

Breast cancer detection using machine learning approaches: a comparative study

Muawia A. Elsadig¹, Abdelrahman Altigani², Huwaida T. Elshoush³

¹Deanship of Scientific Research, Imam Abdulrahman Bin Faisal University, Dammam, Kingdom of Saudi Arabia

²Department of Computer Science, College of Computer Science and Information Technology,
Imam Abdulrahman Bin Faisal University, Dammam, Kingdom of Saudi Arabia

³Department of Computer Science, Faculty of Mathematical Sciences and Informatics, University of Khartoum, Khartoum, Sudan

Article Info

Article history:

Received Nov 6, 2021

Revised Jul 21, 2022

Accepted Aug 14, 2022

Keywords:

Breast cancer detection

Deep learning

Machine learning

Classification algorithms

Breast cancer diagnosis

ABSTRACT

As the cause of the breast cancer disease has not yet clearly identified and a method to prevent its occurrence has not yet been developed, its early detection has a significant role in enhancing survival rate. In fact, artificial intelligent approaches have been playing an important role to enhance the diagnosis process of breast cancer. This work has selected eight classification models that are mostly used to predict breast cancer to be under investigation. These classifiers include single and ensemble classifiers. A trusted dataset has been enhanced by applying five different feature selection methods to pick up only weighted features and to neglect others. Accordingly, a dataset of only 17 features has been developed. Based on our experimental work, three classifiers, multi-layer perceptron (MLP), support vector machine (SVM) and stack are competing with each other by attaining high classification accuracy compared to others. However, SVM is ranked on the top by obtaining an accuracy of 97.7% with classification errors of 0.029 false negative (FN) and 0.019 false positive (FP). Therefore, it is noteworthy to mention that SVM is the best classifier and it outperforms even the stack classifier.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Muawia A. Elsadig

Cybersecurity, Deanship of Scientific Research, Imam Abdulrahman Bin Faisal University

Dammam, Kingdom of Saudi Arabia

Email: muawiasadig@yahoo.com

1. INTRODUCTION

The rapid spread of breast cancer is clearly noticeable, especially in some developed countries. It is considered to be one of the most significant reasons of death among females [1], [2]. In other words, in most cancer-affected women, breast cancer is the most frequent cancer [3] and the primary cause of death [4]. In 2018, about two million new cases were reported [5]. The complications of the breast cancer diagnosis process and its late discovery are the reasons behind the low survival rate. Its early detection prevents its progression and thus reduces the risk. The early detection and treatment of this disease improves the survival rate [6]–[10]. However, the absence of the breast cancer symptoms at the beginning stage of this disease makes its early discovery harder [11]. Therefore, the computer-aided diagnosis (CAD) methods are highly required as they are very effective in prediction process. CAD methods have the potential of being valuable tools to help radiologists [12].

In comparison with other types of cancer and regarding the number of new cases, a recent report has shown that breast cancer among females received the highest number and it is the leading cause of death [13]. Figures 1 and 2 show the mortality and incidence respectively for female breast cancer in 2018. It may

be clearly noticed that female breast cancer dominates in both cases, number of new cases and death. Moreover, other sources indicated that woman's breast cancer is said to be the second largest cause of death worldwide [14]; therefore, computer-aided detection techniques which help radiologists in detecting abnormal behavior are much needed to enrich survival rate through early detection of this deadly disease. Early detection and treatment are important for patients with breast cancer. It is possible for 95% of patients with early breast cancer to be cured completely [15].

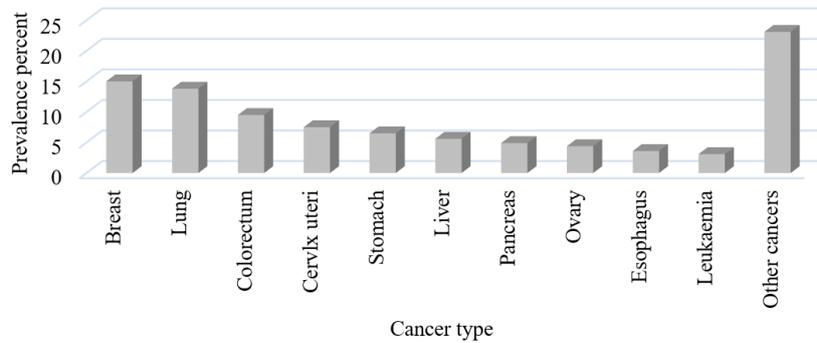


Figure 1. The death rate of some common cancers among females worldwide for the year 2018 [13]

Moreover, to this end, a method to prevent the occurrence of breast cancer has not yet been presented therefore cancer tumor in the beginning periods could be treated effectively [16], [17]. This reflects the need to develop adequate methods to support the early diagnosis of this disease. Developing such methods is mainly based on taking benefits of advanced computing techniques to enrich the diagnostic capabilities.

Generally, CAD assists radiologists to screen patient images and therefore increase the detection accuracy. It has been widely used to diagnose different kinds of disease [18]. Interested readers are referred to the study in [19] which sheds light on CAD methods that use mammograms to diagnosis or detect breast cancer.

