Bulletin of Electrical Engineering and Informatics
Vol. 12, No. 2, April 2023, pp. 1047~1054
ISSN: 2302-9285, DOI: 10.11591/eei.v12i2.4838

A case study of microarray breast cancer classification using machine learning algorithms with grid search cross validation

Nursabillilah Mohd Ali[1,2], Rosli Besar[2], Nor Azlina Ab Aziz[2]
[1]Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia

Article Info

Article history:

Received Sep 26, 2022
Revised Nov 2, 2022
Accepted Nov 23, 2022

Keywords:

Breast Cancer
Classification
Grid search CV
Microarray

ABSTRACT

Breast cancer is one of the leading causes of death and most frequently diagnosed cancer amongst women. Annually, almost half a million women do not survive the disease and die from breast cancer. Machine learning is a subfield of artificial intelligence (AI) and computer science that uses data and algorithms to mimic how humans learn, and gradually improving its accuracy. In this work, simple machine learning methods are used to classify breast cancer microarray data to normal and relapse. The data is from the gene expression omnibus (GEO) website namely GSE45255 and GSE15852. These two datasets are integrated and combined to form a single dataset. The study involved three machine learning algorithms, random forest (RF), extra tree (ET), and support vector machine (SVM). Grid search cross validation (CV) is applied for hyperparameter tuning of the algorithms. The result shows that the tuned SVM is best among the tested algorithms with accuracy of 97.78%. In the future it is recommended to include feature selection method to get the optimal features and better classification accuracies.

Corresponding Author:

Nursabillilah Mohd Ali
Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka
Durian Tunggal, Melaka, Malaysia
Email: nursabillilah@utem.edu.my

1. INTRODUCTION

Breast cancer is the most prevalent disease among women worldwide. Many women are affected by this life-threatening cancer. It is the second biggest factor in female cancer-related fatalities [1], [2]. Breast cancer is a malignant tumor caused by the breast's cells growing and dividing out of control thus creating a lump of tissues. However, not all lumps are cancerous, benign tumors are non-cancerous growths that are treatable with medication and are not life-threatening [2]. Whereas malignant tumors are cancerous growths that can be fatal if left untreated. Early diagnosis is important if such a lump appears in the patients' breast, they must discuss with a medical doctor for early diagnoses and medical treatment [3], [4].

One of the most essential technologies in bioinformatics research is the gene chip, commonly known as the DNA microarray [5]–[7]. A great amount of biological information is available in gene expression microarray data. This is contributed by the rapid development of sequencing technologies [5], [6]. Breast cancer gene expression profiles are among information available in microarray data, which is important in prognosis of breast cancer patients [7]-[9]. The expression variables in the microarray dataset are often organised as a *MxN* matrix, with column containing several features [10] (also known as genes) and each row representing a sample, as illustrated in Figure 1 [6].

In recent years, researchers have shown a great deal of interest in the detection and classification of cancer through microarray data using machine learning algorithms. The classification of microarray data

classifies cancer samples according to their class based on their gene expression profiles. Meanwhile, machine learning is a subset of artificial intelligence (AI) that enables systems to learn from the training data and get better over time. According to Almugren and Alshamlan [8], a machine learning algorithm known as support vector machine (SVM) is hybridized with firefly algorithm for classification of several type of cancers through selection of microarray features. There are several challenges in classification of microarray data. In a gene expression study, thousands of genes features are obtained from a smaller number of samples [9]. This is led to what is known as high dimensionality problem in microarrays [10]. In addition, gene expression also contains numerous ineffective and unnecessary attributes, and just a handful of the assessed genes may have a meaningful impact on cancer classification. Therefore, the classification of microarray data is still challenging and difficult due to the small samples number and high dimensionality problem [11], [12].

