

*Breast Cancer Diagnosis, Feature Selection,
Neural Network, Grid Search, Machine Learning*

Nawazish NAVEED [0000-0002-9619-4799]*,
Hayan T. MADHLOOM [0000-0002-6228-5721]*,
Mohd Shahid HUSAIN [0000-0003-4864-9485]*

BREAST CANCER DIAGNOSIS USING WRAPPER-BASED FEATURE SELECTION AND ARTIFICIAL NEURAL NETWORK

Abstract

Breast cancer is commonest type of cancers among women. Early diagnosis plays a significant role in reducing the fatality rate. The main objective of this study is to propose an efficient approach to classify breast cancer tumor into either benign or malignant based on digitized image of a fine needle aspirate (FNA) of a breast mass represented by the Wisconsin Breast Cancer Dataset. Two wrapper-based feature selection methods, namely, sequential forward selection(SFS) and sequential backward selection (SBS) are used to identify the most discriminant features which can contribute to improve the classification performance. The feed forward neural network (FFNN) is used as a classification algorithm. The learning algorithm hyper-parameters are optimized using the grid search process. After selecting the optimal classification model, the data is divided into training set and testing set and the performance was evaluated. The feature space is reduced from nine feature to seven and six features using SFS and SBS respectively. The highest classification accuracy recorded was 99.03% with FFNN using the seven SFS selected features. While accuracy recorded with the six SBS selected features was 98.54%. The obtained results indicate that the proposed approach is effective in terms of feature space reduction leading to better accuracy and efficient classification model.

1. INTRODUCTION

Breast cancer is the most common cancer among women. it is also considered as the second most common cancer worldwide (Dhungel, Carneiro & Bradley, 2015). Early detection and accurate diagnosis of breast cancer can tremendously contribute to the reduction of fatality rate and remarkably important for the reduction of its morbidity and mortality (Addeh, Demirel & Zarbakhsh, 2017; Moodley et al., 2018). A cost-effective computer-aided detection/diagnosis technique can play a crucial part in reducing interpretation error and provide an automated diagnosis of breast cancer. this can hugely assist physicians

* University of Technology and Applied Sciences, CAS-Ibri, Dept. of IT, Sultanate of Oman,
nawazish.ibr@cas.edu.om, hayan.ibr@cas.edu.om, mshahid.ibr@cas.edu.om

by providing a second opinion which can ease the process of making the final decision. In this work, artificial neural network is used to build a classification model for breast cancer diagnosis based on fine needle aspiration modality. The main purpose is to correctly recognize the sample type as either benign or malignant. The Wisconsin breast cancer dataset (WBCD) is used for training and testing the proposed model. The accuracy of the classifier highly depends on the features used for classification. We have used the wrapper feature selection to extract the most useful features for the diagnosis purpose. Our approach shows encouraging results and can be developed in a fully automated cad system.

2. RELATED WORKS

Several researchers studied the performance of various prediction algorithms in classifying breast cancer data. (Senturk & Kara, 2014) compared the performance of seven different classification algorithms on the WBCD including discriminant analysis, artificial neural network (ANN), decision tree, logistic regression, support vector machine (SVM), Naïve Bayes (NB), and k-nearest neighbor (KNN). these algorithms were tested on all the nine features provided in the dataset. the best performance was obtained using the SVM with accuracy of 96.5%. another study (Barna & Khan, 2019) conducted to test the performance of six different classifiers, namely, logistic regression, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), KNN, NB and ANN. the researchers concluded that the ANN outperformed the other classifiers when the data was separated in to training and testing with the ratio of 75:25. the reported accuracy of ANN was 97.21%.

The work presented by (Vijayalakshmi & Priyadarshini, 2017) focused on using neural network to classify breast cancer tumors. The effectiveness of two different neural networks were compared, namely, the Radial Basis Function (RBF), and the back-propagation neural (BPN). It was found that the RBF neural network achieved higher accuracy than the BPN. The RBF accuracy was 98.26% while the BPN was 90.42%.

Khan et al. (2019) proposed a novel embedded approach based on accelerated particle swarm optimization and cuckoo search named HACPSO. Along with other datasets, they tested their approach on Wisconsin Breast Cancer dataset and achieved 98.08% classification accuracy that outclass the existing methods. Mushtaq et al. (2019) compared five different classification algorithms combined with principle-component analysis (PCA) to classify breast cancer tumor. The highest accuracy of 99.20% was obtained with sigmoid based NB classifier.

