Comparative Study of Supervised Machine Learning Algorithms on Thoracic Surgery Patients based on Ranker Feature Algorithms

Hezha M.Tareq Abdulhadi[1], Hardi Sabah Talabani[2]

[1]Department of Information Technology, National Institute of Technology (NIT), Sulaymaniyah, KRG, Iraq, [2]Department of Applied Computer, College of Medical and Applied Sciences, Charmo University, Sulaymaniyah, KRG, Iraq

ABSTRACT

Thoracic surgery refers to the information gathered for the patients who have to suffer from lung cancer. Various machine learning techniques were employed in post-operative life expectancy to predict lung cancer patients. In this study, we have used the most famous and influential supervised machine learning algorithms, which are J48, Naïve Bayes, Multilayer Perceptron, and Random Forest (RF). Then, two ranker feature selections, information gain and gain ratio, were used on the thoracic surgery dataset to examine and explore the effect of used ranker feature selections on the machine learning classifiers. The dataset was collected from the Wroclaw University in UCI repository website. We have done two experiments to show the performances of the supervised classifiers on the dataset with and without employing the ranker feature selection. The obtained results with the ranker feature selections showed that J48, NB, and MLP's accuracy improved, whereas RF accuracy decreased and support vector machine remained stable.

Index Terms: Ranker feature selection, Information gain, Gain ratio, Supervised machine learning algorithms, Thoracic surgery, Cross-validation

1. INTRODUCTION

Tracking health results is fundamental to reinforce quality initiative, managing health care, and educating consumer. At present, employing computer applications in medical fields have had a direct impact on doctor's productivity and accuracy. Health results measurement is one of these applications. Health outcomes are playing an increasing role in health-care purchasing and administration. Nowadays and in most countries, cancer is becoming one of the leading causes of death. At present, lung cancer is the most common presage for thoracic surgery [1].

In the last several decades, there has been a lot of study in the field of medical science that has used various computing approaches. In the case of medical care, new approaches to data abstraction make data extraction quick and accurate, providing a larger opportunity to work with data for measuring health results. Cancer is a serious health threat that the world is confronting, thus knowing how to anticipate results is essential [2].

Selecting attribute and features in a massive amount of data and using machine learning approaches in recent medical technique might cause the computing process faster and decrease the amount of redundant data. Removing unnecessary data are advantageous since it decreases the difficulty of data processing. Attribute classifier of the data is significant, in the case of thoracic cancer, it leads to the extraction of varied information regarding a specific case of a patient. To reduce and control the victims of lung cancer and

Corresponding author's e-mail: Hezha M.Tareq Abdulhadi, Department of Information Technology, National Institute of Technology (NIT), Sulaymaniyah, KRG, Iraq. E-mail: Hezha.Abdulhadi@nit.edu.krd

thoracic surgery patients, ranker feature selection techniques became an important and necessary method, because it can challenge and solve this kind of problems. In general, machine learning and ranker algorithms are a technique for classifying patient and disease datasets and separate the data to relevant and irrelevant. There are several studies worked on thoracic surgery. Therefore, this work shed a light on the success rate of machine learning algorithms with ranker feature selections in classifying thoracic surgery patients. The major goal is to obtain an accurate prediction of the result after employing different approaches [3].

This research is done by a famous tool which is WEKA, used for analyzing and classifying data with famous machine learning algorithms. Five different machine learning algorithms employed in this study which are J48, Random Forest (RF), Naïve Bayes, Multilayer Perceptron, and Support Vector Machine (SVM) with two famous ranker feature selections algorithms, information gain and gain ratio (GR). We have performed a classification on the thoracic surgery dataset through machine learning techniques and ranker algorithms.

The rest of this paper is organized as follows: Section 2 describes some background concepts relevant to our review. Section 3 describes the problem and proposed method. Section 4 will present the experiments and results, and finally, the conclusion is stated in Section 5.