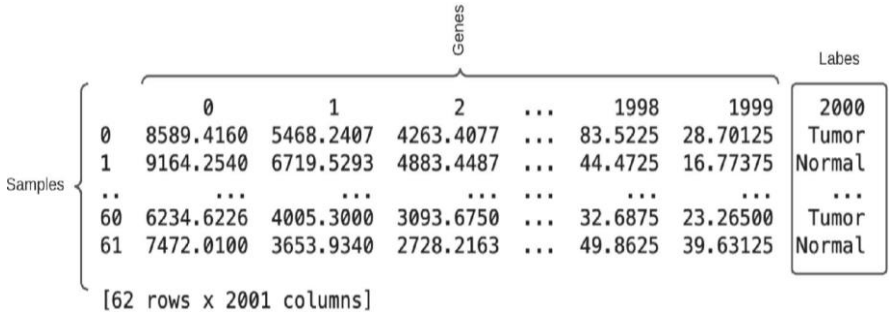

Figure 1. Data matrix of gene expression profiles [6]

In this study, we apply simple machine learning algorithms to classify high dimensional microarray breast cancer data. The three machine learning methods applied namely are, random forest (RF), extra tree (ET) and SVM. The classification models are applied on the data without using any feature selection methods. The hyperparameters of the three machine learning models are selected using grid search cross validation (CV) method. This study aims to determine the best classifier among the three after performing grid search CV.

2. THE MACHINE LEARNING ALGORITHMS

Classification is a data mining technique that identifies or assigns categories to a set of data to enable more accurate analysis. Supervised classification is a type of learning in which labels are determined [11], [13]. There are two steps involved in constructing a classifier: i) the learning phase, during which the model or classifier is constructed based on a set of training data and paired with a class label and ii) predicting the accuracy of the model on unseen data. Three common machine learning methods [13]–[15], RF, ET and SVM with grid search CV are applied in this work. These three machine learning models were chosen as a classifier technique in this study for several reasons namely: i) they are fast and ii) they are able to deal with high dimensional dataset. Grid search CV was used to aid in tuning hyperparameters and fitting the model to the training data using the optimal parameters. This study implements kfold cross-validation (CV), with the number of folds is set to 10.

2.1. Random forest

RF method is a collection of tree-structured learning classifiers. It categorizes a fresh sample using the most frequently occurring prediction produced by these algorithms. The trees are grown via feature selection, and at each node, random features are chosen for splitting. This helps to reduce over-fitting and, as a result, RF classification is quick [16].

2.2. Extremely randomized tree

For classification, a group of many decision trees is utilized. This depicts a forest of decision trees like the RF method, but are constructed differently [17]. Every decision tree chooses the best feature from a set of K randomly chosen qualities to divide each node based on some chosen criterion. Using the training dataset, the ET algorithm generates unpruned trees and numerous decision trees. This algorithm averages the predictions for regression and majority voting to produce final predictions for all decision trees.

2.3. Support vector machine

SVM [17] focuses on locating a hyperplane that best divides the tuples of one class from those of another. Using the support vector and the margin, the hyperplane is identified. The support vector is calculated using the hyperplane's vectors (data points). The margin is the closest point to the hyperplane (on two sides). However, when the data is linearly separable, the hyperplane is the line that divides the data into two pieces, with each portion ultimately belonging to a single class. Maximizing the margin, which is the distance between the nearest data point (called the support vector) in each class, enables the identification of the optimal hyperplane. SVM Kernels (linear and radial basis function (RBF)), the C (cost), and the gamma values were all tuned to achieve the best SVM model [18].

2.4. Grid search cross-validation for hyperparameter tuning

With the right combination of hyperparameters, a machine learning model that is resilient and accurate can be built [14], [18]. Hyperparameter tuning refers to the process of selecting the optimal set of parameters. To increase the performance metric, the dataset must be trained using all machine learning methods and different combinations of hyperparameters. The dataset can be trained using a variety of machine learning methods using the CV technique. Here are some of the common terms that should be considered when using grid search CV (GridSearchCV).

- Estimator: this term is used in scikit-learn to set up the estimator interface. This parameter gets the classifier that needs to be trained.
- Parameter grid: parameter names and settings are in a Python key-value dictionary. All parameters are checked for most accurate results.
- CV: this establishes the CV splitting approach. Resampling the available data is a technique called CV that is used to assess machine learning models. The major objective of this is to assess how well machine learning models perform on new data. It operates by first randomly shuffling the dataset. Then, k groups are created from the complete dataset. While the other groups are utilized as training data, each group is used as a test group. Each sample is utilized k-1 times and only appears once in the testing results.

