QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Comparative Evaluation of Machine Learning Algorithms in Breast Cancer

Bikhtiyar Friyad Abdulrahman
Department of Technical Information Systems Engineering, College of Erbil Technical Engineering, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq
bikhtiyar.abdulrahman@epu.edu.iq

Roojwan Sc Hawezi
Department of Technical Information Systems Engineering, College of Erbil Technical Engineering, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq
roojwan.hawezi@epu.edu.iq

Saja Mohammed Noori M.R
Department of Computer Network, College of Engineering and Computer Science, Lebanese French University, Erbil, Kurdistan Region, Iraq
saja.mohammed@lfu.edu.krd

Zhwan mohammed khalid
Department of Technical Information Systems Engineering, College of Erbil Technical Engineering, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq
Shahab.kareem@epu.edu.iq

Shahab Wahhab Kareem
Department of Technical Information Systems Engineering, College of Erbil Technical Engineering, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq
zhwan.mei20@epu.edu.iq

Zanyar Rzgar Ahmed
Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Erbil, Kurdistan Region, Iraq
Zanyar.ahmed@lfu.edu.krd

ARTICLE INFO

Article History:
Received: 15/12/2021
Accepted: 24/1/2022

ABSTRACT

Breast cancer is one of the world's leading causes of mortality in women and is due to uncontrollable breast cell growth. Early detection and proper care are the only means of

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Published: Winter 2022

Keywords: *Machine Learning, Classification algorithms, Neural Network, Breast Cancer, Dataset.*

Doi: 10.25212/lfu.qzj.7.1.34

avoiding deaths from breast cancer. The precise characterization of tumors is a critical task in the medical profession. Because of their high precision and accuracy, machine learning methods are commonly used in identifying and classifying various forms of cancer. In this review article, the authors have tested different machine learning algorithms and implemented them, which can be used by doctors to identify cancer cells in an early and accurate way. This article introduces several algorithms, including support vector machine (SVM), Nave Bayes Classifier (NBC), artificial neural network (ANN), Random Forest (RF), decision tree (DT), and k-Nearest-Neighbor (KNN). These algorithms are trained with a collection of data containing tumor parameters for a person with breast cancer. After comparing the results, we found the highest accuracy of the Support Vector Machine and Random Forest and the highest precision of the Naive Bayes Classifier (NBC). In addition, we review the number of researches that provide machine learning algorithms for detecting breast cancer.

1. Introduction

The machine learning sector consists of several computational, probabilistic, and optimization strategies to "read" machines from past examples and detect hard-to-data models from large, noisy, or complicated datasets. These characteristics, depending on complex proteomic and genomic measures, are particularly suitable for medical use (S. Kharya and S. Soni, 2016) (S. H. Ismael, S. W. Kareem, and F. H. Almukhtar, 2020). Machine learning approaches are used to diagnose cancer and identify cancer like the support vector network, the Bayesian confidence network, and the artificial neural network (P. Suryachandra and P. V. S. Reddy, 2016) (S. Kareem and M. C. Okur, 2017). Computer training has most recently extended to the diagnosis and forecast of cancer. The survey has shown that certain algorithms are best used to test dataset functionality (S. W. Kareem and M. C. Okur, 2019).

To retrieve useful data bits from individual data to aid decision-making, a variety of essential data extraction technologies are being improved and deployed in a variety of real-world applications (e.g., healthcare, bioscience, and industry) (N. Mallios, E. Papageorgiou, and M. Samarinas, 2011), in machine learning environments has A reasonably wide array of data comprising of actual medical cases of men diagnosed with prostate cancer that receive medical attention is used for the systematic comparison of procedure in this research (Syah, et al., 2021). Machine learning methods are software programs that forecast something (conduct, the form of disease, the picture of stock price volatility, etc.), based on the circumstances that triggered what happened in the past (Karplus, 2012) (A. S. Mohammed, S. W. Kareem, A. Al Azzawi, and M. Sivaram, 2018).