Kumar studied (Kumar, 2021) the performance of seven different classifiers (logistic regression, KNN, decision tree, SVM, Naïve Bayes, random forest and ANN) and reported that KNN is the best classifier with 97% accuracy and decision tree classifier is the worst performer with 94% accuracy for breast cancer diagnosis. Another study (Islam et al., 2020) reported that ANN is the best classifier in terms of accuracy (98.57%), and precision (97.82%) compare to SVM, KNN, logistic regression and random forest when tested on WBCD.

In the above mentioned studies, the authors focused on comparing the performance of various classification algorithms without giving importance to find the best feature subset. It was assumed that all the features are equally important for the classification, hence, no feature selection was performed. In medical application especially in cancer classification

related problems, it is quite significant to find the best feature subset and also to understand the features dependencies. This will help the domain experts to understand the most effective tumor characteristics and the relationship between them.

On the other hand, some other studies used feature selection methods to find the best attributes which can produce the highest discrimination results. A rule based classification system with PCA was proposed by (Douangnoulack & Boonjing, 2018). Only 7 features were used for the classification and the J48 classifier achieved the accuracy of 97.36% on the WBCD. Kumari & Singh (2018) proposed a system for the prediction of breast cancer. A combination of feature selection using Correlation-Based Measures with classification using several algorithms including linear Regression (LR), SVM, and KNN algorithm. The validation was performed using 10-fold cross validation and the accuracy of the model was 99.2%. Nevertheless, the optimal feature subset was not reported.

Ed-Daoudy & Maalmi (2020) applied Association Rules (AR) to eliminate irrelevant features in the WBCD. Four out of nine features were selected. Several classification algorithms were used. The support vector machine with threefold cross-validation produced the highest classification accuracy (98.00%).

3. MATERIALS AND METHODS

3.1. Wisconsin breast cancer dataset

The Wisconsin breast cancer database (WBCD) was created by Dr. Wolberg from the University of Wisconsin Hospitals and was donated and made publicly available online by Mangasarian (WBCD, 1995). The data was collected over a period of two years starting from 1989 until 1991, and it is used as a standard dataset for classification and other machine learning purposes by several researchers. It represents the observations of breast mass cell nuclei obtained by a fine-needle aspiration modality. The cytological samples were converted into digital images in order to extract the characteristics of the cell nuclei using image processing techniques.

Fine needle aspiration (FNA) is a type of biopsy procedure. Basically, it is one of the various modalities used in the process of breast masses diagnosis. In this procedure a small needle (21 to 25 gauge) is used to acquire a sample of the tissue and fluid from the breast (Casabon, Tomlinson-Hansen & Regan, 2020) Total number of 699 samples are available in the WBCD, each sample represented by nine different nuclei features. Furthermore, the diagnosis of each sample as a benign or malignant was also provided in the dataset. 458/699 observations were flagged as benign, while 241/699 were flagged as malignant. The details and the names of the feature set are described in Tab. 1. Each feature has a grade between 1 and 10, where the value 1 indicates that the feature is in most normal condition and the value 10 indicates most abnormal condition.

It has been noticed that there are 16 observations with missing values. All the missing values are related to the Bare Nuclei feature. 15 samples are from the benign class, and one sample is from the malignant class. Since the number of missing values are small and in the interest of maintaining data consistency, these 16 samples were removed.

Tab. 1. Dataset details with feature names

Feature Index	Feature Name	Value Range	$\mu \pm \sigma$
1	Clump thickness	1–10	3.42±2.82
2	Uniformity of cell size	1–10	3.13±3.05
3	Uniformity of cell shape	1–10	3.20±2.97
4	Marginal adhesion	1–10	2.80±2.86
5	Single epithelial cell size	1–10	3.21±2.21
6	Bare nuclei	1–10	3.46±3.64
7	Bland chromatin	1–10	3.43±2.44
8	Normal nucleoli	1–10	2.87±3.05
9	Mitoses	1–10	1.59±1.71

3.2. Feature Selection

Feature selection is a commonly used data preprocessing procedure in data classification. It is mainly used for reducing and eliminating irrelevant and redundant attributes from any dataset (Tang, Alelyani & Liu, 2014; Foithong, Srinil & Pinngern, 2017). Additionally, it plays a significant role in enhancing data comprehensibility, data visualization as well as reducing the time to train a classification model, and improves the prediction results (Jain & Singh, 2018).