3. PROCEDURE

3.1. Microarray breast cancer dataset

Two sets of breast cancer datasets were downloaded from the gene expression omnibus (GEO) [19], [20] for this study. GSE45255 and GSE15852 are the accession numbers, and the chip platform is GP96. GSE45255 only included 139 breast cancer patients. GSE15852, on the other hand, has 43 paired normal and breast cancer patients. These two datasets were combined together to form an integrate dataset with 182 breast cancer patients and 43 normal cases, each sample with 22,215 genes [21]. From this point forward, the combined dataset is referred as grating-outcoupled surface-emitting (GSE_integrate). In the dataset, when various platform of the probes was indicated to the same genes, the average of the probes was taken from a specific dataset, and the probes that started with "AFFX" were deleted as this data had no related genes for these probes [22]. The train and test data are split into an 80:20 ratio in this study.

Before classification is applied, some pre-processing method is essential. Two processing steps were implemented for the BC dataset. First, all sample were split into binary class where relapses were represented as set 1 and non-relapse were represented as 0 (a good prognosis). Second, the input features or gene values were normalized and standardized to the interval of [0,1]. The following is min-max normalization method [11].

$$x_n = \frac{x - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where Xn represents the normalized from input features data X, and $Xmin$ and $Xmax$ are the minimum and maximum number respectively. Format of original microarray breast cancer profiles before and after pre-processing methods are as shown in Tables 1 and 2.

Table 1. Example of original data before standardization and normalization pre-processing step

Probe ID_REF	Sample 1	Sample 2	Sample 3	Sample 4
1,007_s_at	1.80132	1.918762	2.097932	1.70628
1,053_at	0.12803	0.315474	-0.0551	-0.06975
117_at	0.274269	0.205618	0.247629	0.106031
121_at	0.580448	0.610526	0.670929	0.562654
1,255_g_at	-0.40625	-0.40902	-0.38888	-0.46145

Table 2. Example of format dataset used after standardization and normalization preprocessing step

Probe ID_REF	Sample 1	Sample 2	Sample 3	Sample 4
1007_s_at	0.4429	0.4854	0.5503	0.4085
1053_at	0.3875	0.6073	0.1728	0.1556
117_at	0.4312	0.3212	0.3886	0.1617
121_at	0.6002	0.6792	0.8379	0.5534
1255_g_at	0.2396	0.2355	0.2648	0.1595

3.2. Method

Three-machine learning model with CV (grid search) are investigated for classifying BC microarray data. The flow-chart is shown in Figure 2. The following steps describes the procedure of the methods:

- Dataset is split into training and testing data with ratio of 80:20 (80% for training data and 20% for testing data).
- During data splitting, a stratify method is applied to ensure that the training and testing ratio having an equally balance amount during training and testing the dataset. A scikit learn package from python library was used for module splitting and stratifying.
- The datasets were classified using SVM, RF and ET using k-fold CV method, in which *K* represents as 10. Using 10-fold CV, the data is split into 10 subsets, in which each fold had 9 subsets that used as training set, and the remaining subset will be used for the testing set.
- A hyperparameter tuning was applied for the machine learning model. Hyperparameters store the information that governs the training process and cannot be learned during the training process because it can increase capability of a model and results overfitting. Before running the experiments, a set of hyperparameters value need to be set. GridSearchCV was applied from scikit learn package in python to determine the best hyperparameters for the models. After this, the optimal hyperparameters gained from the GridSearchCV were used to re-train the model on the training set and to predict the accuracy value on the test. The optimal range gathered from hyperparameters value are different depending on the trained datasets and the models used. The output obtained from the dataset can be predicted to identify the performance on each dataset.