2. LITERATURE REVIEW

Various studies have been published that emphasize the significance of methodology in the realm of medical diagnosis. This research used various methods to the problem and obtained reasonable classification accuracies. Following are some examples:

Several studies have been implemented in the medical field for analyzing data to discover patterns and predict outcomes. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE) are used to rectify the unbalanced data. Various measures are used for predicting results. For balancing the data by oversampling the minority class, the comparison between prediction methods such as Artificial Neural Network (ANN), Naive Bayes techniques, and Decision Tree Algorithm is explained in [3] by employing 10-fold cross-validation and SMOTE. The receiver operating characteristics summed the classifier performance based on the true positives and true negatives error rates; the ANN achieves the highest accuracy in this scenario. Another 10

folds cross-validation study in life expectancy prediction was conducted by [1] using Naïve Bayes, Logistic Regression, and SVM with the RF concept, which uses the tree classification technique to average deep multiple trees that are trained using different fragments of the current training set.

Jahanvi Joshi *et al.* offered the detailed proof that K-nearest neighbor (KNN) provides preferable accuracy than expectation-maximization classification technique. Employing the Farthest first algorithm, they showed that 80% of patients were healthy and 20% of patients were sick, which are very close to KNN technique outcome [4].

Vanaja *et al.* explained that each feature selection approach has its effects and weak points inclusion of greater characteristics reduces accuracy. This survey was demonstrated that the feature selection algorithms improve the classifier accuracy consistently [5].

Zieba *et al.* employed boosted SVM to estimate post-operative life expectancy in their study. During the research, an Oracle-based technique to extract decision rules from the boosted SVM for solving problems with unbalanced data had been used [6].

Sindhu *et al.* analyzed thoracic surgical data using six classification techniques (Naive Bayes, J48, PART, OneR, Decision Stump, and RF). An experiment was done and discovered that RF provides the greatest classification accuracy with all split percentages [1].

Another research evaluated the performance of four machine learning algorithms (Nave Bayes, Simple logistic regression, Multilayer perceptron, and J48) with their boosted variants using various measures. The outcomes showed that the boosted simple logistic regression approach outperforms or is at least competitive with the other four machine learning techniques, with an average score of 84.5% [7].

In this work, four various machine learning algorithms will be used for post-life expectancy estimation after thoracic surgery, by employing two novel metrics which are information gain (IG) and GR that can be used to improve the accuracy of the algorithms and provide a reasonable result.

3. METHODOLOGY

In this work demonstrated in Fig. 1, the thoracic surgery dataset is used and pre-processed to remove unbalanced and useless data, then filling missing values. The pre-processed
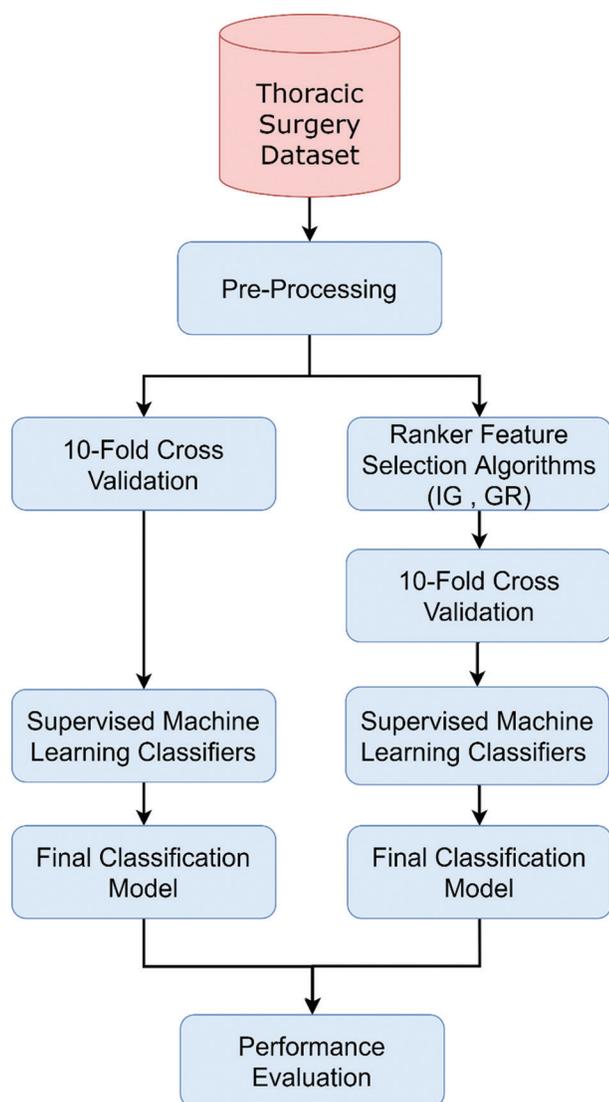
Fig. 1. Flowchart of the proposed method.