There exist numerous applications of relevant feature identification techniques in healthcare sector. Filter methods, wrapper methods, ensemble methods and embedded methods are some of the popularly used techniques used for variable selection (Kohavi & John, 1997; Guyon et al., 2008).

In this paper two wrapper feature selection methods are used, namely, the sequential forward selection (SFS) and the sequential backward feature (SBS) selection. The wrapper feature selection methods outperform other existing methods such as filter methods. It finds the most “useful” features and does optimal selection of features for the learning algorithm (Kumar & Minz, 2014); furthermore, the wrapper methods give more accurate results as it considers the features dependencies (Ang et al., 2015). It has been stated that the Naïve-Bayes learning algorithm is robust when it is used to remove noisy features (Kohavi & John, 1997). This is because the performance of the Naïve-Bayes degrades very slowly as more irrelevant features are added (Kohavi & John, 1997). For that reason, the Naïve-Bayes learning algorithm is used with both SFS and SBS.

In this research, the ultimate goal of performing the feature selection process is not limited to obtaining the highest classification accuracy. However, it is also related to the detection of the most clinically significant features as this optimal set of features can help the specialist objectively focus on these features during a routine manual diagnosis process.

Both SFS and SBS are iterative methods. The SFS starts with an empty set and in each iteration a new unseen feature is added. For each added feature, performance is evaluated using the induction algorithm. Only the feature producing the highest increase of performance is added to new feature subset. Then a new iteration is started with the new generated subset. On the other hand, the SBS starts with full feature set and at each iteration one feature is removed. In both methods the searching process stops when there is no further σ improvement is detected by the induction algorithm. Fig. 1 illustrates the feature selection process.

According to Fig. 1, the feature selection process for both SFS and SBS starts by generating a feature subset. The performance of the subset is evaluated with Naïve-Bayes using 10-fold cross validation process. For each subset, if the induction algorithm performance increase, the final optimal feature subset is updated. The process continues to evaluate the features until no further enhancement in the performance detected.

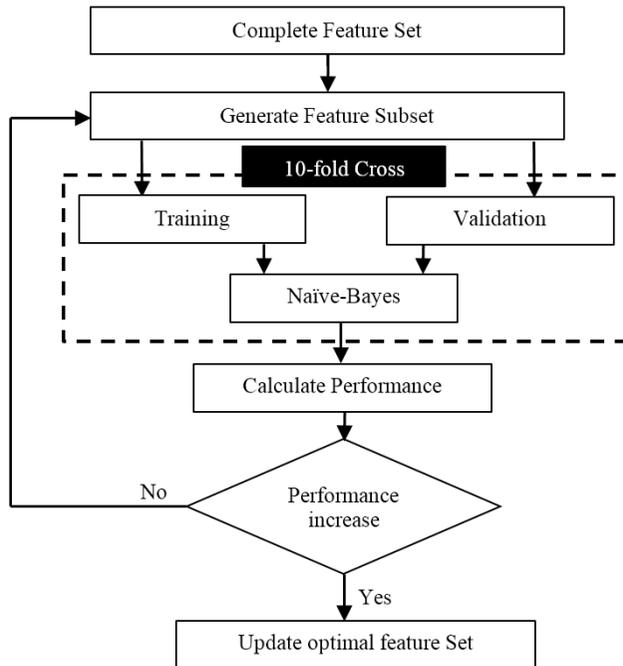


Fig. 1. Feature Selection Process

3.3. Classification

In this research, the ANN is used to classify the breast cancer samples in the WBCD into either benign or malignant. ANN was used intensively in the diagnosis and classification of many medical conditions such as leukemia (Wahhab, 2015), prostate cancer (Wu, Zhuang & Tan, 2020), lung cancer (Hsu et al., 2020), liver cancer (Patsadu, Tangchitwilaikun & Lowsuwankul, 2021) and many others. There are various ANN architectures (Agrawal & Agrawal, 2015). However, one of the most widely used is the multilayer feed-forward neural network (FFNN) with a back-propagation learning algorithm (Zarei et al., 2020). In FFNN, there are a number of parameters need to be tuned in order to obtain the best classification performance. These parameters include the number of hidden layers, the activation function, the number of neurons, the learning rate, and the epochs. The number of hidden layer is set to one as usually single hidden layer is sufficient for various kinds of classification problem (Guliyev & Ismailov, 2018). Regarding the activation function, in fact, there are many options available. Nevertheless, previous researches (Bonakdari et al., 2020; Shenouda, 2006) have established that the sigmoid activation function produced a better result in medical and non-medical applications compared to other activation functions. Hence, for this experiment, the sigmoid activation function is chosen.