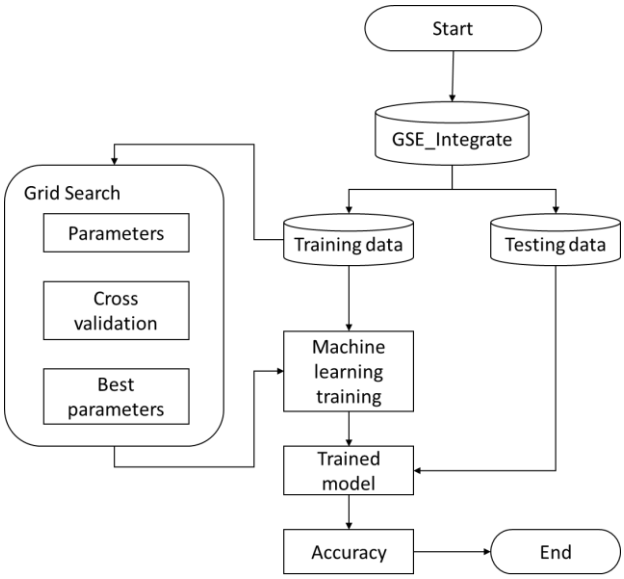

Figure 2. Workflow of BC microarray classification

3.3. Performance metric

Performance of all classifiers are evaluated by different measure metrics such as classification accuracy, f1-score, sensitivity, and specificity [11], [22].

3.3.1. Classification accuracy

Classification accuracy [11] is a commonly used evaluation criterion for a standard classification system and can be calculated using the following.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

TP represents true positive and correctly classified positive samples. TN represents true negative and correctly classified negative samples. FP represents false positive and misclassified negative samples and, FN represents false negative that misclassified positive samples.

3.3.2. F1-score

F1-score measures model's classification ability. The F1-score combines a classifier's precision and recall into a single metric by taking their harmonic mean. Its principal function is to compare the performance of two classifiers. Assume classifier A has a higher recall but classifier B has a higher precision. In this case, the F1-scores for both classifiers can be used to assess which one delivers superior results A perfect model has f1-score equivalent to 1. The formula of f1-score is in equation (3).

$$f1 = {2 * \left(\frac{TP}{TP+FP}\right) * \left(\frac{TP}{FN+TP}\right)} \Big/ {\left(\frac{TP}{TP+FP}\right) + \left(\frac{TP}{FN+TP}\right)} \quad (3)$$

3.3.3. Sensitivity and specificity

Sensitivity is also known as true positive rate (TPR) or recall. Sensitivity evaluates how well a model can recognize the classifier. It identifies proportion of accurately classified positive samples to total samples. Whereas the ability of a test to correctly identify person who do not have the disease is referred to as specificity.

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

$$Specificity = \frac{TN}{TN+FN} \quad (5)$$

4. SIMULATION RESULT AND DISCUSSION

This section discusses results obtained from all 3 classifiers models namely RF, ET and SVM, for binary class microarray breast cancer dataset. All the classifiers are implemented in the following environment, operating system: Windows 10, CPU: Intel Core i5-10210U (2.11 GHz), and memory: 8GB RAM. Table 3 shows the hyperparameters and their range tuned by the GridSearchCV. Hyperparameters setting that are not stated in this table were set to default values.

Table 4 show the best classification accuracies demonstrated by all three models with GridSearchCV [22] for the microarray BC dataset, GSE_integrated. The best result is obtained by SVM with 97.78% accuracy, 99% f1 score, 97% sensitivity and 100% specificity. This followed by RF and ET with both obtaining 93.33% accuracy. However, the accuracy obtained is lesser than 100%. This is due to the dataset does not have equal class ratios. This is known as imbalanced datasets. Although, the dataset is a binary data which has only two possible class: zero for normal and one for relapse, the imbalanced dataset makes it more challenging to train and predict. The lower sensitivity and higher specificity confirm the problem of imbalance data. All three algorithms achieve 100% specificity which indicates all samples classified as negative (normal) are correctly classified. This is due to the significantly lesser normal samples in the GSE_integrate. Figure 3 show the results of area under curve (AUC) of three models respectively RF, ET, and SVM. The results are significantly good, suggesting no overfitting. The time cost for overall tuning parameter takes around 1-2 minutes. However, in the future, the feature selection method will be used to choose relevant features that contribute to the characteristics of the BC microarray dataset [23]-[25].