results will be compared with the other test according to some measurement criteria. The second test will be done on the dataset using the attribute ranking methods (IG and GR) to eliminate the redundant and irrelevant attributes from the original set of attributes and to evaluate the importance of an attribute by measuring the IG and GR with regard to the class. After attribute evaluation, the dataset will be separated randomly by applying 10-fold cross-validation and then the classification process will begin with the supervised classifiers to find the best performance among them. The final classification model of both tests will be evaluated and compared based on some performance criteria explained in the next chapter.

3.1. Thoracic Surgery Corpus
The dataset used in this paper was collected from the information of patients who were suffering from lung cancer and underwent lung resections in 2007 and 2011 at the Center for Thoracic Surgery in Wroclaw, which, in turn, is affiliated with the Lower Silesian Center for Pulmonary Diseases and the Department of Thoracic Surgery at the University of Wroclaw medical. It is worth noting that this dataset has been extracted from Wroclaw Thoracic Surgery Centre that has been gathered by the National Lung Cancer Registry of the Polish Institute of Lung Diseases and Tuberculosis in Warsaw [8]. In general, the dataset consists of 17 attributes (14 nominal and three numeric) with 470 records, which are detailed in Table 1.

3.2. Pre-Processing
The dataset is pre-processed removing unbalanced and useless data through SMOTE, a bootstrapping algorithm to solve this issue (SMOTE). Other methods, ROS, are also being tested (random over sampler) for that issue. In this work, several new features are designed to better describe the underlying connections among different dataset features, resulting in enhanced model performance [9]. The operations of correcting discrepancies in the data reducing noise in outliers and filling in missing values using one of the data preprocessing methods called (data cleansing).

3.3. Ranker Feature Selection
The two basic principles of ranker-based feature selection algorithms are as follows: First, the evaluation of features related to their impact on the process of data classification or analysis. Second, building a ranking list based on its score using the desired features (the most influential on the accuracy of the algorithm performance) that were identified to create a subset. Among the different types of rank-based feature selection algorithms, two main types. GR and IG

dataset will be used in two different tests. The two main purposes of this paper are as follows: First, to analyze the effect of number of attributes on accuracy of machine learning to solve the problem for prediction of the post-operative life in lung cancer patients reducing the number of attributes and increasing the accuracy is required to minimize the computational time of prediction techniques. Second, to make a comparison between the supervised classifiers performances before and after using ranker feature algorithms with employing 10-fold cross-validation technique for splitting the dataset. Notably, cross-validation is a method to evaluate a predictive model by partitioning the original sample into a training set to train the model and a validation/test set to evaluate it. The first test will be done on the dataset employing supervised machine learning classifiers then the

TABLE 1: Descriptions of thoracic surgery dataset attributes

Attribute ID	Attribute name	Attribute type	Attribute description
1	DGN	Nominal	Diagnosis-specific combination of ICD-10 codes for primary and secondary as well multiple tumors if any (DGN3, DGN2, DGN4, DGN6, DGN5, DGN8, and DGN1)
2	PRE4	Numeric	Forced vital capacity – FVC
3	PRE5	Numeric	Volume that has been exhaled at the end of the first second of forced expiration – FEV1
4	PRE6	Nominal	Performance status – Zubrod scale (PRZ2, PRZ1, and PRZ0)
5	PRE7	Nominal	Pain before surgery (T,F)
6	PRE8	Nominal	Hemoptysis before surgery (T,F)
7	PRE9	Nominal	Dyspnea before surgery (T,F)
8	PRE10	Nominal	Cough before surgery (T,F)
9	PRE11	Nominal	Weakness before surgery (T,F)
10	PRE14	Nominal	T in clinical TNM – size of the original tumor, from OC11 (smallest) to OC14 (largest) (OC11, OC14, OC12, and OC13)
11	PRE17	Nominal	Type 2 DM – diabetes mellitus (T,F)
12	PRE19	Nominal	MI up to 6 months (T,F)
13	PRE25	Nominal	PAD – peripheral arterial diseases (T,F)
14	PRE30	Nominal	Smoking (T,F)
15	PRE32	Nominal	Asthma (T,F)
16	AGE	Numeric	Age at surgery
17	Risk1Y	Nominal	1 year survival period – (T)rue value if died (T,F)

were adopted and applied to check whether they had a positive effect in increasing the performance accuracy of the supervised algorithms used in this paper. Indeed, and through the obtained results, it was proved that after their application, there was a relative increase in the performance of the algorithms [10].

