Comparative evaluation of data mining algorithms in breast cancer

Mustafa Qahtan Alsudani[1,2], Hassan Falah Fakhruldeen[1,3,4], Israa Fayez Yousif [5]

[1]Department of Computer Techniques Engineering, Faculty of Information Technology, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq
[2]Faculty of Medicine, Jabir Ibn Hayyan Medical University, Najaf, Iraq
[3]Department of Electrical Engineering, Faculty of Engineering, University of Kufa, Kufa, Najaf, Iraq
[4]Department of Computer Technical Engineering, College of Technical Engineering, The Islamic University, Najaf, Iraq
[5]Department of Materials Engineering, Faculty of Engineering, University of Kufa, Kufa, Najaf, Iraq

Article Info

ABSTRACT

Unchecked breast cell growth is one of the leading causes of death in women globally and is the cause of breast cancer. The only method to avoid breast cancer-related deaths is through early detection and treatment. The proper classification of malignancies is one of the most significant challenges in the medical industry. Due to their high precision and accuracy, machine learning techniques are extensively employed for identifying and classifying various forms of cancer. The authors of this review studied numerous data mining algorithms and implemented them such that clinicians might use them to accurately detect cancer cells early on. This article introduces several techniques, including support vector machine (SVM), K star (K*) classifier, additive regression (AR), back propagation (BP) neural network, and Bagging. These algorithms are trained using a set of data that contains tumor parameters from breast cancer patients. Comparing the results, the authors found that SVM and Bagging had the highest precision and accuracy, respectively. Also assess the number of studies that provide machine learning techniques for breast cancer detection.

Corresponding Author:

Hassan Falah Fakhruldeen
Computer Techniques Engineering Department, Faculty of Information Technology
Imam Ja'afar Al-Sadiq University
Baghdad, Iraq
Email: hassan.fakhruldeen@gmail.com

1. INTRODUCTION

The data mining industry utilizes computational, statistical, and optimization techniques to "read" data from historical instances and find difficult-to-data models from vast, noisy, or complex datasets. These properties, which depend on complicated proteomic and genomic measurements, are ideally suited for medicinal applications [1], [2]. Cancer is diagnosed and identified using data mining techniques such as the support vector network, the Bayesian confidence network, and the artificial neural network (ANN) [3], [4]. Fluorodeoxyglucose (FDG) positron emission tomography (PET) [5] enable clinicians to analyze the role of the tumor in the human body by recovering radiation symbol tracers. In recent years, a variety of machine learning [6]-[9], bio-inspired computation techniques [10]-[16], and deep learning [17] have been applied to make medical predictions. Recently, computer training has expanded to include cancer detection and prognosis.

The survey revealed that some techniques are optimal for testing the usefulness of datasets [18]. Various essential data extraction technologies are being improved and deployed in a variety of real-world

applications (e.g., healthcare, bioscience, and industry) to extract useful data bits from individual data to aid in decision-making [19]. In machine learning environments, a reasonably large amount of data consisting of actual medical cases of men diagnosed with prostate cancer who receive medical attention is used for the systematic comparison of procedure. Methods of machine learning (ML) are software programs that predict anything (behavior, the form of a disease, and the picture of stock price volatility) based on the conditions that led to prior events [20], [21]. Breast cancer is the most prevalent malignancy in women worldwide. Breast cancer is caused by the early growth of specific breast cells. Several methods have been developed for detecting breast cancer. Breast imaging is a type of mammography [22]. Breast imaging is sometimes known as mammography. Is a method for diagnosing breast cancer. X-rays are used to assess the condition of female nipples. It is nearly impossible to detect breast cancer in the observable external cancer cell in its first stages. Mammography can detect cancer at an early stage and just takes a few minutes.

Dynamic magnetic resonance imaging (MRI) [23] has established the detection strategy for breast deformation. The method forecasts the growth rate of tumor angiogenesis. Magnetic reasoning imaging led to a reduction in contrast metastases in breast cancer patients. Ultrasound is a well-known tool for detecting symptoms within a sound wave traveling through the body [24]. A transducer generating sound is put on the skin, and the sound waves capture the tissue reflections [25]. Electrography is a freshly developed technology based on images [26]. This method is utilized when breast cancer tissues are larger than the neighboring normal parenchyma. A sample compression color map distinguishes between benign and malignant types. Although numerous approaches have been validated, none of them can produce an accurate and consistent outcome. In mammography, the doctors must interpret a large amount of picture data, which reduces accuracy.