The rest of the FFNN hyper-parameters is tuned with the grid search optimization using 10-fold cross validation. The grid search algorithm traverses a given combination of parameters. Later, the parameters resulted in the best performance can be used to train the final model and tested using the test set. In grid search the performance of the model is verified using a statistical method called cross validation (CV) (Liu at al., 2017). The cross validation divides the dataset into two parts, namely, training and validation. On each hyper-parameter combination, the FFNN is trained and the accuracy is verified. Eventually, the model which produced the highest performance is used for the final classification test. Tab. 2 presented the range of each hyper-parameter used in the optimization process.

Tab. 2. Hyper-parameters values

Hyper-parameter	Values
Learning rate	[0.001, 0.01, 0.1]
Number of neurons in the hidden layer	[2,4,6,8,10, 12, 14]
Training cycle	Max(2000)

3.4. Performance Measure

The performance of the proposed breast cancer classification model is evaluated using the confusion matrix. It is used to calculate true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Accuracy is the most empirical metric used to assess effectiveness of a classifiers. Other important metrics are the precision and recall. The precision is calculated as the correct positive prediction over all the samples classified as positive, while recall is used to test the classifier ability to identify the positive cases. These metrics are calculated as follow:

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

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

4. RESULTS AND DISCUSSION

The proposed work is performed in three folds, first, the best features were selected using two wrapper methods, namely, sequential forward selection and sequential backward selection. Second, the hyper-parameters optimization using grid search, and finally the FFNN training and testing with optimal set of features. The feature selection and classification analysis were performed using RapidMiner. RapidMiner is an open-source data mining and machine learning tool and it provides the largest coverage of healthcare data mining requirements compared to other tools such as R, Scikit-learn and Spark (Santos-Pereira, Gruenwald & Bernardino, 2021).

From the data presented in Fig. 2, it can be seen that both wrapper methods performance is equal at feature number 3 and feature number 5. However, the optimal performance of SFS was obtained at 7th feature. On the other hand, the SBS reached its highest performance at the 6th feature. The best selected features for both methods are listed in Tab. 3.

Based on this result, the top 7 features selected using SFS were utilized to construct the FFNN model.