3.3.1. GR
It is an enhancement version of IG. It calculates the GR in connection with the class. Whereas the IG selects the feature with a huge number of value, this method's objective is to maximize the feature IG while decreasing the value numbers [11].

$$Gain\ Ratio\ (Feature) = \frac{Gain\ (Feature)}{SplitInfo\ (Feature)} \quad (1)$$

In the following, the value for splitting information is shown. It is the result of splitting the training dataset D into v partitions, each corresponding to v outcomes on the attribute feature:

$$SplitInfo\ (A) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} . log_2 \frac{D_j}{D} \quad (2)$$

3.3.2. IG
The attribute values are evaluated by the IG method with the calculation of IG concerning the class which calculated the difference in information between cases where the feature's value is known and cases unidentified. Each feature will get an assigned score, indicating how much more information about the class is fetched when that feature is used [11].

$$InfoGain\ (Feature) = H\ (Class) - H\ (Class\ |Feature) \quad (3)$$

Where, H refers to entropy is:

$$H\ (X) = -\sum_{i-1}^{n} P\ (x_i) . log_2 P\ (X_i) \quad (4)$$

3.4. 10-Fold Cross-Validation
Cross-validation is one of the standard machine learning techniques used in Weka workbench. Ten-fold cross-validation is a mechanism for evaluating predictive models by dividing the original dataset into two subsets: The training set and the test set in which the used dataset is randomly divided into 10 equal-sized of subparts, one subpart is kept as validation data for testing, and the remaining nine parts are used as training data. Hence, iterating the cross-validation process 10 times, the results for 10-fold can then be averaged to produce one evaluation. The advantage of this technique is that all the datasets will be used in both training set and testing set [12]. The reason for the selection of the cross-validation technique is that it reduces the variance in the estimation a lot more than the other techniques. Accordingly, the dataset used in this paper has been separated according to this technique. This ensures that we will obtain the necessary estimations as well as monitor the performance of the classifiers.

3.5. Supervised Machine Learning Classifiers
Supervised learning mechanism is a type of machine learning in which machines are trained employing labelled training data. In other words, when the used dataset is divided into Training and testing. The supervised learning mechanism is used on a training dataset consisting of known input data (X)

and output variable (Y) to build a module and implement it to predict the output variables (Y) of the testing data [13]. The following are the supervised learning algorithms that have been used in this paper.

3.5.1. RF

A RF algorithm, as its name suggests, is made up of a large number of individual decision trees that act as a set. Each tree in the RF emerges from the prediction of the class and becomes the class with the most votes the basic principle behind the RF algorithm is a simple but powerful concept – the wisdom of the majority crowd. In data science, the reason the RF model is so successful is that a large number of relatively uncorrelated (trees) models acting as a committee will outperform any of the single-component models. The low correlation coefficient between the models is key. Just like how investments with a low coefficient of correlation are aggregated, uncorrelated models can produce aggregate forecasts that are more accurate than any individual forecasts. The reason for this wonderful effect is that trees protect each other from their mistakes (as long as they don't all err in the same direction constantly). While some trees may be wrong, many others will be right so that the trees as a group can move in the right direction [14]. The mathematical formula of the algorithm is as follows [15].

$$RFfi_i = \frac{\sum_{j \in all\ trees} norm\ fi_{ij}}{T} \tag{5}$$

Where, RFfi sub(i)= the significance of feature i calculated from all trees in the RF model

normfi sub(ij) = the normalized feature importance for I in tree j.