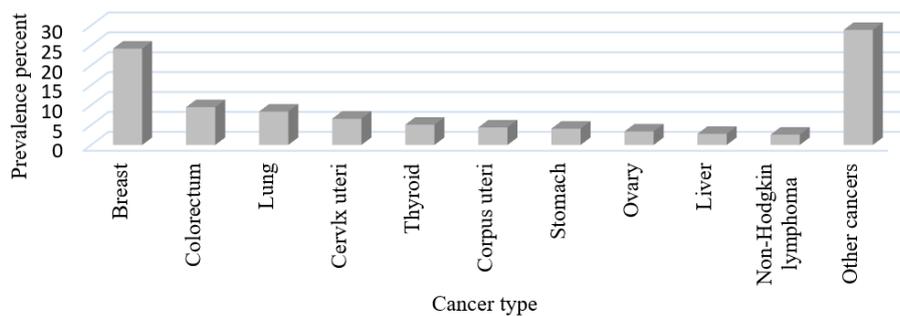


Figure 2. The rate of new cases of some common cancers among females worldwide for the year 2018 [13]

In this paper an enhanced version of Wisconsin breast cancer diagnosis (WBDC) dataset has been used to train and test eight popular machine learning models that are commonly used in medical image classification. One of these classifiers is an ensemble classifier based on stacking techniques. It takes the output of all other classifiers to form an ensemble classification model. The dataset is thoroughly improved through applying five feature selection methods that include Chi-square (X2), ReliefF, Analysis of variance (ANOVA), Gini index, and Gain ratio. The purpose of applying these efficient feature selection methods is to reduce the data dimensionality and therefore neglect all unweighted features and consider only features that have considerable impact in the classification process. When a model learns from irrelevant features, this will affect its accuracy, so an important key of success is to have only the relevant features and ignore others. Using the relevant features reduces the computational cost and improves model performance.

A thorough review on the common classification models to predict breast cancer occurrence has been given by this paper. These classifiers are carefully selected, taking into consideration their capabilities to attain high prediction accuracy. In addition, these classifiers include two classifiers that are based on an ensemble approach; one uses bagging technique and the other uses stacking technique. These variations are being considered to base our work on different techniques and scenarios to enrich our findings. These classifiers will be trained and tested under an enhanced version of a common dataset (which is another contribution of this work) to pick the classifiers that are capable of achieving high detection accuracy, and this is the main aim of this paper. Generally, we can conclude that many phases of enhancement have been applied to reflect trusted and reliable findings, and these phases will be thoroughly illustrated throughout this paper.

The paper is organized: the spread of breast cancer worldwide has been thoroughly discussed in the first section and many statistics supporting that have been given. In addition, this section sheds light on computer-aided techniques and their important role in detection and diagnosis capabilities. The following section provides a review of relevant classification models that are commonly used to predict breast cancer. The limitations and achievements of these models are given. Mainly, this section gives more focus on eight classification methods that have ensured their capabilities in obtaining high performance to predict breast cancer tumors. Our proposed method is given in section 3 which gives a clear description of the method and highlights its effectiveness in achieving the expected goals. Comprehensive details about our enhanced dataset, the applied feature selection methods, implementation tool and validation techniques are fully illustrated by this section. Section 4 shows the experimental work and thoroughly discusses the results obtained. Finally, the conclusion is summarized in section 5.

2. RELATED WORK

It is noteworthy that computer aided techniques such as machine learning and deep learning methods are widely becoming common techniques in the medical field. The way to examine and evaluate patient data is one of the most important factors that influences the diagnostic process, so the automation of this process can greatly help to gain accurate results. Machine learning classification models have a great influence in minimizing the errors that can be caused by inexperienced physicians and they are capable of giving accurate results in a short time. However, some problems arise with machine learning techniques that may happen due to using inefficient validation techniques, inadequate classification models, and redundant unweighted features. In addition, and due to the complexities of breast tissues, classification and prediction of breast cancer in medical imaging is said to be a critical task [20]. Therefore, utilizing machine learning and deep learning techniques is a vital part in improving the diagnosis process, if these techniques are implemented perfectly considering all phases of work flow starting from using efficient tools for image reprocessing and selecting proper features, up to applying adequate validation methods. The rest of this section gives a comprehensive review of the classifiers that are selected to be under test by this work.

Gao *et al.* [21] pointed out the potential solutions that CAD techniques can offer compared to traditional methods. Generally, the advanced development of computer techniques in machine learning, data mining, and deep learning, has played a huge role in improving clinical care systems and supporting early diagnosis of many diseases and therefore the survival rate.

Segmentation is one of the essential CAD system components [22] that play an important role in classification accuracy. A comparative analysis for three segmentation techniques that include: spatial fuzzy c-means (SFCM) with level set, selective level set, and spatial neutrosophic distance regularized level set (SNDRLS) was conducted in [23]. The performance of these techniques was evaluated on a dataset of breast cancer images. Their results showed that SFCM with level set works effectively to extract cancerous cells compared to other techniques. Kaushal and Singla [24] proposed an automated segmentation technique for breast cancer images. The authors listed some advantages of this technique that lead to enhance segmentation process and therefore the proposed technique can identify cancerous cells correctly. The proposed technique was evaluated and showed its effectiveness.

An experiment that uses the logistic regression (LR) classifier model has been carried out. The dataset used in this experiment is Wisconsin diagnosis breast cancer (WDBC), and the authors highlight that the proper selection for feature combination is important to improve classification accuracy. Their findings show that the logistic classifier accuracy can reach up to 96.5% in case of selecting maximum perimeter and maximum texture as two characteristics [25]. On the other hand, another study has compared LR and decision tree classifiers to point out that the decision tree (DT) algorithm has performed slightly better than LR; however, both classifiers obtained high accuracy rate [26].

A study was conducted to compare two classifiers-multi-layer perceptron (MLP) and radial basis function (RBR). Both are neural network classifiers. Their results reflect that MLP has outperformed ring-

between-ring (RBR) by attaining high accuracy compared to RBR [27]. Zheng *et al.* [28] outlined that neural network classifiers have become a popular method to classify cancer data. In addition to that, another study pointed out that MLP is the most effective classifier [29]. This result is based on their results obtained by comparing three classifiers that include MLP, naïve Bayes (NB) and C4.5 tree.

NB classification model is built on Bayes theorem. A comparison study was conducted on three classifiers to indicate that MLP is more effective; however, it has poor interpretability. In spite of violation of one of NB's assumption, NB has attained good performance with good interpretability [29].

Karabatak [30] indicates that NB is one of the most powerful classification models; however, it has some drawbacks. To overcome these drawbacks, Karabatak proposed a new NB classifier (weighted NB). The author claimed that, the conducted experiments showed that weighted NB has obtained better accuracy than the regular NB. On the other hand, the researcher stated some drawbacks for this new classifier to be overcome in future research.