Table 3. Hyperparameter setting

Classifier	Hyperparameter	Range
SVM	Kernel	[linear, poly, rbf]
	C	[0.1, 1, 10, 100, 1000]
	Gamma	[1, 0.1, 0.01, 0.001]
RF	Number of estimators	[100, 200, 300, 400, 500]
	Maximum depth	[2, 4, 6, 8, 10]
	Minimum samples leaf	[1, 5, 10]
ET	Number of estimators	[100, 200, 300, 400, 500]
	Criterion	[gini, entropy]

Table 4. Example of format dataset used after preprocessing step

Dataset	Classifier	Accuracy (%)	F1 (%)	Sensitivity (%)	Specificity (%)
GSE_integrated	SVM	97.78	99	97	100
	RF	93.33	96	92	100
	ET	93.33	96	92	100

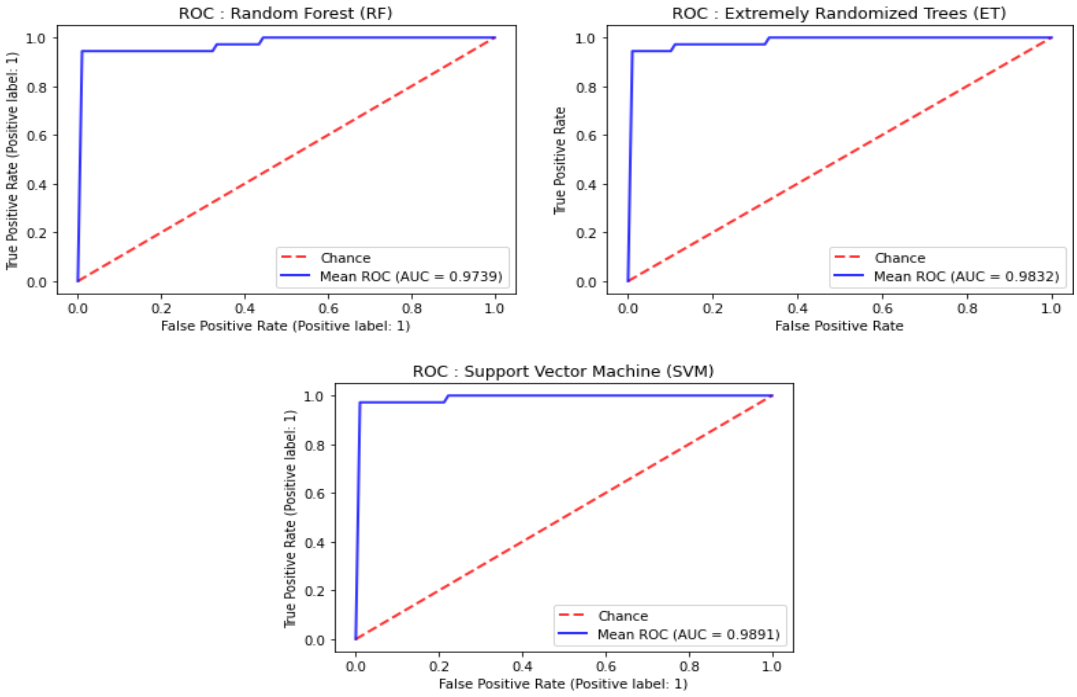

Figure 3. The receiver operating characteristic (ROC) and AUC curves and values of classifiers

5. CONCLUSION

Microarray data has thousands of features. The features are informative in diagnosis and prognosis of diseases including breast cancer. Machine learning algorithms are suitable for analysis of this type of data. They offer automated and faster system. Thus, this study applies RF, ET and SVM with simple parameter tuning based on GridSearchCV. The performance of the machine learning methods is compared using several performance metrics, accuracy, f1-score, sensitivity, and specificity. The data used, GSE_integrate, has 182 cancer/relapse samples and 43 normal samples. The result shows, the SVM method is the best model compared to RF and ET. In the future, the usage of the feature selection method to select relevant features that contribute to characteristics of the BC microarray dataset is to be investigated. Additionally, data balancing techniques are to be incorporated to tackle problem observed due to imbalance data.

ACKNOWLEDGEMENTS

This work is financially supported by Universiti Teknikal Malaysia Melaka, the Ministry of Higher Education Malaysia (MoHE) through SLAB Sponsorship Awards. The authors would like to thank Fisabilillah Research & Development Grant, Tabung Amanah Zakat under Funding Number: MMUE/180060 and Centre for Research and Innovation Management (CRIM) (UTeM) for financial supporting the paper. Also, thanks to the referred reviewers for their valuable comments and suggestions that made considerable improvement to our work.

REFERENCES

[1] C. Nalini and D. Meera, "Breast cancer prediction system using data mining methods," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 12, pp. 10901–10911, 2018.