3.5.2. J48

The process of classification using a decision tree uses gain information to divide the tree. The first step is to gain information for each attribute. The attribute with the largest amount of IG will be the node root of the decision tree. The decision tree technique aims to divide the database with a specific goal that has already been determined, and the presence of a certain element in one of the groups, which is represented here by the branches, becomes a result because it achieved the series of conditions set down to this branch and not only because it is similar to the rest of the elements [16]. Although, it has not been defined similarity in this case. The J48 and the algorithms that are used to produce it can be complex, but the results that lead to it can be shown in a simple, easy-to-understand form, and with a high level of utility. The algorithm steps are as follows:

First: If the instances belong to the same class, the leaf is tagged with a comparable class.

Second: The prospective data for each attribute will be calculated, and the data gain from the attribute test will be calculated.

Third: Eventually, based on the current selection parameter, the best attribute will be selected.

3.5.3. Naive Bayes

It is a classification model in machine learning fields which based on probability. A Naive Bayesian model is simple to construct and does not require iterative parameter estimation, making it ideal for huge datasets [17]. From P(c), P(x), and P(x|c), the Bayes theorem may be used to get the posterior probability, P(c|x). The effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors, according to the Naive Bayes classifier. The following is the formula of the model.

$$P(c \mid x) = \frac{P(x \mid c) P(c)}{P(x)} \tag{6}$$

$P(c \mid X) = P(x_1 \mid c)\ P(x_2 \mid c).P(x_n \mid c)\ P(c)$

$P(c \mid x)$: Rear probability of class(target)given predictor(attribute).

$P(c)$: The prior probability of class.

$P(x \mid c)$: Likelihood which is the probability of predictor given class.

$P(x)$: The prior probability of predictor.

3.5.4. Multilayer perceptron

It is a category of feedforward ANN which creates a set of outputs from a set of inputs. The perceptron, which comprises numerous inputs Xi multiplied by a scalar value known as weight Wij and a bias bj, was one of the earliest PEs constructed [18]. A specified activation function f is used to process the acquired result, which may be explained as follows:

$$Y_j = f\left[\sum_i (W_{ij} * X_{i} + b_i)\right] \tag{7}$$

$$Y_j = f\left[\sum_i (W_{ij} * X_{i} + b_i)\right] \tag{8}$$

The hyperbolic tangent function tanh, which is represented as follows, is the most frequent activation function f utilized in perceptron.

$$\tan h\left(x\right) = 2.\sigma\left(2x\right) - 1,$$

$$where\ \sigma\left(x\right)is\ \sigma\left(x\right) = \frac{e^x}{1 + e^x} \tag{9}$$

The MLP network is used to solve nonlinear separation issues by connecting numerous perceptions in one or more hidden layer topologies. The aim is to discover the error function with the lowest possible error in proportion to the connection weights. The error function is explained as follows:

$$E = \frac{1}{2}\sum_{j\in M}(y^{\hat{}}_m - y_m)^2 \tag{10}$$

with $y^{\hat{}}_m$ being the desired output of m'th y_m.

3.5.5. SVM

The SVM algorithm classifies data for two divisions by taking input data and generating output predictably. The best way for implementing this technique is to build a model to text corpuses while any training sample belonged to one of the classes. After that, the data will be divided into two categories with the way of constructing an N-dimensional hyperplane. To separate data, SVM will build two hyperplanes but they should be paralleled in both sides of the hyperplane while the separated hyperplane will increase the space between other hyperplanes [19]. SVM is capable of conducting regression analyze and extending it while performing a numerical calculation. The formula of the algorithm is shown below:

$$K\left(x,y\right) = \left(x.y + c\right) \tag{11}$$

4. EXPERIMENTS AND RESULTS

In machine learning, and specifically in the field of data classification, there are many commonly accepted criteria for measuring the classification performance for the machine learning algorithms. In this research, the scales shown in the following tables were used to explain the difference in the performance of the algorithms used to classify the data. Then, the performance of each algorithm is compared before and after applying each of the classifier feature selection algorithms GR and IG.