It takes time, and in some of the worst cases, incorrectly diagnoses the disease. This article compares multiple machine learning algorithms for disease diagnosis using data. To accurately detect the condition, six supervised machine-learning approaches were employed. The breast cancer dataset was classified using the decision tree, Bayes, and neural net techniques [27]. The experiment concludes that the neural net classifies breast tumors with greater sensitivity and accuracy than the decision tree and Naive Bayes methods. A machine has been created in [28] to aid in the differentiation between malignant and benign cancers. The backward elimination (BE) method was combined with the random forest tree to determine functionality. The dataset was gathered from the predictive network in Wisconsin. The accuracy of this hybridization approach is roughly 95%, and the number of variables has been lowered from 33 to 17 to 18 It is utilized to evaluate three related algorithms, namely support vector machine (SVM), ANN, and decision tree (DT) [29]. This investigation utilizes a database collected from the Iranian Center (ICBC). Algorithm SVM produced up to 95% accuracy using a total of 8 predictor variables. Rana [30] developed the SVM collection with a decision tree (C5.0) model for breast cancer detection. The dataset was produced by integrating 32 parts of the prognostic dataset from Wisconsin. Rank-based role selection was utilized to achieve the variable reduction. The performance of the radial base function for five characteristics is 92.59%. Several statistical models of contemporary machine learning technologies used to detect cancer progression were addressed [30].

The author has reviewed numerous ML-related literature in this study. Each category and classification of the papers varies with respect to the dataset and its characteristics. The precision of mammographic data is as high as 83%, while the precision of other datasets is as high as 71%. Frequently, these papers contain up to 14 variables of mammograms, with the precision of mammographic data as high as 83%. Comparative experiments in [31], [32] featured numerous machine learning algorithms for breast cancer estimate and diagnosis, including SVM, logistic regression, Naive Bayes, and the k-nearest neighbors (KNN). 95.6% and 68% of the breast cancer recurrence and non-recurrence data from the Wisconsin prognostic breast cancer data repository were utilized for the analysis, respectively. Several classification algorithms, such as director tree (AD), decision tree (j48) algorithm, and best first tree (B+ tree), were executed. The dataset was obtained from the diagnostic center of Swami Vivekananda in Chennai. It contains 220 medical data and is used to evaluate nine characteristics. The outcome indicates that 99% of four algorithms are j48 [33].

A breast cancer solution was developed that differentiates between various forms of breast cancer. The method focuses on the diagnosis and estimation of breast cancer in Wisconsin, as well as the identification of multiple types of breast cancer [34]. Utilizing and analyzing two distinct migratory topologies in Iceland allows for a more precise and time-efficient training technique. A prognosis of illness status was presented utilizing a hybrid method for anticipating improvements and their repercussions, which are crucial to deadly infections [35]. Their strategy to alerting the public about the severity of diseases consists of two primary components: i) treatment and extraction of informational choices; and ii) decision based on the tree-support hybrid model for predictions DT-SVM. To construct accurate predictive models for breast cancer utilizing data mining approaches, they studied Wisconsin machine learning datasets from unique client identifier (UCI). Table 1 shows then list of publications comparing parameters, accuracy, and various algorithms for breast cancer diagnosis

Table 1. Presents a list of publications comparing parameters, accuracy, and various algorithms for breast cancer diagnosis

Method	Accuracy	Objective	Features
Neural networks	96.14%	Implementation of breast cancer prediction classification techniques [25]	9
Random forest tree	99%	Random forest classification along with breast cancer detection and prognostic function collection [26]	17
Neural Network SVM and DT	95%	Three machine learning models for predicting breast cancer recurrence [27]	8
SVM	92.59%	Breast cancer identification using decision tree and support vector ensemble with reduced function subset [28]	5
various predictive model	83%	Applications for machine learning for cancer prognosis [29]	14
(SVM), Logistic Regression, Naïve Bayes and (KNN)	95.6%	ML models for identifying breast cancer and forecasting recurrence [30]	----
j48, Best First Tree (B+ tree), and AD tree	99%	Performance analyzes of breast cancer decision tree algorithms [31]	4
ISLAND APPROACH TO NEUROL NETWORK DIFFERENTIAL	99.97%	A neural network approach to characterization and breast cancer diagnosis [32]	9
Support Vector Machine and Decision tree	95%	Big-data extraction: DT-SVM hybrid platform for breast cancer prediction [33]	9
Fuzzy inference system	93%	Breast cancer risk identification Mamdani fuzzy inference method [34]	4
Support Vector Machine	97%	A significance vector machine learning technique is used to classify cancer [35]	4
Naïve Bayes	92%	Bayes Weighted classification: a breast cancer detection predictive plan [36]	9
CNN	82.73%	In mammogram imaging applications, convolutional neural networks are used to detect breast cancer [37]	
SVM and ANN	97.14%	An examination of the effect of artificial neural networks and vector-supporting devices on the diagnosis of breast cancer [38]	subset
backpropagation Artificial Neural Network	91.7%	Textural feature-based mammogram classification using ANN [39]	
c-SVM and v-SVM	97.68%	A vector-based ensemble algorithm supports the detection of breast cancer [40]	
RepTree	81.3%	prediction Particle swarm	
Naive Bayes	80%	Selection for breast cancer recurrence	
k-NNs	75%	Optimization [41]	
Normalized Multilayer Perceptron Neural Network	99%	Created a model employing a normalized multilayer perceptron neural network. The findings obtained are excellent [42]	14
CNN	98%	For breast cancer classification, a new deep Feed forward NN model with four AFs has been proposed: Swish, hidden layer 1; LeakyReLU, hidden layer 2; ReLU, hidden layer 3; and naturally Sigmoidal final feature layer [43]	4
SVM	90%	four distinct classification models, including SVM, KNN, Naïve Bayes,	
KNN	87%	and DT, with characteristics picked at varying threshold levels. The	
Naïve Bayes	85%	suggested approaches were applied to separate gene expression datasets	
DT	87%	for performance evaluation and validation [44]	