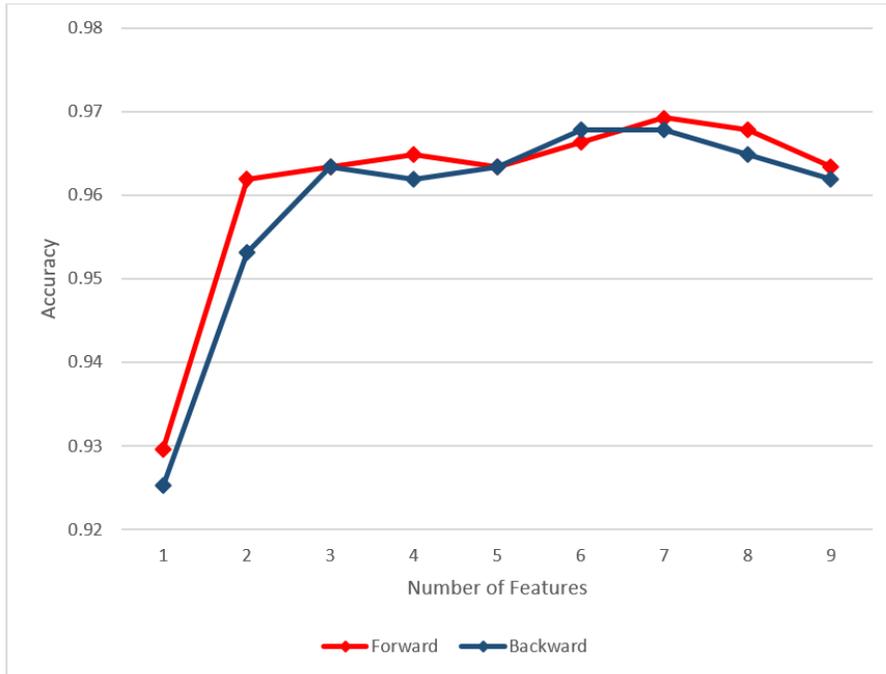


Fig. 2. SFS and SBS performance

Tab. 3. The best feature subset selected using SFS and SBS

Feature Index	Optimal Features (SFS)	Optimal Features (SBS)
F1	Bare Nuclei	Bare Nuclei
F2	Normal Nucleoli	Normal Nucleoli
F3	Clump Thickness	Clump Thickness
F4	Uniformity of Cell Shape	Uniformity of Cell Shape
F5	Bland Chromatin	Bland Chromatin
F6	Marginal Adhesion	Marginal Adhesion
F7	Uniformity of Cell Size	

During the grid search optimization, 18 FFNN architecture was trained and validated using 10-fold cross validation. Each architecture was constructed with different set of parameters as mentioned in Tab. 3.

As shown in Fig. 3, the best result was obtained when number of neurons in the hidden layer was equal to 6 and learning rate equal to 0.01 with 100 epochs.

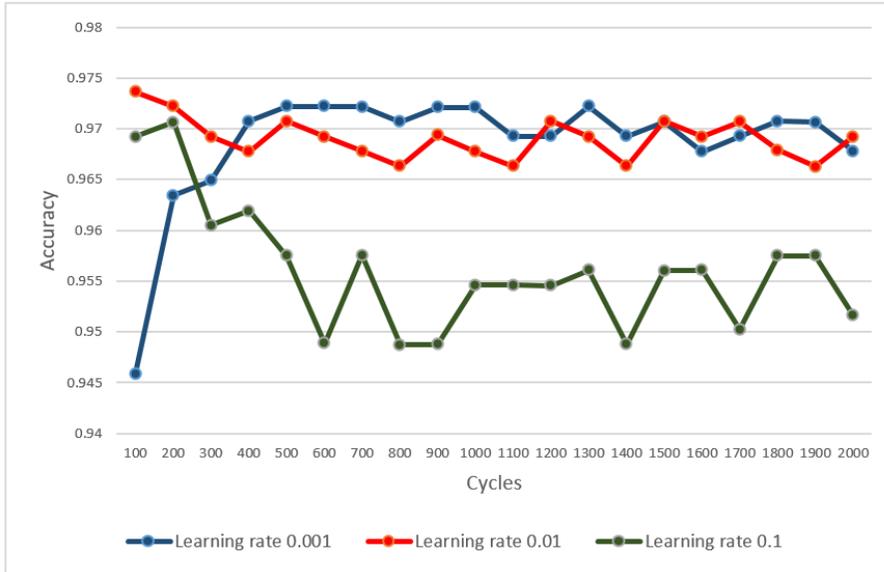


Fig. 3. The performance of the three different learning rates with six hidden neurons

After finding the optimal hyper-parameters, the original dataset was divided into two sets, the training set and the test set with the ratio of 70% for training and 30% for testing. Based on that, and in order to compare the classification results obtained using both feature selections methods, the classification experiment was performed twice. The first experiment considered the 7 features selected using SFS, whilst, the second experiment was performed on the 6 features selected using SBS. The final classification results are presented in Tab. 4, and Tab. 5.

It is apparent from the data presented in Tab. 4 that the FFNN classification performance using both feature subsets obtained using SFS and SBS is almost the same as in both experiments all the benign instances were correctly classified. Nevertheless, there is a slight difference in the classification result of the malignant instances. The SFS subset resulted in classifying 69 malignant instances out of 71, whereas the SBS subset wrongly classifies 3 malignant instances as benign.

Tab. 4. Classification results with SFS and SBS feature subsets

Experiments	Correctly classified instances	Incorrectly classified instances	TP	TN	FP	FN
SFS + FFNN (7 Features)	204	2	69	135	0	2
SBS + FFNN (6 Features)	203	3	68	135	0	3

As illustrated in Tab. 5, The overall classification accuracy of FFNN using the SFS features subset is outperformed the accuracy obtained using the SBS features subset. In the first experiment the accuracy was 99.03 % with 100% precision and 97.18% recall. While in the second experiment the accuracy was 98.54 % with 100% precision and 95.77% recall.