In general, through the results obtained in Tables 2 and 3 with Figs 2-5, it is clear that there is a difference in the stability and instability in the classification performance of algorithms

with or without ranker feature selections in the process of classifying thoracic surgery datasets. To begin with regard
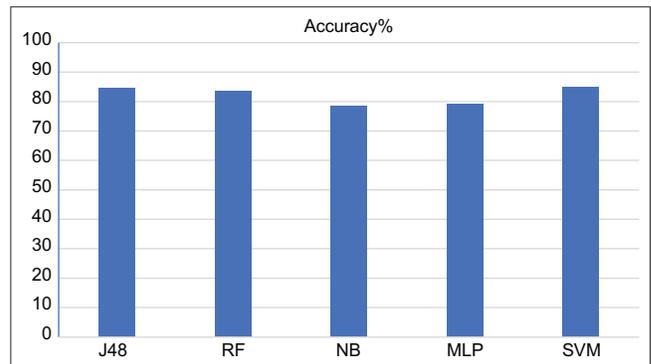

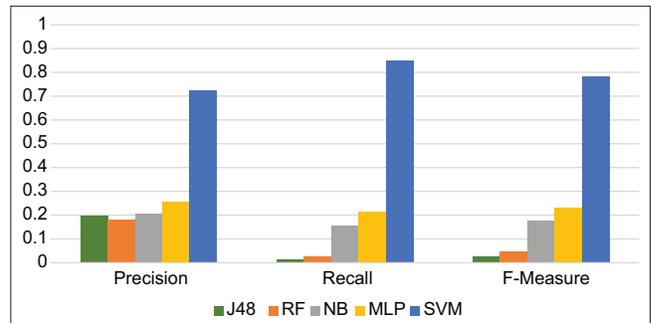**Fig. 2.** Accuracy of the classifiers before feature selections.

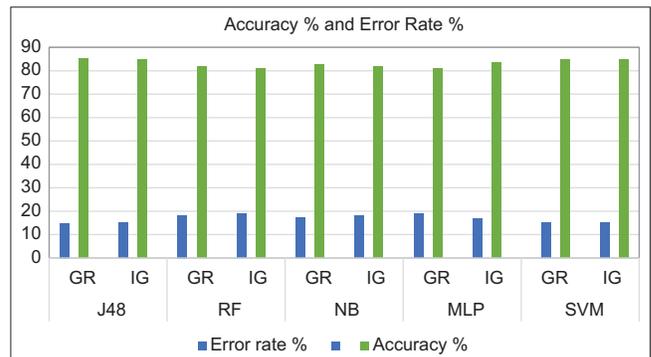**Fig. 3.** Precision/recall and F-measure of the classifiers before feature selections.

Fig. 4. Accuracy and error rate of the classifiers after using feature selections.

TABLE 2: Performance measurements before implementing ranker attribute evaluators

Supervised algorithms	Precision	Recall	F-measure	Accuracy
J48	0.200	0.014	0.027	84.46%
RF	0.182	0.029	0.049	83.62%
NB	0.208	0.157	0.179	78.51%
MLP	0.259	0.214	0.234	79.14%
SVM	0.724	0.849	0.782	84.89%

TABLE 3: Performance measurements after implementing ranker attribute evaluators (Gr)/(IG)

Performance measurements	J48		RF		NB		MLP		SVM	
	GR	IG	GR	IG	GR	IG	GR	IG	GR	IG
Precision	0.724	0.724	0.750	0.751	0.744	0.745	0.767	0.799	0.724	0.724
Recall	0.851	0.849	0.817	0.811	0.828	0.819	0.811	0.834	0.849	0.849
F-measure	0.783	0.782	0.777	0.776	0.777	0.776	0.785	0.811	0.782	0.782
Error rate %	14.893	15.106	18.297	18.936	17.234	18.085	18.936	16.595	15.106	15.106
Accuracy %	85.106	84.893	81.702	81.063	82.766	81.914	81.063	83.404	84.893	84.893

TABLE 4: Classification/time measurements before implementing ranker attribute evaluators

Classification measurements	J48	RF	NB	MLP	SVM
Correctly classified instances	397	393	369	372	399
Incorrectly classified instances	73	77	101	98	71
Time (milliseconds)	30	210	9	1820	90

TABLE 5: Classification/Time measurements after implementing ranker attribute evaluators (GR)/(IG)

