007/s10462-019-09750-3.

12. De Silva, K., Jönsson, D., and Demmer, R. T. "A Combined Strategy of Feature Selection and Machine Learning to Identify Predictors of Prediabetes." Journal of the American Medical Informatics Association, vol. 27, no. 3, pp. 396-406, 2020.

13. Tumar, I., Hassouneh, Y., Turabieh, H., and Thaher, T. "Enhanced Binary Moth Flame Optimization as a Feature Selection Algorithm to Predict Software Fault Prediction." IEEE Access, vol.8, pp.8041-8055, 2020.

14. Magesh, G., and Swarnalatha, P. "Optimal Feature Selection through a Cluster-Based DT Learning (CDTL) in Heart Disease Prediction." Evolutionary Intelligence, pp. 1-11, 2020, https://doi.org/10.1007/s12065-019-00336-0.

15. Liu, Y., Ju, S., Wang, J., and Su, C. "A New Feature Selection Method for Text Classification Based on Independent Feature Space Search." Mathematical Problems in Engineering, vol. 2020, no. 6076272, pp. 1-14, 2020.

16. Balogun, A. O., Basri, S., Abdulkadir, S. J., and Hashim, A. S. "Performance Analysis of Feature Selection Methods in Software Defect Prediction: a Search Method Approach." Applied Sciences, vol. 9, no. 13, pp. 1-20, 2019.

17. Hoque, N., Singh, M., and Bhattacharyya, D. K. "EFS-MI: an Ensemble Feature Selection Method for Classification." Complex and Intelligent Systems, vol. 4, no. 2, pp. 105-118, 2018.

18. Hancer, E., Xue, B., and Zhang, M. "A Survey on Feature Selection Approaches for Clustering." Artificial Intelligence Review, vol. 53, pp. 4519–4545, 2020, https://doi.org/10.1007/s10462-019-09800-0.

19. Nitta, G. R., Rao, B. Y., Sravani, T., Ramakrishiah, N., and Balaanand, M. "LASSO-Based Feature Selection and Naïve Bayes Classifier for Crime Prediction and Its Type." Service Oriented Computing and Applications, vol. 13, no. 3, pp. 187-197, 2019.

20. Lin, C., Lin, S., Liu, H., and Tseng, K. "Applying the Naïve Bayes Classifier to HVAC Energy Prediction Using Hourly Data." Microsystem Technologies, pp.1-15, 2019, https://doi.org/10.1007/s00542-019-04479-z.

21. Halakatti, S.T., Halakatti, S.T. "Identification of Iris Flower Species Using Machine Learning." International Journal of Computer Science, vol.5, no. 8, pp. 59–69, 2017.

22. Ala’raj, M., Majdalawieh, M. and Abbod, M.F. "Improving Binary Classification Using Filtering Based on K-NN Proximity Graphs." Journal of Big Data, vol. 7, no. 15, pp. 1-18, 2020.
23. Conrads, T. P., Fusaro, V. A., Ross, S., Johann, D., Rajapakse, V., Hitt, B. A., Steinberg, S.M. "High-Resolution Serum Proteomic Features for Ovarian Cancer Detection." Endocrine-related cancer, vol. 11, no. 2, pp.163-178, 2004.

24. Dalianis, H. "Evaluation Metrics and Evaluation." In: Clinical Text Mining. Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-78503-5_6.

25. T. A. Al-asadi and A. J. Obaid, "Object Based Image Retrieval Using Enhanced SURF," Asian Journal of Information Technology, vol. 15, no. 16, pp. 2756-2762, 2016.

Retracted