Showrov *et al.* [31] conducted a comparative study among three classifiers-artificial neuron network (ANN), support vector machine (SVM), and NB-using WDBC dataset of nine features for breast cancer prediction. In terms of classification accuracy, their results reflected that linear SVM topped the other classifiers while radius basis function neural network (RBF NN) comes next and then Gaussian naïve Bayes. Moreover, another study has shown that support vector machine has precise diagnosis capacity [32].

Using extracted features from mammography images-after applying some image processing techniques Al-Hadidi *et al.* [33] trained and tested two classification algorithms using MATLAB software and reported their results accordingly. These two classifiers are logistic regression (LR) and back propagation neural network (BPNN). Their observation reflects that a good regression value exceeding 93% has been obtained using BPNN with no more than 240 features.

The random forest (RF) classifier is an ensemble approach which contains multiple algorithms. Each one can be implemented in a decision tree. The combined result of these multiple decision trees leads to enhanced classification accuracy [34]. In other words, the RF is a combination of multiple decision trees that represent an ensemble classifier to promote performance and prediction accuracy. A classifier that is based on RF was developed in [35]. Their model has been trained using two different datasets and has obtained promising results with high accuracy.

Using the WDBC, a comparison study has been carried out for three classifiers, including NB, RF and k-nearest-neighbor (KNN). These classifiers were trained and tested to examine their prediction accuracy of breast cancer tumors based on the aforementioned dataset. Accordingly, the authors conducting this comparison observed that KNN outperformed the other classifiers as it has obtained higher accuracy compared to NB and RF classifiers and generally all classifiers have obtained detection accuracy rates above 94%. The KNN classifier does just obtain the best accuracy among the others, it also outperforms them in terms of precision and F1 score values [36]. Price and Lindqvist indicate that ANN classifier has performed well when applying feature selection methods compared to SVM, NB, and decision trees classifiers. It attained considerable improvement in its performance that reaches 51% increase [37].

Using a small dataset of 275 instances, the authors in [38] compared the performance of two machine learning classifier models, extreme gradient boosting (XGBoost) and RF. Their results show that RF has performed better than XGBoost in terms of detection accuracy for breast cancer; however, using a small dataset may reflect unreliable results, so the authors stated that a large dataset is required to support their findings. A recent study compared nine classification models that include LR, Gaussian NB, RBF SVM, linear SVM, DT, RF, XGBoost, KNN and gradient boosting. These models were trained and tested using Wisconsin diagnosis cancer dataset. The obtained results indicate that KNN for supervised learning and LR for semi-supervised learning achieved the highest accuracy [39].

Ensemble learning approach is one of the most partial ways to offer a trade-off between variance and bias. Many studies show that combining single classifiers to build an aggregated classification model can improve classification performance compared to the performance that can be obtained by any one of these classifiers. The three basic techniques of ensemble classification are stacking, boosting and bagging. In the stacking approach, the outputs from different classification models are to be aggregated into a new dataset [40]. Readers interested for more information are referred to [41]–[46].

Hosni *et al.* [47] based on their review study, highlighted that most classifiers that are frequently used to build ensemble classification models include artificial neural networks, support vector machines, and trees. In addition, the dataset most frequently used by researchers is WBCD. The results of this thorough review have motivated us to base our work on WBCD which is a trusted dataset by the research community.

According to our mentioned survey work, seven classification models have been chosen to be under investigation throughout this work. Additionally, an ensemble classifier based on stacking technique will be used to include all the aforementioned seven classifiers. Therefore, our study will include eight classifiers: RF, SVM, logistic regression, DT, KNN, MLP, NB and stack.

3. RESEARCH METHOD

Eight classification models have been selected carefully to be under test throughout this work. These classifiers are: LR, ANN, RF, SVM, MLP, NB, DT, and stack. Two of these classifiers are ensemble classifiers that are based on different ensemble techniques: RF classifier and stack classifier. RF depends on bagging technique. It is a collection of tree-structured classifiers. While the stack classifier is based on stacking technique, it takes the outputs of different classification models. This variation of choosing different classifiers that are based on different techniques and scenarios is to boost up our findings and therefore lead to trusted results. Each classifier has been trained and tested under four different train-test sets that taken from the enhanced dataset that described below.

3.1. Dataset

In this work an enhanced dataset has been used to train and test the classifiers that are under test. The enhanced dataset has been developed based on the WBCD, the well-known dataset that available in UCI machine learning repository website [48]. The WBCD include 569 instances with 30 features. The enhanced process that applied to this dataset has reduced the number of features to only 17 features. The features have been reduced by applying five feature selection techniques and accordingly the top 17 features that have considerable wight have been selected and the other features (redundant and unweighted features) have been neglected. This process is considered a huge contribution of this paper which results in improving the classification accuracy and at the same time reduces the classification process overheads. It is so smart to gain high accuracy and to reduce the computation overhead simultaneously. Figure 3 shows the revised dataset with 17 features.

Four different sets of our revised dataset have been used to train and test all classifiers. These sets are: set 1: 60% to train and 40% for test; set 2: 70% to train and 30% for test; set 3: 80% to train and 20% for test; and set 4: 90% to train and 10% for test. It is commonly known that applying different training sizes leads to having in-depth experiments that give ensured trusted results.