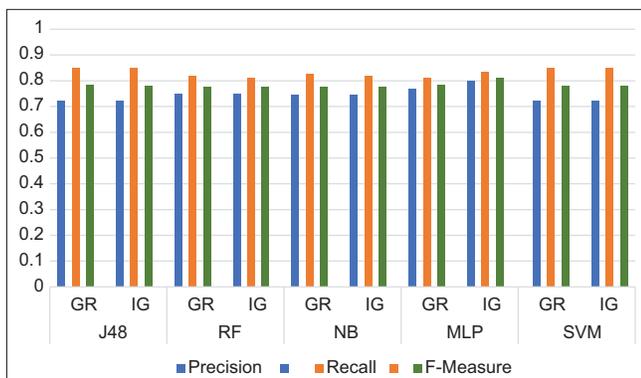| Classification measurements	J48		RF		NB		MLP		SVM	
	GR	IG	GR	IG	GR	IG	GR	IG	GR	IG
Correctly classified instances	400	399	384	381	389	385	381	392	399	399
Incorrectly classified instances	70	71	86	89	81	85	89	78	71	71
Time (milliseconds)	40	10	140	90	10	**9**	1150	1290	30	40

Fig. 5. Precision/recall and F-measure of the classifiers after ranker evaluators.

in J48, NB, and MLP algorithms, we noticed an increment in accuracy of the classification performance, in which the accuracy of J48 is 84.46% without using ranker feature selections, as shown in Table 2 and Fig. 2, this performance has been improved using ranker feature selections GR and IG to 85.106% and 84.893%, respectively, as shown in Table 3 and Fig. 4. Furthermore, the classification performance accuracy of NB is 78.51% without ranker, as shown in Table 2 and Fig. 2, the performance is raised with ranker GR and IG to 82.766% and 81.914%, respectively, as shown in Table 3 and Fig. 4. Moreover, the classification performance

accuracy of MLP is 79.14% without ranker feature selections, as shown in Table 2 and Fig. 2, this accuracy is enhanced with ranker GR and IG to 81.063% and 83.404%, respectively, as shown in Table 3 and Fig. 4.

Another point to consider is with regard to the RF algorithm, we notice a decrement in performance accuracy of RF which was 83.62% without ranker feature selections, as shown in Table 2 and Fig. 2, the accuracy is raised after employing ranker GR and IG to 81.702% and 81.063%, respectively, as shown in Table 3 and Fig. 4. Whereas, in testing SVM algorithm, there are no changes observed in the accuracy during classification as it remains equal in both cases and its performance did not change with both feature selections, the accuracy without ranker selections was 84.89%, as shown in Table 2 and Fig. 2, and it remains stable with no any effectiveness with ranker selections with accuracy 84.89% for both feature selection algorithms GR and IG, as shown in Table 3 and Fig. 4.

In Table 4, it is clear that SVM is the most accurate algorithm in classifying instances correctly with 399 instances out of a total of 470 instances units without ranker feature selections. However, it is not the fastest in constructing the model, as it took 0.09 seconds for classifying the whole dataset records.

Besides, MLP is the slowest algorithm among the other algorithms in the classification process as it took 1.82 seconds without using ranker feature selections. In contrast, NB is the lowest in classifying instances correctly with 369 instances out of a total of 470 instances without using franker feature selections. However, it is the fastest in building the model, as it took 0.00 seconds to classify the whole dataset records.

In Table 5, a drastic change can be observed, it is clear that J48 is the most accurate algorithm in classifying instances correctly with 400 instances out of a total of 470 instances units with ranker feature. However, it is one of the fastest algorithms in constructing the model using IG which took 10 milliseconds for classifying the whole dataset records. In contrast, both RF using IG and MLP using GR are the lowest in classifying instances correctly with 381 instances out of a total of 470 instances without ranker feature. Furthermore, MLP remained the slowest in building the model, as it took 1290 milliseconds to classify the whole dataset records using IG. The NB remained the fastest algorithm among the others in the classification models as it took 0.00 seconds with IG. In contrast, both RF using IG and MLP using GR are the lowest in classifying instances correctly with 381 instances out of a total of 470 instances without franker feature selections. However, MLP remained the slowest in building the model, as it took 1290 milliseconds in classifying the whole datasets using IG. Finally, NB remained the fastest algorithm among the other algorithms in classifying the dataset as it took 9 milliseconds with using IG.