[2] M. Akram, M. Iqbal, M. Daniyal, and A. U. Khan, "Awareness and current knowledge of breast cancer," *Biological Research*, vol. 50, no. 1, pp. 1–23, Dec. 2017, doi: 10.1186/s40659-017-0140-9.

[3] B. N. Hellquist, K. Czene, A. Hjälm, L. Nyström, and H. Jonsson, "Effectiveness of population-based service screening with mammography for women ages 40 to 49 years with a high or low risk of breast cancer: Socioeconomic status, parity, and age at

birth of first child," *Cancer*, vol. 121, no. 2, pp. 251–258, 2015, doi: 10.1002/cncr.29011.
[4] A. Thompson *et al.*, "Evaluation of the current knowledge limitations in breast cancer research: A gap analysis," *Breast Cancer Research*, vol. 10, no. 2, pp. 1–25, Apr. 2008, doi: 10.1186/bcr1983.
[5] R. Kumar, A. Sharma, and R. Tiwari, "Application of microarray in breast cancer: An overview," *Journal of Pharmacy and Bioallied Sciences*, vol. 4, no. 1, pp. 21–26, 2012, doi: 10.4103/0975-7406.92726.
[6] H. Almazrua and H. Alshamlan, "A comprehensive survey of recent hybrid feature selection methods in cancer microarray gene expression data," *IEEE Access*, vol. 10, pp. 71427–71449, 2022, doi: 10.1109/ACCESS.2022.3185226.
[7] H. Alshamlan, G. Badr, and Y. Alohali, "mRMR-ABC: A hybrid gene selection algorithm for cancer classification using microarray gene expression profiling," *BioMed Research International*, pp. 1–15, 2015, doi: 10.1155/2015/604910.
[8] H. C. King and A. A. Sinha, "Gene expression profile analysis by DNA microarrays: Promise and pitfalls," *JAMA*, vol. 286, no. 18, pp. 2280–2288, Nov. 2001, doi: 10.1001/jama.286.18.2280.
[9] S. Alagukumar and R. Lawrance, "Classification of microarray gene expression data using associative classification," in *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)*, Jan. 2016, pp. 1–8, doi: 10.1109/ICCTIDE.2016.7725362.
[10] N. Almugren and H. M. Alshamlan, "New bio-marker gene discovery algorithms for cancer gene expression profile," *IEEE Access*, vol. 7, pp. 136907–136913, 2019, doi: 10.1109/ACCESS.2019.2942413.
[11] P. Mohapatra, S. Chakravarty, and P. K. Dash, "Microarray medical data classification using kernel ridge regression and modified cat swarm optimization based gene selection system," *Swarm and Evolutionary Computation*, vol. 28, pp. 144–160, Jun. 2016, doi: 10.1016/j.swevo.2016.02.002.
[12] H. Alshamlan, G. Badr, and Y. Alohali, "A comparative study of cancer classification methods using microarray gene expression profile," in *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*, 2014, pp. 389–398, doi: 10.1007/978-981-4585-18-7_44.
[13] M. Maniruzzaman *et al.*, "Statistical characterization and classification of colon microarray gene expression data using multiple machine learning paradigms," *Computer Methods and Programs in Biomedicine*, vol. 176, pp. 173–193, Jul. 2019, doi: 10.1016/j.cmpb.2019.04.008.
[14] D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine learning classification techniques for breast cancer diagnosis," *IOP Conference Series: Materials Science and Engineering*, vol. 495, no. 1, pp. 1–16, Jun. 2019, doi: 10.1088/1757-899X/495/1/012033.
[15] D. Lakshmi, S. R. Gurrela, and M. Kuncharam, "A comparative study on breast cancer tissues using conventional and modern machine learning models," in *Smart Innovation, Systems and Technologies*, 2021, pp. 693–699, doi: 10.1007/978-981-16-0878-0_67.
[16] Ö. Akar and O. Güngör, "Classification of multispectral images using random forest algorithm," *Journal of Geodesy and Geoinformation*, vol. 1, no. 2, pp. 105–112, 2012, doi: 10.9733/jgg.241212.1.
[17] M. R. Karim, A. Rahman, J. B. Jares, S. Decker, and O. Beyan, "A snapshot neural ensemble method for cancer-type prediction based on copy number variations," *Neural Computing and Applications*, vol. 32, no. 19, pp. 15281–15299, Oct. 2020, doi: 10.1007/s00521-019-04616-9.
[18] Irawansyah, Adiwijaya, and W. Astuti, "Comparative analysis of support vector machine (SVM) and random forest (RF) classification for cancer detection using microarray," in *2021 9th International Conference on Information and Communication Technology (ICoICT)*, Aug. 2021, pp. 650–656, doi: 10.1109/ICoICT52021.2021.9527458.
[19] T. Barrett *et al.*, "NCBI GEO: Archive for functional genomics data sets—update," *Nucleic Acids Research*, vol. 41, no. 1, pp. 991–995, Nov. 2012, doi: 10.1093/nar/gks1193.
[20] M. Sun, T. Ding, X.-Q. Tang, and Y. Keming, "An efficient mixed-model for screening differentially expressed genes of breast cancer based on LR-RF," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 1, pp. 124–130, Jan. 2019, doi: 10.1109/TCBB.2018.2829519.
[21] A. Saini, J. Hou, and W. Zhou, "Breast cancer prognosis risk estimation using integrated gene expression and clinical data," *BioMed Research International*, pp. 1–15, 2014, doi: 10.1155/2014/459203.
[22] G. N. Ahmad, H. Fatima, S. Ullah, A. S. Saidi, and Imdadullah, "Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV," *IEEE Access*, vol. 10, pp. 80151–80173, 2022, doi: 10.1109/ACCESS.2022.3165792.
[23] N. M. Ali, R. Besar, and N. A. Ab. Aziz, "Hybrid feature selection of breast cancer gene expression microarray data based on metaheuristic methods: A comprehensive review," *Symmetry*, vol. 14, no. 10, pp. 1–33, Sep. 2022, doi: 10.3390/sym14101955.
[24] H. Motieghader, A. Najafi, B. Sadeghi, and A. Masoudi-Nejad, "A hybrid gene selection algorithm for microarray cancer classification using genetic algorithm and learning automata," *Informatics Med. Unlocked*, vol. 9, pp. 246–254, 2017, doi: 10.1016/j.imu.2017.10.004.
[25] H. Lu, J. Chen, K. Yan, Q. Jin, Y. Xue, and Z. Gao, "A hybrid feature selection algorithm for gene expression data classification," *Neurocomputing*, vol. 256, pp. 56–62, 2017, doi: 10.1016/j.neucom.2016.07.080.