| 1 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 | F11 | F12 | F13 | F14 | F15 | F16 | F17 |
|----|-------|-------|-------|-------|-------|--------|---------|-------|-------|---------|--------|--------|-------|-------|---------|--------|-------|
| 2 | 17.33 | 25.38 | 184.6 | 122.8 | 17.99 | 0.2654 | 0.1471 | 2019 | 1001 | 0.3001 | 0.7119 | 1.095 | 153.4 | 8.589 | 0.2776 | 0.6656 | 10.38 |
| 3 | 23.41 | 24.99 | 158.8 | 132.9 | 20.57 | 0.186 | 0.07017 | 1956 | 1326 | 0.0869 | 0.2416 | 0.5435 | 74.08 | 3.398 | 0.07864 | 0.1866 | 17.77 |
| 4 | 25.53 | 23.57 | 152.5 | 130 | 19.69 | 0.243 | 0.1279 | 1709 | 1203 | 0.1974 | 0.4504 | 0.7456 | 94.03 | 4.585 | 0.1599 | 0.4245 | 21.25 |
| 5 | 26.5 | 14.91 | 98.87 | 77.58 | 11.42 | 0.2575 | 0.1052 | 567.7 | 386.1 | 0.2414 | 0.6869 | 0.4956 | 27.23 | 3.445 | 0.2839 | 0.8663 | 20.38 |
| 6 | 16.67 | 22.54 | 152.2 | 135.1 | 20.29 | 0.1625 | 0.1043 | 1575 | 1297 | 0.198 | 0.4 | 0.7572 | 94.44 | 5.438 | 0.1328 | 0.205 | 14.34 |
| 7 | 23.75 | 15.47 | 103.4 | 82.57 | 12.45 | 0.1741 | 0.08089 | 741.6 | 477.1 | 0.1578 | 0.5355 | 0.3345 | 27.19 | 2.217 | 0.17 | 0.5249 | 15.7 |
| 8 | 27.66 | 22.88 | 153.2 | 119.6 | 18.25 | 0.1932 | 0.074 | 1606 | 1040 | 0.1127 | 0.3784 | 0.4467 | 53.91 | 3.18 | 0.109 | 0.2576 | 19.98 |
| 9 | 28.14 | 17.06 | 110.6 | 90.2 | 13.71 | 0.1556 | 0.05985 | 897 | 577.9 | 0.09366 | 0.2678 | 0.5835 | 50.96 | 3.856 | 0.1645 | 0.3682 | 20.83 |
| 10 | 30.73 | 15.49 | 106.2 | 87.5 | 13 | 0.206 | 0.09353 | 739.3 | 519.8 | 0.1859 | 0.539 | 0.3063 | 24.32 | 2.406 | 0.1932 | 0.5401 | 21.82 |

F1: texture_worst
F5: radius_mean
F9: area_mean
F13: area_se
F17: texture_mean

F2: radius_worst
F6: concave points_worst
F10: concavity_mean
F14: perimeter_se

F3: perimeter_worst
F7: concave points_mean
F11: concavity_worst
F15: compactness_mean

F4: perimeter_mean
F8: area_worst
F12: radius_se
F16: compactness_worst

Figure 3. The enhanced dataset

3.2. The feature selection approach

Using an adequate feature selection approach is an important step to improving classification accuracy. This process reduces the features by taking into consideration the features with impact weight into the classification process and ignoring other features that have not. It has two benefits: it boosts classifier predictability and reduces its computational overheads. There are many feature selection methods that are commonly known and have great impact in improving classification performance. In this paper five selection methods are applied to our dataset to take only features that received considerable wights by these methods. These feature selection methods include Chi-square (X2) [49]–[51], ReliefF [52], [53], ANOVA [54], Gini index [55], [56], and Gain ratio [57]. Choosing these feature selection methods allows these methods to use different metrics to select optimal features. For example, Chi-square (X2) computes the chi-score to rank the features, and information gain is ranking the features depend on their value.

Using these five feature selection methods, the weight for any given feature has been calculated, and then the feature has been ranked accordingly. A feature that received high rank by all feature selection methods will be selected automatically, while for the other features that received variant ranking, the average

of the weights will be calculated to select only those have considerable average of weights. By applying this scenario, only 17 features have received the best ranking by the five features selection methods and the other feature received poor weights and therefore has been neglected. In other words, the average is calculated when the features selection methods give variant weights, but in case they all agreed to give high rank for a given feature, then this feature will be selected automatically. Table 1 shows a part of the calculated weights that were given by the five selection feature methods for two features of the dataset. According to this important phase of enhancement, the following 17 features were selected as they received the top weights compared to the rest of the features: *texture_worst*, *radius_worst*, *perimeter_worst*, *perimeter_mean*, *radius_mean*, *concave points_worst*, *concave points_mean*, *area_worst*, *area_mean*, *concavity_mean*, *concavity_worst*, *radius_se*, *area_se*, *perimeter_se*, *compactness_mean*, *compactness_worst*, *texture_mean*. This method perfectly enhances our dataset and therefore reflects a high degree of detection accuracy.

Table 1. Example of the wights given by the five features selection methods for two features

| Feature | Chi-square (X2) | ReliefF | ANOVA | Gini index | Gain ratio |
|----------------------|-----------------|---------|---------|------------|------------|
| Concave points_worst | 279.705 | 0.089 | 964.385 | 0.308 | 0.293 |
| Radius_worst | 290.486 | 0.086 | 860.782 | 0.320 | 0.310 |

3.3. Validation methods

Validation is an essential phase for any model to gain acceptance. In our experiments and for attaining realistic and reliable results, the random sampling validation technique has been repeated 10 times for each classification model. The classification accuracy for each model has been calculated through the common accuracy in (1):

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})} \quad (1)$$

where a true positive (TP) refers to positive instances that are predicted correctly by a classification model. A true negative (TN) refers to negative cases that are predicted correctly by a classification model. A false positive (FP) refers to negative cases that are predicted incorrectly by a classification model. A false negative (FN) refers to positive cases that are predicted incorrectly by a classification model. Then the confusion matrix has been used to evaluate our models' performance. The confusion matrix measures the classification error in terms of false negatives (FN) and false positives (FP).

3.4. Models implementation tool

Orange data mining software has been chosen to conduct all the experiments throughout this work. It is a very powerful tool that has huge capabilities to visualize data efficiently and professionally. It is open source that is very useful to implement different machine learning classification algorithms.