Three classification technologies in the Weka program are equivalent, with the DT-SVM being more predictive than naive classification and sequential minimal optimization. Fuzzy logic was utilized in [36] to detect the existence of breast cancer. This study collects information from the UCI learning repositories. The objective is to detect breast cancer by lowering the causes of the disease and reducing the time required for diagnosis. The localiser type directional Aid (LDA) procedure was utilized to choose the feature, while the Fuzzy Mamdani method was employed to teach it. A fuzzy deduction is a technique for inference. Fuzzy logic was utilized to evaluate the outcomes. 93% of the findings were made available to the public. Tan *et al.* [37] developed an effective machine learning strategy for cancer classification that could improve cancer classification accuracy. The project consists of two phases. Utilizing the Ariance analysis (ANOVA) scoring scheme to select the essential genes is the initial stage. The classification task in the second step requires the application of a suitable classifier. Two of the most effective machine learning classifiers were relevance vector machine (RVM) learning and fuzzy support vector machine (FSVM) learning. To compute the testing values, three origin points, Lymphoma, Leukemia, and Small round blue cell tumors (SRBCT) data sets, have been defined. The Naive Bayes classifier explored the performance criterion of the machine learning method employing a new weighted approach to breast cancer classification in [38]. Weighted ideas are implemented to expand and improve the performance of standard Naive Bayes models. Breast cancer dataset-based domain-awareness weight assignment utilizing the machine learning library at UCI. The experiments show that an approach to heavy naive berries is preferable to the naive method. A convolutional neural network (CNN)-based approach was developed in [39] to aid in the rapid diagnosis of problems by physicians. Using enhanced mammography pictures, the classifier constructed a model to detect the tumor in cancer patients. The proposed approach features a high degree of accuracy and a quick diagnostic time. SVM and ANN are suggested as templates for mixed feature series in [40].

The classification tasks were accomplished utilizing a distinct combination of feature subsets by determining the optimal parameters and dividing the results. The SVM displayed superior classification of data with a precision of 97,1388%, compared to the ANN's accuracy of 96,7096%. Sakri *et al*. [41] a suggestion to identify breast cancer using law's texture energy measurement (ITEM) instrument. The backpropagation artificial neural network (BPANN) classifies malignant, normal, and natural tissue sections using an approach. 90,9% of the typical irregular grouping is responsive to the proposed technique, with 94.4% accuracy. The accuracy for benign versus malignant classification is 91.7% and 66.6%, respectively. SVM weighted area, unless the recipient's operational curve includes a variation of the breast cancer learning ensemble, the breast cancer learning ensemble is not included area under curve (AUC) [42]. Six C-SVM and v-SVM kernel functions can be added to the standard model package. It was demonstrated that the proposed model considerably improves breast cancer diagnosis. The accuracy of the model was 97.68. They published a report regarding the relationship between classifications without using a feature selection scheme. 70%, 76.33%, and 66.33% of the time, respectively, are generated by the Naive Bayes, RepTree, and K-NNs. They process their results using the Weka platform. With the deployment of particle swarm optimization (PSO), the four most advantageous characteristics for this categorization function were determined. With PSO, the precision rates for Naive Bayes, RepTree, and K-NN were 81.3%, 80%, and 75.3%, respectively. The researchers constructed a model utilising a normalized multilayer perceptron neural network to accurately identify breast cancer [43], [44].

The obtained results are exceptional (accuracy is 99.27%). In comparison to other studies employing ANN, this result is extremely encouraging. As a control test, breast cancer Wisconsin (Original) was utilized. To build models for identifying the two types of breast cancer, researchers employed four distinct classification models, including SVM, KNN, Naive Bayes, and DT, with attributes chosen at varying threshold levels [45]. The suggested methodologies were applied to distinct gene expression datasets in order to evaluate and validate their performance. The SVM algorithm effectively classified breast cancer into triple negative and non-triple negative subtypes with less misclassification errors than the other three algorithms evaluated.