Tab. 5. Performance evaluation of FFNN with SFS and SBS feature subsets

Experiments	Accuracy (%)	Precision (%)	Recall (%)
SFS + FFNN (7 Features)	99.03	100	97.18
SBS + FFNN (6 Features)	98.54	100	95.77

As shown in Tab. 4, the first experiment classified 2 malignant cases as benign and in the second experiment there are 3 malignant cases classified as benign. Although the number of misclassified instances are insignificant, in medical application classifying an unhealthy case as healthy is more dangerous than classifying healthy case as unhealthy. This is because in the latter situation the patient can undergo further investigation and then the misclassification can be ruled out.

In our experiment, the reason of getting misclassified instances could be due to the imbalance data distribution of the WBCD as the ratio of majority (benign) to minority (malignant) is approximately 1:2. The number of benign instances used in the training phase (309) is almost double in size compared to the number of malignant instances (168). Although the majority to minority ratio is not tremendously high, the possibility of the learning algorithm overwhelmed by the majority class could not be ruled out. Nevertheless, the proposed approach to classify breast cancer provides an outstanding accuracy with high precision and recall in comparison with the previous experimental results in the literature. Moreover, the proposed approach identified the optimal discriminative set of features using the SFS feature selection algorithm. It has been found that the two features, namely, Single epithelial cell size and Mitoses are not contributing to the diagnosis of breast cancer using the fine-needle aspiration modality.

In the literature, several methods have been proposed for the diagnosis of breast cancer based on the WBCD. Tab. 6 demonstrates the classification accuracy achieved by previous studies compared to our proposed approach.

Tab. 6. Comparison of the proposed method with previous researches in the literature

Authors	Feature Selection Reduction	No of feature selected	Classification method	Reported Accuracy (%)
(Vijayalakshmi & Priyadarshini, 2017)	Not Reported	Not Reported	RBF neural network	98.26
(Yi & Yi, 2017)	Pearson chi-square test	4	Decision Tree	94.30
(Douangnoulack & Boonjing, 2018)	PCA	7	Decision Tree (J48)	97.36
(Kumari & Singh, 2018)	Pearson's linear correlation	Not Reported	KNN	99.28
(Mushtaq et al., 2019)	Not Reported	Not Reported	Naïve Bayes with sigmoid PCA	99.20
(Ed-daoudy & Maalmi, 2020)	association rules	4	SVM	98.00
Proposed Approach	SFS	7	FFNN	99.03

It can be seen that the proposed method outperformed many of the works previously done. On the other hand, the work performed by (Kumari & Singh, 2018; Mushtaq et al., 2019) achieved slightly higher accuracy at 99.28 and 99.2 respectively compared to the proposed method. However, both research works did not focus on finding the most clinically significant. Furthermore, the number of selected features are not reported. Most of the feature selection methods used in the previous experiments focused on filter feature selection methods. Usually filter methods do not consider dependencies between the features. However, in this research the SFS and SBS are used to get the optimal subset of features. The main advantage of these methods is that it will take the interaction between the features into consideration. One interesting finding is that a well optimized FFNN architecture can produce better classification accuracy compared to some of the most popular machine learning algorithms such as SVM and J48.

It is worth mentioning that in medical application especially in cancer classification related problems, it is highly important to select the best set of features that can produce the optimal model performance. This will help the domain experts to understand the most effective tumor characteristics and the relationship between them.

5. CONCLUSION

Data mining and machine learning techniques are extensively used to explore patterns in medical data, which can be used for many purposes such as diagnosis and prognosis. Many researches have been conducted in the medical field to accurately diagnose several diseases such as cancer. One of the most important step in the context of computer-aided diagnosis is features reduction. Certainly, there are some features non-informative and redundant features. These features make the classification algorithms ineffective. Hence, features selection will considerably enhance the performance of the classification algorithm.