4. RESULTS AND DISCUSSION

After a thorough investigation on eight classifiers that are mostly used in classification of medical images, this paper comes up with different findings that include nominating the two classifiers as the best in diagnosing breast cancer; these classifiers are SVM and MLP, while SVM is dominant. SVM has attained the highest classification accuracy followed by MLP. The results illustrated in Table 2 show the accuracy of each classifier. The accuracy is read after investigating each classifier under four different training-testing sets. SVM has outperformed all classifiers under all testing scenarios by obtaining classification accuracies of 97.7%, 97.5%, 97% and 97% over the four sets 80%-20%, 90%-10%, 60%-40%, and 70%-30% respectively. It noticed that the highest accuracy is obtained when the training size is 80%. SVM outperforms two types of ensemble classifiers-stack and RF. Both stack and RF use different ensemble techniques, bagging and stacking respectively. It is noteworthy to mention that SVM has outperformed Stack classifier in spite of the fact that stack classifier takes the outputs of all other classifiers-under our investigation including SVM-to enhance prediction accuracy. MLP and stack come next, next to each other in the second position after SVM.

Moreover, Table 3 shows the classification errors, false positives (FP) and false negatives (FN) for all classifiers under four training-test sets. These results indicate that SVM outperformed other classifiers by casing the least classification errors, 0.029 FN and 0.019 FP. Generally, the classifiers under our investigation can be classified into three groups: Group 1 which include SVM, MLP, and stack (ensemble classifier); this group obtain an excellent prediction accuracy and topped the other groups. Group 2 which include LR and RF. These classifiers achieved very good accuracy while Group 3, which includes KNN, DT

and NB, gained good prediction accuracy and took the bottom level of this classification. Figure 4 clearly reflects that. Based on our findings and this thorough evaluation, the paper concludes that SVM ranks at the top; however, MLP is the closest competitor. SVM outperformed even the stack classifier which is an ensemble classifier that is based on the stacking technique.

Table 2. Comparison of eight classifiers’ accuracy over four training-testing sets

| Classifier | Accuracy | | | |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | Training testing size | Training testing size | Training testing size | Training testing size |
| | 60%-40% | 70%-30% | 80%-20% | 90%-10% |
| KNN | 93.6% | 92.9% | 93.5% | 94.0% |
| DT | 94.2% | 92.3% | 93.2% | 93.9% |
| SVM | 97.0% | 97.0% | 97.7% | 97.5% |
| Random forest | 95.7% | 95.5% | 95.4% | 95.8% |
| MLP (neural network) | 96.9% | 96.5% | 96.8% | 96.8% |
| NB | 93.8% | 93.0% | 93.8% | 93.2% |
| Logistic regression | 95.0% | 94.4% | 95.1% | 96.0% |
| Stack | 97.0% | 96.3% | 96.4% | 97.0% |

Table 3. Classification errors, false positives (FP) and false negatives (FN) for all classifiers under four training-testing sets

| Training set | 60% | | 70% | | 80% | | 90% | |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| | FP | FN | FP | FN | FP | FN | FP | FN |
| DC | 0.036 | 0.095 | 0.050 | 0.123 | 0.054 | 0.090 | 0.036 | 0.105 |
| LR | 0.034 | 0.076 | 0.036 | 0.089 | 0.036 | 0.071 | 0.028 | 0.062 |
| SVM | 0.020 | 0.045 | 0.027 | 0.034 | 0.019 | 0.029 | 0.017 | 0.038 |
| RF | 0.029 | 0.068 | 0.026 | 0.077 | 0.033 | 0.069 | 0.028 | 0.067 |
| NB | 0.055 | 0.074 | 0.064 | 0.080 | 0.060 | 0.067 | 0.061 | 0.080 |
| MLP | 0.027 | 0.038 | 0.034 | 0.036 | 0.028 | 0.038 | 0.017 | 0.057 |
| KNN | 0.035 | 0.112 | 0.036 | 0.130 | 0.042 | 0.105 | 0.036 | 0.100 |
| STACK | 0.023 | 0.042 | 0.028 | 0.053 | 0.031 | 0.045 | 0.025 | 0.038 |

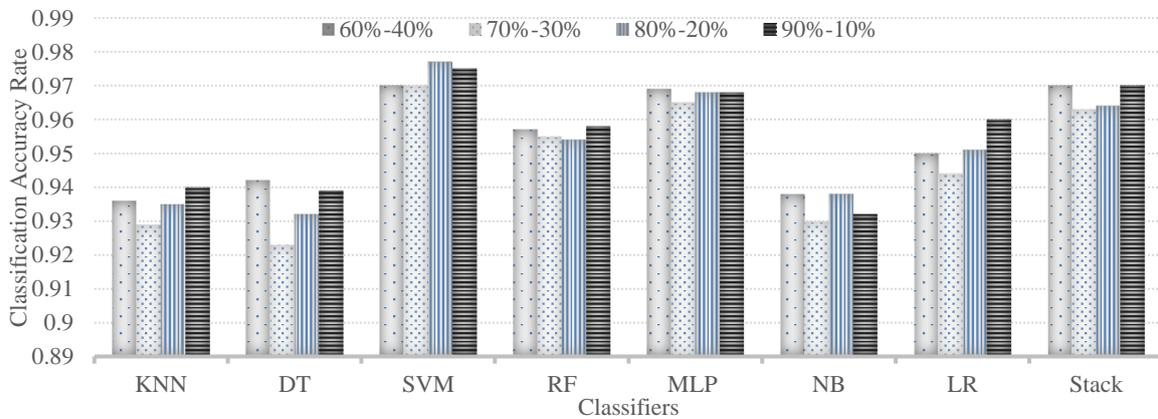


Figure 4. The classification performance in terms of accuracy obtained by each classifier over four training sets: 60%, 70%, 80% and 90%

5. CONCLUSION

It is noteworthy to mention that machine learning techniques have ensured their efficiency to discover and define patterns from large amount of medical images and therefore this will assist to classify and sort out these images accordingly. In return this will highly enrich the detection process. However, choosing a trusted dataset, an adequate machine learning approach, a proper feature selection method and accurate validation technique is a key factor to introduce a reliable and efficient detection scheme. This paper applies weighty enhancements to enrich its findings. Many phases of improvements have been implemented that include, but are not not limited to: i) using an enhanced version of publicly trusted dataset by applying five different feature selection methods and, accordingly, the features which have the greatest influence on the prediction process have been selected to improve the prediction accuracy while the other features that

have no weighted influence are discarded. Decreasing the number of the features reduces the computational overheads. Therefore, this phase of enhancement has two important improvements: i) it enriches classification accuracy and decreases the model computation cost; ii) applying four sets of training-test scenarios to boost the trustworthiness; and iii) using variant evaluation methods.