2. METHOD
2.1. Preprocess of data mining
ML approaches can separate the learning process into two groups: unsupervised and supervised. Diverse data instances are employed and tagged to achieve optimal performance for training the system for unsupervised instruction. However, there are no predetermined knowledge sets available in education, making the goal impossible to attain. The results are not anticipated. Classification is one of the most prevalent forms of regulated schooling. It utilizes previous data to establish a benchmark for future forecasts. Utilizing historical data. In the realm of medicine, clinics and hospitals keep huge databases containing patient histories and symptom diagnosis. Using this knowledge, researchers then develop categorization models based on historical events. Thus, medical inference with the aid of computers has gotten simpler, given the sheer volume of medical data available today [46].

2.2. Neural network
An ANNs reduced sight is a knowledge-based computational algorithm. The meaning and function of ANN are identical to those of the human brain. The observation of raw data reveals correlations and broad patterns [47]. There are relationships between the network's weight and nodes. There are four hidden layers in the neural network: center, input, and output. Each of these layers is connected to the neural network using a weight connector [48]. This research presents a neural network that employs a multilayered vision and a back propagation. In the back spreading network, there are three distinct layers (input, hidden, and output) in which a signal traverses one path such that it does not return to its source after conveying the neuronal output from the input neuron. Figure 1 illustrates the neural network.
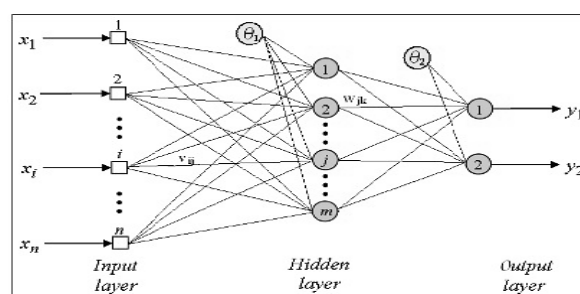

Figure 1. Back propagation neural network [48]

2.3. K-star classification algorithm

K-star, often known as K*, is an instance-based classifier. Using a correlation function, this approach attempts to determine if the instance is connected to any of the training datasets. This method differs from other instance-based learners since it uses an entropy-based function. This function classifies the circumstance by allocating it to a predefined and classed data set model. The crucial aspect of this hypothesis is that similar circumstances impart similar categories [49].

2.4. Additive regression

A meta-classifier that enhances the visual appeal of a regression-based base classifier. Every repeat offers a pattern to the residuals generated by the classifier during the previous iteration. Attaching the predictions of any reliable classifier yields the forecast. Overcoming the reduction (learning rate) parameter facilitates limit overfitting and generates a smoothing effect, but improves learning [50].

2.5. Bagging

Bagging is a method for improving the performance of classification algorithms in machine learning. This technique was described by Leo Breiman, and its name was derived from the term "bootstrap aggregation" [51]. On the basis of a foundational set of example data D, a classification algorithm generates a classifier H: D-1,1 for categorization within a couple of possible categories. The bagging technique generates a sequence of classifiers Hm, m=1,..., M based on the training set's qualities. [52] These classifiers are combined to create a composite classifier.

2.6. Support vector machine

The support vector machine method separates the data set into hyperplane margin-appropriate groups. This method is often used in the field of medicine to diagnose the ailment. Given that a data set may contain many hyper lines, the SVM algorithm attempts to produce a maximal difference between various groups by maximizing the limit [53]. Observe the darkness of this dataset's groups. On a single line, the Wisconsin breast cancer dataset cannot be separated into perfect groups. Figure 2 depicts the shadows.

Using transformation to address this problem and add a Z-axis dimension. When a dataset is shown on the Z-axis, the stark distinction between groups is now readily obvious. The procedure is carried out using kernels. Polynomial and exponential kernels measure a separation line with a higher dimension. Figure 3 illustrates the significance of kernels in deciding the acquisition of dark data.
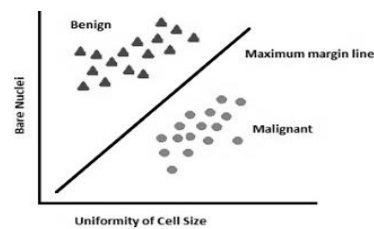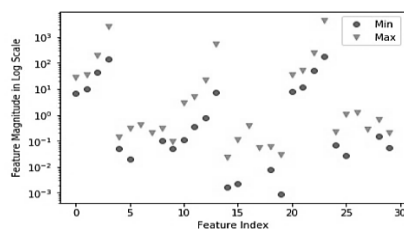

Figure 2. Wisconsin dataset class distribution [52]   Figure 3. Using the kernel for classification [54]