BIOGRAPHIES OF AUTHORS

Nursabillilah Mohd Ali was born in Melaka, Malaysia, in 1985. She received the B.Eng. (Hons.) and M.Sc. degrees from Universiti Teknikal Malaysia Melaka and International Islamic University, Malaysia in 2009 and 2014, respectively, all in mechatronic engineering. She has been an academic staff since 2009, where now she is a Senior Lecturer of Universiti Teknikal Malaysia Melaka. She is a Chartered Engineer of the Engineering Council UK and a Graduate Engineer of the Board of Engineers Malaysia. Currently, she is working toward the Ph.D. degree at the Multimedia University. Her research interests include bioinformatics system, DNA gene expression, optimization algorithm and machine learning. She can be contacted at email: nursabillilah@utem.edu.my.

Rosli Besar is currently Associate Professor of Multimedia University. He received the B.Eng. (Hons) and M.Sc. degrees from the University of Science Malaysia (USM), Malaysia, in 1990 and 1993, respectively and the Ph.D. degree from the Multimedia University, Malaysia, in 2004. His current interests include signal and image processing and medical imaging. He can be contacted at email: rosli@mmu.edu.my.

Nor Azlina Ab Aziz received her Ph.D. degree from University of Malaya, Malaysia. She is currently a senior lecturer with the Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia. She is also the chairperson of the Center for Engineering Computational Intelligence, Multimedia University. She has published and presented numerous scientific papers in international journals and conferences and lead multiple research projects. Her research interests include the fundamental aspects and applications of computational intelligence in wireless communication, bioinformatics, operational research, and affective computing. She can be contacted at email: azlina.aziz@mmu.edu.my.