In this research, an approach to classify breast cancer based on FNA modality has been proposed. Feature selection is a prominent process used for improving the overall classification accuracy as well as understanding the tumor characteristics. In this work, two wrapper feature selection methods, namely, the SFS and SBS were used to extract the optimal subset of tumor attributes. Using SFS, seven important features have been identified from the original nine feature set. Afterward, the FFNN classifier was optimized, trained and tested on the WBCD. The proposed approach performance is evaluated and compared with other previous works. The seven features selected using SFS produced the highest accuracy of 99.03. This research demonstrated wrapper feature selection methods such as SFS can be used for removing the less important features and the proposed FFNN model can be used to obtain efficient automatic diagnostic systems.

Acknowledgment

This work is part of a research project(DSR_IRPS_2019-20_PROP_4) supported by the University of Technology and Applied Science, CAS-Ibri.

REFERENCES

- Addeh, A., Demirel, H., & Zorbakhsh, P. (2017). Early detection of breast cancer using optimized ANFIS and features selection. *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 39–42). IEEE. <http://doi.org/10.1109/CICN.2017.8319352>
- Agrawal, S., & Agrawal, J. (2015). Neural network techniques for cancer prediction: A survey. *Procedia Computer Science*, *60*, 769–774. <http://doi.org/10.1016/j.procs.2015.08.234>
- Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2015). Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. *IEEE/ACM transactions on computational biology and bioinformatics*, *13*(5), 971–989. <http://doi.org/10.1109/TCBB.2015.2478454>
- Barna, S. D., & Khan, S. (2019). Performance Evaluation of Classification Learning Models for Wisconsin Breast Cancer Data Repository. *7th International Conference on Data Science and SDGs: Challenges, Opportunities and Realities (EC-50)*. Bangladesh.
- Bonakdari, H., Moradi, F., Ebtehaj, I., Gharabaghi, B., Sattar, A. A., Azimi, A. H., & Radecki-Pawlik, A. (2020). A Non-Tuned Machine Learning Technique for Abutment Scour Depth in Clear Water Condition. *Water*, *12*(1), 301. <http://doi.org/10.3390/w12010301>
- Casaubon, J. T., Tomlinson-Hansen, S., & Regan, J.-P. (2020). *Fine Needle Aspiration of Breast Masses*. StatPearls. StatPearls Publishing.
- Dhungle, N., Carneiro, G., & Bradley, A. P. (2015). Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests. *International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1–8). IEEE. <http://doi.org/10.1109/DICTA.2015.7371234>
- Douangnoulack, P., & Boonjing, V. (2018). Building Minimal Classification Rules for Breast Cancer Diagnosis. *2018 10th International Conference on Knowledge and Smart Technology (KST)* (pp. 278–281). IEEE. <http://doi.org/10.1109/KST.2018.8426198>
- Ed-Daoudy, A., & Maalmi, K. (2020). Breast cancer classification with reduced feature set using association rules and support vector machine. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *9*(1), 34. <http://doi.org/10.1007/s13721-020-00237-8>
- Foithong, S., Srinil, P., & Pinngern, O. (2017). Min-Uncertainty & Max-Certainty Criteria of Neighborhood Rough-Mutual Feature Selection. *Walailak Journal of Science and Technology*, *14*(4).
- Guliyev, N. J., & Ismailov, V. E. (2018). On the approximation by single hidden layer feedforward neural networks with fixed weights. *Neural Networks*, *98*, 296–304. <http://doi.org/10.1016/j.neunet.2017.12.007>
- Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2008). *Feature extraction: foundations and applications* (Vol. 207). Springer. <http://doi.org/10.1007/978-3-540-35488-8>
- Hsu, Y.-C., Tsai, Y.-H., Weng, H.-H., Hsu, L.-S., Tsai, Y.-H., Lin, Y.-C., Hung, M.-S., Fang, Y.-H., & Chen, C.-W. (2020). Artificial neural networks improve LDCT lung cancer screening: a comparative validation study. *BMC Cancer*, *20*(1), 1023. <https://doi.org/10.1186/s12885-020-07465-1>
- Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M., & Kabir, M. N. (2020). Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. *SN Computer Science*, *1*(5), 290. <https://doi.org/10.1007/s42979-020-00305-w>
- Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*, *19*(3), 179–189. <https://doi.org/10.1016/j.eij.2018.03.002>
- Khan, A., Shah, R., Imran, M., Khan, A., Bangash, J. I., & Shah, K. (2019). An alternative approach to neural network training based on hybrid bio meta-heuristic algorithm. *Journal of Ambient Intelligence and Humanized Computing*, *10*(10), 3821–3830. <https://doi.org/10.1007/s12652-019-01373-4>
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, *97*(1–2), 273–324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- Kumar, V. (2021). Evaluation of computationally intelligent techniques for breast cancer diagnosis. *Neural Computing and Applications*, *33*(8), 3195–3208. <https://doi.org/10.1007/s00521-020-05204-y>
- Kumar, V., & Minz, S. (2014). Feature selection: a literature review. *SmartCR*, *4*(3), 211–229.
- Kumari, M., & Singh, V. (2018). Breast Cancer Prediction system. *Procedia Computer Science*, *132*, 371–376. <https://doi.org/10.1016/j.procs.2018.05.197>
- Liu, X., Li, B., Shen, D., Cao, J., & Mao, B. (2017). Analysis of Grain Storage Loss Based on Decision Tree Algorithm. *Procedia Computer Science*, *122*, 130–137. <https://doi.org/10.1016/j.procs.2017.11.351>
- Moodley, J., Walter, F., Scott, S., & Mwaka, A. (2018). Towards timely diagnosis of symptomatic breast and cervical cancer in South Africa. *South African Medical Journal*, *108*(10), 803–804. <https://doi.org/10.7196/SAMJ.2018.v108i10.13478>