3.   RESULTS AND DISCUSSION

The training data collection used in this work was taken from the breast cancer data of Wisconsin. This data set is available in the ML repository at UCI [54]. Its data collection is multivariate and has more than 569 examples. Based on more than 10 features, a digital image's cell nuclei are classified as malignant or benign [55]. The 10 qualities are (Area, Compactness: (p*p/a-1), where p is the perimeter and area, concave points: the number of concave contour sections, and a coastline approximation is a type of fractal. Concavity is the endpoint of a concave contour component. This paper includes the Anaconda software as a tool for teaching machines. Anaconda is a Python-based open-source program that was first released under the new Berkeley source distribution (BSD) License in 2012. It contains several machine learning algorithms and approaches, including the algorithms covered in this article. Python and programming are free and open-source languages for scientific computation [56]. This program also offers data science, large-scale data management, and predictive analytics. After comparing the outcomes, the parameters and results of five distinct data mining models are displayed. The authors discovered that SVM and K* star had the highest accuracy at 98.6%, followed by Bagging at 97.3%, BP at 97%, and additive regression at 93.6%. In addition, we examine the number of studies that have developed machine learning algorithms for breast cancer detection. Identification rate is often referred to as accuracy. The definition of the accuracy measure is the number of instances properly identified divided by the total number of instances in the data collection.

The precision for various sets can be altered and is highly dependent on the classification threshold [57]. The precision may be determined by (1).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

TP is the true positive, whereas TN is the true negative. The complete shape of P is also positive, signifying cancer cells, whereas N denotes negatives and benign non-cancerous cells. Precision is frequently equated with assurance. Precision is determined by the rates of true positive cases and true positive instances. Precision demonstrates the classifier's ability to deal with good events, but has minimal bearing on unfavorable scenarios. Precision and recall are proportional to one another [58]. This parameter can be determined by (2).

$$\text{Precision} = \frac{TP}{TP+FN} \tag{2}$$

The recall is depicted as false-negative and overly hopeful examples. This metric is utilized in the medical industry since it provides information on the correct classification of the number of malignant and benign cases. The model will locate all cases in the dataset that are pertinent. With the equation, one can calculate the recall. Table 2 shows the comparison and contrasts the recall and accuracy of the Wisconsin dataset for five machine learning techniques.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

Table 2. The comparison and contrasts the recall and accuracy of the Wisconsin dataset for five machine learning techniques

Comparison of Wisconsin breast cancer classification algorithms			
Method	Accuracy	Precision	Recall
Support Vector machine	98.6%	97.5%	96.9%
K star	98.6%	97.3%	97%
Bagging	97.3%	97.2%	96.8%
Neural network	97%	97.2%	96.9%
Additive regression	93.5%	96.2%	93.2%

4.   CONCLUSION

Cancer of the breast is the most common form of the disease in the world. A woman chosen at random has a 13% probability of having the condition diagnosed. A significant number of lives can also be saved with the early identification of breast cancer. SVM, K*, BP, Bagging, and additive regression are the five methods of machine learning that are discussed in this article for predicting breast cancer. We looked at five different data mining methodologies and rated them based on their recall, precision, and accuracy. Comparisons of the efficacy of the various algorithms have also been made, using the Wisconsin dataset as the basis. In their research, the authors found that the SVM and K* have the highest degree of precision.

REFERENCES

[1]   S. Kharya, D. Dubey, and S. Soni, "Predictive machine learning techniques for breast cancer detection," *International Journal of Computer Science and Information Technologies*, vol. 4, no. 6, pp. 1023–1028, 2013.

[2]   S. Ismael, S. Kareem, and F. Almukhtar, "Medical image classification using different machine learning algorithms," *AL-Rafidain Journal of Computer Sciences and Mathematics*, vol. 14, no. 1, pp. 133–145, Jun. 2020, doi: 10.33899/csmj.2020.164682.

[3]   P. Suryachandra and P. V. S. Reddy, "Comparison of machine learning algorithms for breast cancer," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, 2016, vol. 2016, pp. 1–6, doi: 10.1109/INVENTIVE.2016.7830090.

[4]   S. Kareem and M. Cudi Okur, "Evaluation of bayesian network structure learning," in *2nd International Mediterranean Science and Engineering Congress*, 2017, no. Imsec, pp. 1–7.

[5]   S. W. Kareem and M. C. Okur, "Pigeon inspired optimization of bayesian network structure learning and a comparative evaluation," *Journal of Cognitive Science*, vol. 20, no. 4, pp. 535–552, Dec. 2019, doi: 10.17791/jcs.2019.20.4.535.

[6]   N. Mallios, E. Papageorgiou, M. Samarinas, and K. Skriapas, "Comparison of machine learning techniques using the WEKA environment for prostate cancer therapy plan," in *Proceedings of the 2011 20th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE 2011*, 2011, pp. 151–155, doi: 10.1109/WETICE.2011.28.

[7]   A. Karplus, "Machine learning algorithms for cancer diagnosis," *Santa Cruz County Science Fair*, 2012.

[8]   A. S. Mohammed, S. W. Kareem, A. K. Al Azzawi, and M. Sivaram, "Time series prediction using SRE-NAR and SRE-ADALINE," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 10, no. 12 Special Issue, pp. 1716–1726, 2018.