The main contribution of this paper is ranking the SVM as the best classifier. It obtained a classification accuracy reaching 97.7% with the least classification errors 0.029 false negatives (FN) and 0.019 false positives (FP). This was followed by MLP and stack classifiers. Stack is an ensemble classifier based on stacking technique. Also, the paper presented a comparative study that classifies these classifiers into three levels or groups based on their performance. While excluding the stack classifier-as it is an ensemble classifier that is based on all other classifiers-the top group includes SVM and MLP followed by the next group that includes LR and RF and then the last group that includes DT, KNN and NB. On the other hand, the paper contributes by introducing an enhanced version of the dataset that has been improved by applying the five-feature selection methods to improve prediction accuracy and reduce computational cost. Generally, SVM has topped all classifiers, even the stack classifier which is an ensemble classifier of all classifiers under investigation of this paper.

ACKNOWLEDGEMENTS

The author would like to extend his gratitude and appreciation to the Deanship of Scientific Research of Imam Abdulrahman bin Faisal University to their support for the project No. 2019348.

REFERENCES

- [1] V. Lahoura *et al.*, "Cloud computing-based framework for breast cancer diagnosis using extreme learning machine," *Diagnostics*, vol. 11, no. 2, Feb. 2021, doi: 10.3390/diagnostics11020241.
- [2] A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: A systematic literature review," *Artificial Intelligence in Medicine*, vol. 127, May 2022, doi: 10.1016/j.artmed.2022.102276.
- [3] C. Kaushal and A. Singla, "Analysis of breast cancer for histological dataset based on different feature extraction and classification algorithms," in *Advances in Intelligent Systems and Computing*, Springer Singapore, 2021, pp. 821–833.
- [4] G. Meenalochini and S. Ramkumar, "Survey of machine learning algorithms for breast cancer detection using mammogram images," *Materials Today: Proceedings*, vol. 37, pp. 2738–2743, 2021, doi: 10.1016/j.matpr.2020.08.543.
- [5] N. Kumari, "A survey on various machine learning approaches used for breast cancer detection," *International Journal of Advanced Research in Computer Science*, vol. 10, no. 3, pp. 76–79, Jun. 2019, doi: 10.26483/ijarcs.v10i3.6438.
- [6] D. Sharma, R. Kumar, and A. Jain, "A systematic review of risk factors and risk assessment models for breast cancer," in *Mobile Radio Communications and 5G Networks*, Springer Singapore, 2021, pp. 509–519.
- [7] O. Ginsburg *et al.*, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, no. S10, pp. 2379–2393, May 2020, doi: 10.1002/cncr.32887.
- [8] Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, "Comparative analysis of breast cancer detection using machine learning and biosensors," *Intelligent Medicine*, vol. 2, no. 2, pp. 69–81, May 2022, doi: 10.1016/j.imed.2021.08.004.
- [9] N. M. Hassan, S. Hamad, and K. Mahar, "Mammogram breast cancer CAD systems for mass detection and classification: a review," *Multimedia Tools and Applications*, vol. 81, no. 14, pp. 20043–20075, Jun. 2022, doi: 10.1007/s11042-022-12332-1.
- [10] A. Ahmed and O. M. B. El Sadig, "Heterogeneous multi-classifier method based on weighted voting for breast cancer detection," *International Journal of Advances in Science Engineering and Technology*, vol. 7, no. 4, pp. 36–41, 2019.
- [11] S. A. Medjahed, T. Ait Saadi, and A. Benyettou, "Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules," *International Journal of Computer Applications*, vol. 62, no. 1, pp. 1–5, Jan. 2013, doi: 10.5120/10041-4635.
- [12] K. Drukker, C. A. Sennett, and M. L. Giger, "Automated method for improving system performance of computer-aided diagnosis in breast ultrasound," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 122–128, Jan. 2009, doi: 10.1109/TMI.2008.928178.
- [13] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.
- [14] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, "Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review," *Expert Systems with Applications*, vol. 167, Apr. 2021, doi: 10.1016/j.eswa.2020.114161.
- [15] A. Hizukuri, R. Nakayama, M. Nara, M. Suzuki, and K. Namba, "Computer-aided diagnosis scheme for distinguishing between benign and malignant masses on breast DCE-MRI images using deep convolutional neural network with Bayesian optimization," *Journal of Digital Imaging*, vol. 34, no. 1, pp. 116–123, Feb. 2021, doi: 10.1007/s10278-020-00394-2.
- [16] H. Zerouaoui and A. Idri, "Reviewing machine learning and image processing based decision-making systems for breast cancer imaging," *Journal of Medical Systems*, vol. 45, no. 1, Jan. 2021, doi: 10.1007/s10916-020-01689-1.
- [17] H. Fang, H. Fan, S. Lin, Z. Qing, and F. R. Sheykhahmad, "Automatic breast cancer detection based on optimized neural network using whale optimization algorithm," *International Journal of Imaging Systems and Technology*, vol. 31, no. 1, pp. 425–438, Mar. 2021, doi: 10.1002/ima.22468.
- [18] Y. Wang, L. Zhang, X. Shu, Y. Feng, Y. Zhang, and Q. Lv, "Feature-sensitive deep convolutional neural network for multi-instance breast cancer detection," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, p. 1, 2021, doi: 10.1109/TCBB.2021.3060183.
- [19] S. Z. Ramadan, "Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review," *Journal of Healthcare Engineering*, pp. 1–21, Mar. 2020, doi: 10.1155/2020/9162464.
- [20] I. Hirra *et al.*, "Breast cancer classification from histopathological images using patch-based deep learning modeling," *IEEE Access*, vol. 9, pp. 24273–24287, 2021, doi: 10.1109/ACCESS.2021.3056516.
- [21] Y. Gao, K. J. Geras, A. A. Lewin, and L. Moy, "New frontiers: an update on computer-aided diagnosis for breast imaging in the age of artificial intelligence," *American Journal of Roentgenology*, vol. 212, no. 2, pp. 300–307, Feb. 2019, doi: 10.2214/AJR.18.20392.