- Mushtaq, Z., Yaqub, A., Hassan, A., & Su, S. F. (2019). Performance Analysis of Supervised Classifiers Using PCA Based Techniques on Breast Cancer. *2019 International Conference on Engineering and Emerging Technologies (ICEET)* (pp. 1–6). IEEE. <https://doi.org/10.1109/CEET1.2019.8711868>
- Patsadu, O., Tangchitwilaikun, P., & Lowsuwankul, S. (2021). Liver Cancer Patient Classification on a Multiple-Stage using Hybrid Classification Methods. *Walailak Journal of Science and Technology*, *18*(10). <https://doi.org/10.48048/wjst.2021.9169>
- Santos-Pereira, J., Gruenwald, L., & Bernardino, J. (2021). Top data mining tools for the healthcare industry. *Journal of King Saud University – Computer and Information Sciences*, in press. <https://doi.org/https://doi.org/10.1016/j.jksuci.2021.06.002>
- Senturk, Z. K., & Kara, R. (2014). Breast Cancer Diagnosis Via Data Mining: Performance Analysis of Seven Different algorithms. *Computer Science & Engineering: An International Journal (CSEIJ)*, *4*(1), 35–46. <https://doi.org/10.5121/cseij.2014.4104>
- Shenouda, E. A. M. A. (2006). A Quantitative Comparison of Different MLP Activation Functions in Classification. In: J. Wang, Z. Yi, J. M. Zurada, B. L. Lu & H. Yin (Eds.), *Advances in Neural Networks. Lecture Notes in Computer Science* (vol. 3971). Springer. https://doi.org/10.1007/11759966_125
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. In *Data classification: Algorithms and applications* (chapter 2). Chapman and Hall/CRC. <https://doi.org/10.1201/b17320>
- Vijayalakshmi, S., & Priyadarshini, J. (2017). Breast Cancer Classification using RBF and BPN Neural Networks. *International Journal of Applied Engineering Research*, *12*(15), 4775–4781.
- Wahhab, H. T. A. (2015). *Classification of acute leukemia using image processing and machine learning techniques*. University of Malaya.
- WBCD. (1995). Retrieved January 20, 2021 from [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)).
- Wu, J., Zhuang, Q., & Tan, Y. (2020). Auxiliary Medical Decision System for Prostate Cancer Based on Ensemble Method. *Computational and Mathematical Methods in Medicine*, *2020*, 6509596. <https://doi.org/10.1155/2020/6509596>
- Yi, L., & Yi, W. (2017). Decision Tree Model in the Diagnosis of Breast Cancer. In *2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC)* (pp. 176–179). IEEE. <https://doi.org/10.1109/ICCTEC.2017.00046>
- Zarei, M., Ansari, H., Keshavarz, P., & Zerafat, M. (2020). Prediction of pool boiling heat transfer coefficient for various nano-refrigerants utilizing artificial neural networks. *Journal of Thermal Analysis and Calorimetry*, *139*(6), 3757–3768.