[9]   M. Mori *et al.*, "Diagnostic accuracy of contrast-enhanced spectral mammography in comparison to conventional full-field digital mammography in a population of women with dense breasts," *Breast Cancer*, vol. 24, no. 1, 2017, doi: 10.1007/s12282-016-0681-8.

[10]   H. A.-J. Al-Asady, H. F. Fakhruldeen, and M. Q. Alsudani, "Channel estimation of OFDM in C-band communication systems under different distribution conditions," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 23, no. 3, pp. 1778-1782, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1778-1782.

[11] A. G. Mohammed, "A study of scheduling algorithms in LTE -advanced HetNet," *Qalaai Zanist Scientific Journal*, vol. 6, no. 3, Sep. 2021, doi: 10.25212/lfu.qzj.6.3.33.

[12] S. Kareem, "Hybrid public key encryption algorithms for e-commerce," Msc. Thesis, Erbil: University of Salahaddin–Hawler, 2009.

[13] C. S. Park, S. H. Kim, N. Y. Jung, J. J. Choi, B. J. Kang, and H. S. Jung, "Interobserver variability of ultrasound elastography and the ultrasound BI-RADS lexicon of breast lesions," *Breast Cancer*, vol. 22, no. 2, 2015, doi: 10.1007/s12282-013-0465-3.

[14] A. T. Azar and S. A. El-Said, "Probabilistic neural network for breast cancer classification," *Neural Computing and Applications*, vol. 23, no. 6, pp. 1737–1751, Nov. 2013, doi: 10.1007/s00521-012-1134-8.

[15] H. F. Fakhruldeen, T. S. Mansour, F. Ibrahim Jabbar, and A. Alkhayyat, "Multiple inputs all-optical logic gates based on nanoring insulator-metal-insulator plasmonic waveguides," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, pp. 6836-6846, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6836-6846.

[16] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast cancer prediction: A comparative study using machine learning techniques," *SN Computer Science*, vol. 1, no. 5, p. 290, Sep. 2020, doi: 10.1007/s42979-020-00305-w.

[17] I. K. Teeba and I. I. Hamarash, "Model-based quality assessment of internet of things software applications: A systematic mapping study," *International Journal of Interactive Mobile Technologies*, vol. 14, no. 9, 2020, doi: 10.3991/ijim.v14i09.13431.

[18] A. Sharma, A. Miran, and Z. R. Ahmed, "The 3D facemask recognition: Minimization for spreading COVID-19 and enhance security," in *Lecture Notes in Networks and Systems*, vol. 314, 2022, pp. 619–637, doi: 10.1007/978-981-16-5655-2_60.

[19] H. J. M. Al-Alkawi, A. O. Hanfesh, and S. M. N. Mohammed Rauof, "Development of high performance (mechanical and wear properties) of AA 6061-hybrid nano composites via liquid metallurgy route," *Al-Nahrain Journal for Engineering Sciences*, vol. 22, no. 2, pp. 143–150, Jul. 2019, doi: 10.29194/njes.22020143.

[20] A. J. Qasim, R. Din, and F. Q. A. Alyousuf, "Review on techniques and file formats of image compression," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 2, pp. 602–610, 2020, doi: 10.11591/eei.v9i2.2085.

[21] M. K. Hasan, M. M. Islam, and M. M. A. Hashem, "Mathematical model development to detect breast cancer using multigene genetic programming," in *2016 5th International Conference on Informatics, Electronics and Vision, ICIEV 2016*, May 2016, pp. 574–579, doi: 10.1109/ICIEV.2016.7760068.

[22] M. V. Prakash, V. Porkodi, S. Rajanarayanan, M. Khan, B. F. Ibrahim, and M. Sivaram, "Improved conservation of energy in fog IoT services using machine learning model," in *2020 International Conference on Computing and Information Technology, ICCIT 2020*, Sep. 2020, pp. 1–4, doi: 10.1109/ICCIT-144147971.2020.9213719.

[23] S. W. Kareem, R. Z. Yousif, and S. M. J. Abdalwahid, "An approach for enhancing data confidentiality in hadoop," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 20, no. 3, pp. 1547–1555, Dec. 2020, doi: 10.11591/ijeecs.v20.i3.pp1547-1555.

[24] S. Islam Ayon and M. Milon Islam, "Diabetes prediction: a deep learning approach," *International Journal of Information Engineering and Electronic Business*, vol. 11, no. 2, pp. 21–27, Mar. 2019, doi: 10.5815/ijieeb.2019.02.03.

[25] S. Chaurasia, P. Chakrabarti, and N. Chourasia, "An application of classification techniques on breast cancer prognosis," *International Journal of Computer Applications*, vol. 59, no. 3, pp. 6–10, Dec. 2012, doi: 10.5120/9526-3946.

[26] 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. 06, no. 05, pp. 551–560, 2013, doi: 10.4236/jbise.2013.65070.