- [22] C. Kaushal, S. Bhat, D. Koundal, and A. Singla, "Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images," *IRBM*, vol. 40, no. 4, pp. 211–227, Aug. 2019, doi: 10.1016/j.irbm.2019.06.001.
- [23] C. Kaushal, D. Koundal, and A. Singla, "Comparative analysis of segmentation techniques using histopathological images of breast cancer," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, Mar. 2019, pp. 261–266, doi: 10.1109/ICCMC.2019.8819659.
- [24] C. Kaushal and A. Singla, "Automated segmentation technique with self-driven post-processing for histopathological breast cancer images," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 4, pp. 294–300, Dec. 2020, doi: 10.1049/trit.2019.0077.
- [25] L. Liu, "Research on logistic regression algorithm of breast cancer diagnose data by machine learning," in *2018 International Conference on Robots and Intelligent System (ICRIS)*, May 2018, pp. 157–160, doi: 10.1109/ICRIS.2018.00049.
- [26] P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, "Comparative study of machine learning algorithms for breast cancer prediction," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020, pp. 796–801, doi: 10.1109/ICSSIT48917.2020.9214267.
- [27] A. M. S. C. Meenakshi, M. Govindarajan, "Detection of breast cancer using MLP and RBF classifiers," *IMS Manthan*, vol. 5, no. 1, 2010.
- [28] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, "Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis," *IEEE Access*, vol. 8, pp. 96946–96954, 2020, doi: 10.1109/ACCESS.2020.2993536.
- [29] D. Soria, J. M. Garibaldi, E. Biganzoli, and I. O. Ellis, "A comparison of three different methods for classification of breast cancer data," in *2008 Seventh International Conference on Machine Learning and Applications*, 2008, pp. 619–624, doi: 10.1109/ICMLA.2008.97.
- [30] M. Karabatak, "A new classifier for breast cancer detection based on Naïve Bayesian," *Measurement*, vol. 72, pp. 32–36, Aug. 2015, doi: 10.1016/j.measurement.2015.04.028.
- [31] M. I. H. Showrov, M. T. Islam, M. D. Hossain, and M. S. Ahmed, "Performance comparison of three classifiers for the classification of breast cancer dataset," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, Dec. 2019, pp. 1–5, doi: 10.1109/EICT48899.2019.9068816.
- [32] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3240–3247, Mar. 2009, doi: 10.1016/j.eswa.2008.01.009.
- [33] M. R. Al-Hadidi, A. Alarabeyyat, and M. Alhanahnah, "Breast cancer detection using k-nearest neighbor machine learning algorithm," in *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, Aug. 2016, pp. 35–39, doi: 10.1109/DeSE.2016.8.
- [34] B. Dai, R.-C. Chen, S.-Z. Zhu, and W.-W. Zhang, "Using random forest algorithm for breast cancer diagnosis," in *2018 International Symposium on Computer, Consumer and Control (IS3C)*, Dec. 2018, pp. 449–452, doi: 10.1109/IS3C.2018.00119.
- [35] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *Journal of Biomedical Science and Engineering*, vol. 6, no. 5, pp. 551–560, 2013, doi: 10.4236/jbise.2013.65070.
- [36] S. Sharma, A. Aggarwal, and T. Choudhury, "Breast cancer detection using machine learning algorithms," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, Dec. 2018, pp. 114–118, doi: 10.1109/CTEMS.2018.8769187.
- [37] T. Price and N. Lindqvist, "Evaluation of feature selection methods for machine learning classification of breast cancer," *Degree Project, Kth Royal Institute of Technology, School of Electrical Engineering and Computer Science*. Stockholm, Sweden, 2018.
- [38] S. Kabiraj *et al.*, "Breast cancer risk prediction using XGBoost and random forest algorithm," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–4, doi: 10.1109/ICCCNT49239.2020.9225451.
- [39] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer," *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, Feb. 2021, doi: 10.1016/j.amsu.2020.12.043.
- [40] S. Wang, Y. Wang, D. Wang, Y. Yin, Y. Wang, and Y. Jin, "An improved random forest-based rule extraction method for breast cancer diagnosis," *Applied Soft Computing*, vol. 86, Jan. 2020, doi: 10.1016/j.asoc.2019.105941.
- [41] R. Nagarajan and M. Upreti, "An ensemble predictive modeling framework for breast cancer classification," *Methods*, vol. 131, pp. 128–134, Dec. 2017, doi: 10.1016/j.ymeth.2017.07.011.
- [42] X. Tang, L. Cai, Y. Meng, C. Gu, J. Yang, and J. Yang, "A novel hybrid feature selection and ensemble learning framework for unbalanced cancer data diagnosis with transcriptome and functional proteomic," *IEEE Access*, vol. 9, pp. 51659–51668, 2021, doi: 10.1109/ACCESS.2021.3070428.
- [43] V. Chaurasia and S. Pal, "Stacking-based ensemble framework and feature selection technique for the detection of breast cancer," *SN Computer Science*, vol. 2, no. 2, Apr. 2021, doi: 10.1007/s42979-021-00465-3.
- [44] X. Zhang *et al.*, "Deep learning based analysis of breast cancer using advanced ensemble classifier and linear discriminant analysis," *IEEE Access*, vol. 8, pp. 120208–120217, 2020, doi: 10.1109/ACCESS.2020.3005228.
- [45] N. Arya and S. Saha, "Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020, doi: 10.1109/TCBB.2020.3018467.
- [46] S. J. Malebary and A. Hashmi, "Automated breast mass classification system using deep learning and ensemble learning in digital mammogram," *IEEE Access*, vol. 9, pp. 55312–55328, 2021, doi: 10.1109/ACCESS.2021.3071297.
- [47] M. Hosni, I. Abnane, A. Idri, J. M. Carrillo de Gea, and J. L. Fernández Alemán, "Reviewing ensemble classification methods in breast cancer," *Computer Methods and Programs in Biomedicine*, vol. 177, pp. 89–112, Aug. 2019, doi: 10.1016/j.cmpb.2019.05.019.
- [48] D. Dua and C. Graff, "UCI machine learning repository," *Irvine, CA: University of California, School of Information and Computer Science*. 2019.
- [49] X. Jin, A. Xu, R. Bie, and P. Guo, "Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles," in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2006, pp. 106–115.
- [50] Z. Mushtaq, A. Yaqub, S. Sani, and A. Khalid, "Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets," *Journal of the Chinese Institute of Engineers*, vol. 43, no. 1, pp. 80–92, Jan. 2020, doi: 10.1080/02533839.2019.1676658.
- [51] G. I. Salama, M. Abdelhalim, and M. A. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers," *Breast Cancer (WDBC)*, vol. 32, no. 569, 2012.
- [52] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, Sep. 2018, doi: 10.1016/j.jbi.2018.07.014.
- [53] M. R. Mahmood, "Two feature selection methods comparison chi-square and relief-F for facial expression recognition," *Journal of Physics: Conference Series*, vol. 1804, no. 1, Feb. 2021, doi: 10.1088/1742-6596/1804/1/012056.
- [54] N. Seedat and V. Aharonson, "Machine learning discrimination of Parkinson's disease stages from walker-mounted sensors data," in *Explainable AI in Healthcare and Medicine*, Springer International Publishing, 2021, pp. 37–44.