5. CONCLUSION

The comparison made in this paper showed a significant effect of the ranker features on supervised classification algorithms. Through the obtained results, we concluded that the use of ranker feature selections leads to improving the classification performance of particular algorithms, as done with J48, MLP, and NB algorithms. In contrast, ranker feature selection reduced the performance of RF. Moreover, specific algorithms such as SVM remained stable before and after ranker feature selection concerning classification performance. Similarly, as for the speed of building the model, the NB algorithm did not change its speed in both cases by recording the least time for data classification and the fastest among the other algorithms, 9 milliseconds. Eventually, the highest performance in the accuracy of classification was the J48 algorithm using GR, which amounted to 85.1%. Other feature selection algorithms can be employed to improve the used algorithms' performance in future work.

REFERENCES

[1] S. Prabha, S. Veni and S. Prabha. "Thoracic Surgery analysis using data mining techniques". *International Journal of Computer Technology and Applications*, vol. 5, no. 1, pp. 578-586, 2014.

[2] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis and D. I. Fotiadisa. "Machine learning applications in cancer prognosis and prediction". *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8-17, 2015.

[3] A. S.Dusky and L. M. El Bakrawy. "Improved prediction of post-operative life expectancy after Thoracic Surgery". *Advances in Systems Science and Applications*, vol. 16, no. 2, pp. 70-80, 2016.

[4] J. Joshi, R. Doshi and J. Patel. "Diagnosis of breast cancer using clustering data mining approach". *International Journal of Computer Applications*, vol. 101, no. 10, pp. 13-17, 2014.

[5] S. Vanaja and K. R. Kumar. "Analysis of feature selection algorithms on classification: A survey". *International Journal of Computer Applications*, vol. 96, no. 17, pp. 29-35, 2014.

[6] M. Zięba, J. Tomczak, M. Lubicz and J. Świątek. "Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients". *Applied Soft Computing*, vol. 14, pp. 99-108, 2014.

[7] M. U. Harun and N. Alam. "Predicting outcome of thoracic surgery by data mining techniques". *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 1, pp. 7-10, 2015.

[8] M. Lubicz, K. Pawelczyk, A. Rzechonek and J. Kolodziej. "*UCI Machine Learning Repository: Thoracic Surgery Data Data Set*", 2021. Available from: https://archive.ics.uci.edu/ml/datasets/thoracic+surgery+data [Last accessed on 2021 Oct 08].

[9] S. Xu. "*Machine Learning-Assisted Prediction of Surgical Mortality of Lung Cancer Patients*". The IEEE International Conference on Data Mining, 2019.

[10] S. Subbiah and J. Chinnappan. "An improved short term load forecasting with ranker based feature selection technique". *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 5, pp. 6783-6800, 2020.

[11] D. El Zein and A. Kalakech. "*Feature Selection for Android Keystroke Dynamics*". 2018 International Arab Conference on Information Technology, 2018.

[12] H. Talabani and A. V. C. Engin. "*Performance Comparison of SVM Kernel Types on Child Autism Disease Database*". International Conference on Artificial Intelligence and Data Processing, 2018.

[13] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi and J. Akinjobi. "Supervised machine learning algorithms: Classification and comparison". *International Journal of Computer Trends and Technology*, vol. 48, no. 3, pp. 128-138, 2017.

[14] M. Rathi and V. Pareek. "Spam mail detection through data mining a comparative performance analysis". *International Journal of Modern Education and Computer Science*, vol. 5, no. 12, pp. 31-39, 2013.

[15] J. Wong. "*Decision Trees Medium*", 2021. Available from: https://towardsdatascience.com/decision-trees-14a48b55f297 [Last accessed on 2021 Oct 08].

[16] A. Yadav and S. Chandel. "Solar energy potential assessment of Western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model". *Renewable Energy*, vol. 75, pp. 675-693, 2015.

[17] K. Vembandasamy, R. Sasipriya and E. Deepa. "Heart diseases

detection using Naive Bayes algorithm". *International Journal of Innovative Science, Engineering and Technology*, vol. 9, no. 29, pp. 441-444, 2015.

[18] M. Khishe and A. Safari. "Classification of sonar targets using an MLP neural network trained by dragonfly algorithm". *Wireless Personal Communications*, vol. 108, no. 4, pp. 2241-2260, 2019.

[19] H. Talabani and A. V. C. Engin. "*Impact of Various Kernels on Support Vector Machine Classification Performance for Treating Wart Disease*". International Conference on Artificial Intelligence and Data Processing, 2018.