[27] L. Ahmad and A. Eshlaghy, "Using three machine learning techniques for predicting breast cancer recurrence," *Journal of Health & Medical Informatics*, vol. 04, no. 02, 2013.

[28] I. H. Al-Kharsan, A. Al-Khaykan, and H. F. Fakhruldeen, "Expansion of high-voltage overhead transmission lines to remote areas," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 13, no. 4, pp. 2488–2497, Dec. 2022, doi: 10.11591/ijpeds.v13.i4.pp2488-2497.

[29] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, 2015, doi: 10.1016/j.csbj.2014.11.005.

[30] M. Rana, "Breast cancer diagnosis and recurrence prediction using machine learning techniques," *International Journal of Research in Engineering and Technology*, vol. 04, no. 04, pp. 372–376, Apr. 2015, doi: 10.15623/ijret.2015.0404066.

[31] E. Venkatesan, "Performance analysis of decision Tree algorithms for breast cancer classification," *Indian Journal of Science and Technology*, vol. 8, no. 1, pp. 1–8, Jan. 2015, doi: 10.17485/ijst/2015/v8i29/84646.

[32] H. T. Tike Thein and K. M. Mo Tun, "An approach for breast cancer diagnosis classification using neural network," *Advanced Computing: An International Journal*, vol. 6, no. 1, pp. 1–11, Jan. 2015, doi: 10.5121/acij.2015.6101.

[33] K. Sivakami, "Mining big data : Breast cancer prediction using DT - SVM hybrid model," *Internation Journal of Scientific Engineering and Applied Science*, no. 5, pp. 418–429, 2015.

[34] H. F. Fakhruldeen, H. Abdul-Jaleel Al-Asady, T. Mahinroosta, F. Sohrabi, and S. M. Hamidi, "Novel add-drop filter based on serial and parallel photonic crystal ring resonators (PCRR)," *Journal of Optical Communications*, Dec. 2020, doi: 10.1515/joc-2021-0220.

[35] A. Bharathi and K. Anandakumar, "Cancer classification using relevance vector machine learning approach," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 3, pp. 630–634, Jun. 2015, doi: 10.1166/jmihi.2015.1412.

[36] S. Kharya and S. Soni, "Weighted naive bayes classifier: A predictive model for breast cancer detection," *International Journal of Computer Applications*, vol. 133, no. 9, pp. 32–37, Jan. 2016, doi: 10.5120/ijca2016908023.

[37] Y. J. Tan, K. S. Sim, and F. F. Ting, "Breast cancer detection using convolutional neural networks for mammogram imaging system," *Proceeding of 2017 International Conference on Robotics, Automation and Sciences, ICORAS 2017*, vol. 2018-March, pp. 1–5, 2018, doi: 10.1109/ICORAS.2017.8308076.

[38] R. Alyami *et al.*, "Investigating the effect of correlation based feature selection on breast cancer diagnosis using artificial neural network and support vector machines," in *2017 International Conference on Informatics, Health and Technology, ICIHT 2017*, Feb. 2017, pp. 1–7, doi: 10.1109/ICIHT.2017.7899011.

[39] H. A. J. Al-Asady and H. F. Fakhruldeen, "Volts/Hz control of three phase induction machine: MATLAB Simulink," *Journal of Physics: Conference Series*, vol. 1818, no. 1, p. 012089, Mar. 2021, doi: 10.1088/1742-6596/1818/1/012089.

[40] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *European Journal of Operational Research*, vol. 267, no. 2, pp. 687–699, Jun. 2018, doi: 10.1016/j.ejor.2017.12.001.

[41] S. B. Sakri, N. B. Abdul Rashid, and Z. Muhammad Zain, "Particle swarm optimization feature selection for breast cancer recurrence prediction," *IEEE Access*, vol. 6, pp. 29637–29647, 2018, doi: 10.1109/ACCESS.2018.2843443.

[42] E. Alickovic and A. Subasi, "Normalized neural networks for breast cancer classification," in *IFMBE Proceedings*, vol. 73, 2020, pp. 519–524, doi: 10.1007/978-3-030-17971-7_77.

[43] K. Vijayakumar, V. J. Kadam, and S. K. Sharma, "Breast cancer diagnosis using multiple activation deep neural network," *Concurrent Engineering Research and Applications*, vol. 29, no. 3, pp. 275–284, Sep. 2021, doi: 10.1177/1063293X211025105.

[44] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," *Journal of Personalized Medicine*, vol. 11, no. 2, pp. 1–12, Jan. 2021, doi: 10.3390/jpm11020061.

[45] H. F. Fakhruldeen and T. S. Mansour, "Design of plasmonic not logic gate based on insulator – metal – insulator (IMI) waveguides," *Advanced Electromagnetics*, vol. 9, no. 1, pp. 91–94, Apr. 2020, doi: 10.7716/aem.v9i1.1376.