- [55] T. Li, C. Zhang, and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression," *Bioinformatics*, vol. 20, no. 15, pp. 2429–2437, Oct. 2004, doi: 10.1093/bioinformatics/bth267.
- [56] Y. R. Tabar, K. B. Mikkelsen, M. L. Rank, M. C. Hemmsen, and P. Kidmose, "Investigation of low dimensional feature spaces for automatic sleep staging," *Computer Methods and Programs in Biomedicine*, vol. 205, Jun. 2021, doi: 10.1016/j.cmpb.2021.106091.
- [57] A. Thakkar and R. Lohiya, "Attack classification using feature selection techniques: a comparative study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 1249–1266, Jan. 2021, doi: 10.1007/s12652-020-02167-9.

BIOGRAPHIES OF AUTHORS



Muawia A. Elsadig    received the bachelor degree in Computer Engineering, the MSc degree in computer networks, and the Ph.D. degree in computer science (information security). Currently, he is an assistant professor of cybersecurity and researcher, Deanship of Scientific Research, Imam Abdulrahman Bin Faisal University (IAU), Dammam KSA. In addition, he has engaged in teaching different computer courses in the college of Computer Science and Information Technology (CSIT) and the college of applied studies. He worked for different accredited international universities and has a rich record of publications at recognized international journals and conferences. He received many awards for his research achievements. Dr. Elsadig has many years of teaching experience and considerable industry contributions. He contributed as a reviewer for many international reputable journals. His research interests include the area of information security, network security, cybersecurity, wireless sensor networks, bioinformatics, and information extraction; ranging from theory to design to implementation. He can be contacted at email: muawiasadig@yahoo.com.



Abdelrahman Altigani    received the bachelor's degree in mathematics and the master's degree in computer science from the Faculty of Mathematical Sciences, University of Khartoum, Sudan. He is currently pursuing the Ph.D. degree. He has 11 publications in international refereed journals and conferences. He received the award of the best research article in the computing track at the ICCEEE2013 conference. During his studies, he received the International Ph.D. Fellowship scholarship thrice, in recognition of his outstanding academic performance. He participated in reviewing a number of research articles for many journals. His area of interest includes symmetric encryption. He can be contacted at email: aaaltigani@iau.edu.sa.



Huwaida T. Elshoush    received the bachelor's degree in computer science (division 1), the master's degree in computer science, and the Ph.D. degree in information security from the Faculty of Mathematical Sciences and Informatics, University of Khartoum, Sudan, in 1994, 2001, and 2012, respectively. Her M.Sc. dissertation dealt with Frame Relay Security. She is currently an Associate Professor within the Computer Science Department, Faculty of Mathematical Sciences and informatics, University of Khartoum, where she is also acting as the Head of Research office. She has more than 26 publications and some of her publications appeared in *Applied Soft Computing Journal* (Elsevier), *PLOS One Journal*, *IEEE Access*, *Multimedia Tools and Applications*, *PeerJ Computer Science*, *Journal of Information Hiding and Multimedia Signal Processing* and Springer book chapters. Her research interests include the information security, cryptography, steganography, and intrusion detection systems. She is a reviewer of many international reputable journals related to her fields, including *Applied Soft Computing Journal* (Elsevier). Dr. Elshoush's awards and honors include the second-place prize in the ACM Student Research Competition SRC-SAC in 2013 in Coimbra, Portugal. Her article entitled "An Improved Framework for Intrusion Alert Correlation" has been awarded the Best Student Paper Award of the 2012 International Conference of Information Security and Internet Engineering (ICISIE) in WCE 2012. Other prizes were the best student during the five years of her undergraduate study. She can be contacted at email: htelshoush@hotmail.com.