[46] S. M. R. K. Al-Jumur, S. Wahhab Kareem, and R. Z. Yousif, "Predicting temperature of Erbil City applying deep learning and neural network," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 2, pp. 944-952, May 2021, doi: 10.11591/ijeecs.v22.i2.pp944-952.

[47] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89--109, 2001, doi: 10.1016/S0933-3657(01)00077-X.

[48] Y. Yasui and X. Wang, "Statistical learning from a regression perspective by BERK, R. A.," *Biometrics*, vol. 65, no. 4, pp. 1309–1310, Dec. 2009, doi: 10.1111/j.1541-0420.2009.01343_5.x.

[49] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, "Machine learning with applications in breast cancer diagnosis and prognosis," *Designs*, vol. 2, no. 2, p. 13, May 2018, doi: 10.3390/designs2020013.

[50] H. F. Fakhruldeen and T. S. Mansour, "All-optical not gate based on nanoring silver-air plasmonic waveguide," *International Journal of Engineering & Technology*, vol. 7, no. 4, p. 2818, Oct. 2018, doi: 10.14419/ijet.v7i4.18955.

[51] S. W. Kareem and M. C. Okur, "Evaluation of Bayesian network structure learning using elephant swarm water search algorithm," in *Handbook of Research on Advancements of Swarm Intelligence Algorithms for Solving Real-World Problems*, 2020, pp. 139–159, doi: 10.4018/978-1-7998-3222-5.ch008.

[52] S. W. Kareem, "Novel swarm intelligence algorithms for structure learning of bayesian networks and a comparative evaluation," Ph.D. Thesis, Yaşar University, 2020.

[53] S. K. Sarkar, "Identifying patients at risk of breast cancer through decision trees," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 8, pp. 88–91, Aug. 2017, doi: 10.26483/ijarcs.v8i8.4602.

[54] S. Kareem and M. C. Okur, "Bayesian network structure learning using hybrid bee optimization and greedy search," in *3rd International Mediterranean Science and Engineering Congress*, 2018, no. Imsec, pp. 1–7.

[55] N. Louridi, M. Amar, and B. El Ouahidi, "Identification of cardiovascular diseases using machine learning," in *2019 7th Mediterranean Congress of Telecommunications (CMT)*, Oct. 2019, pp. 1–6, doi: 10.1109/CMT.2019.8931411.

[56] B. N. Alsunbuli, H. F. Fakhruldeen, W. Ismail, and N. M. Mahyuddin, "Hybrid beamforming with relay and dual-base stations blockage mitigation in millimetre-wave 5G communication applied in (VIOT)," *Computers and Electrical Engineering*, vol. 100, p. 107953, May 2022, doi: 10.1016/j.compeleceng.2022.107953.

[57] R. A. Al-Kaabi, H. F. Fakhruldeen, and H. A.-J. Al-Asady, "An overview of the status, challenges, and trends of the advanced crypto algorithms to enhance the security of wireless networks," in *AIP Conference Proceedings*, 2023, vol. 2591, no. 1: AIP Publishing LLC, p. 030039.

[58] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, Dec. 2016, pp. 1–4, doi: 10.1109/ICEDSA.2016.7818560.

BIOGRAPHIES OF AUTHORS

Mustafa Qahtan Alsudani 🆔 📚 SC 🔗 received the B.Eng. degree in technical computer engineering from Al-Rafidain University College, Iraq, in 2010 and the M.S. and Ph.D. degrees in computer engineering techniques from National Aerospace University "KhAI," Kharkiv, Ukraine, in 2013 and 2018, respectively. Currently, he is an Associate Professor and Head of the department of computer engineering teachings, imam Jaafar Al-Sadiq University. His research interests include computer vision, system vulnerability, system security, cyber-attacks, wireless networks, and wireless communications. He can be contacted at email: alsudani.m.q@mail.com.

Hassan Falah Fakhruldeen 🆔 📚 SC 🔗 received the B.Eng. degree in communications engineering from Al-Furat Al-Awsat Technical University, Iraq, in 2010 and the M.S. and Ph.D. degrees in electronics and communications engineering from Baghdad University, Iraq, in 2013 and 2020, respectively. Currently, he is an Associate Professor at the Department of Electrical Engineering, University of Kufa. His research interests include photonics, optics, optical fiber communications, nano-photonic devices, plasmonics devices, optical communications, optical fiber networks, plasmonic sensors, all-optical signal processing, 5G communications, communications transmission lines, signal transmission planning, wireless networks, wireless communications, and radio over fiber communications. He can be contacted at email: hassan.fakhruldeen@gmail.com.

Israa Fayez Yousif 🆔 📚 SC 🔗 received the B.Eng. degree in materials engineering from the University of Kufa -engineering college, Iraq, in 2014 and the M.S. degree in materials engineering from university of Kufa -engineering college in 2013 and 2018, respectively. She can be contacted at email: israa.fayez@sadiq.edu.iq.