In women worldwide, breast cancer is the most prevalent cancer. Early development of certain cells in the breast causes breast cancer. Several approaches have been developed for breast cancer detection. Breast imaging, also known as mammography, is a form of mammography (M. Mori, S. Akashi-Tanaka, S. Suzuki, M. I. Daniels, C. Watanabe, M. Hirose, et al, 2017). Is a breast cancer diagnostic procedure. The x-rays monitor the status of female nipples. It is nearly impossible to diagnose breast cancer early in the detectable external cancer cell. At an early stage, cancer is detected by mammography, only takes a few minutes. Dynamic Magnetic Resonance Imaging (MRI) (T. Nagashima, M. Suzuki, H. Yagata, H. Hashimoto, T. Shishikura, N. Imanaka, et al, 2002)has established the breast distortion detection procedure. The modality predicts the rate at which tumor angiogenesis grows. In persons with breast cancer, magnetic reasoning imaging leads to a reduction in contrast metastases. Ultrasound (Mohammed, 2021)is a well-known monitoring method to identify the symptoms within a sound wave via the body. A sound-emitting transducer is mounted on the skin, and the sound waves capture the reflections of the body's tissues (Kareem S. W., 2009). Electrography (C. S. Park, S. H. Kim, N. Y. Jung, J. J. Choi, B. J. Kang, and H. S. Jung, 2015) is a technology based on imagery that was recently developed. This technique applies when the tissues of breast cancer are larger than the adjacent regular parenchyma. A color map of sample compression distinguishes benign and malignant types. The echoes are transformed to grayscale or a numerical value that can be seen on a computer (Alyousuf, Din, & Qasim, 2020). Positron emission tomography photographs of Ffluorodeoxyglucose (PET) (A. T. Azar and S. A. El-Said, 2013) Enable doctors to examine the tumor's role in the human body, relying on the recovery of radiation symbol tracers (Miran & Kadir, 2019).

In recent years many medical predictions have used different machine learning (M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, vol. 1,2020.) (M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza, and M. K. Hasan, 2018), bio-inspired computation techniques (Kh., T. & Hamarash, I, 2022) (Ashish Sharma, 2022) (Al-Alkawi, H. J. M., Hanfesh, A. O., & Mohammed Rauof, S. M. N., 2019) (Qasim, A.J., Din, R.E., Alyousuf, F.Q.A, 2020) (M. K. Hasan, M. M. Islam, and M. Hashem, 2016) (M.V. Prakash, V. Porkodi, S. Rajanarayanan, M. Khan, B.F. Ibrahim, M. Sivaram, 2020) (W. Kareem, R. Z. Yousif, and S. M. J. Abdalwahid, 2020), and deep learning (S. I. Ayon and M. Islam , 2019).

While a variety of methods have been proven, none of them can provide an accurate and stable result. The doctors must read a high amount of imagery data in mammography, which limits accuracy. It takes time, even and diagnoses the disease with the wrong result in some worst cases. This article contrasts several strategies of machine learning for the detection of disease from the data. To diagnose the disease with proper results, six supervised machine-learning methods were used.

This article is arranged as follows. Section one introduction, and Section two reviews the previous literature. Section three, preprocess of machine learning. Software and simulation are presented and present the discussions of the results in section four. Section five Conclusions are presented.

2. Related work

In (S. Chaurasia, P. Chakrabarti, and N. Chourasia, vol. 59, 2012.) It was done with three classification strategies on the breast cancer dataset: Decision Tree, Bayes, and Neural Net. The experiment concludes that the Neural Net performance classification of breast cancers with greater sensitivity and precision is better than the Decision Tree and Naive Bayes classification.

In (C. Nguyen, Y. Wang, and H. N. Nguyen, 2013) A machine has been designed to assist in the distinction between malignant and benign tumors. To choose

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

functionality, the Backward Elimination (BE) approach was paired with the Random Forest Tree. The dataset was obtained from Wisconsin's predictive network. There are 33 variables, the accuracy of this hybridized algorithm is approximately 95%, and it has been reduced or 17 to 18 variables.

In (L. G. Ahmad, A. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. Razavi, 2013)It is used to test 3 related algorithms such as SVM, (ANN), and Decision Tree (DT). The database used was extracted from the Iranian Center in this analysis (ICBC). A total of 8 variables of the predictor was used (SVM) algorithm provided up to 95%.

(Hota, vol. 3, 2014.)Created the Support Vector Machine gathering with a decision tree (C5.0) model for detecting breast cancer. The dataset was generated by combining 32 elements of the Wisconsin prognostic dataset. The variable reduction was accomplished by the use of Rank-based role selection. The performance of the radial base function is 92.59 percent for 5 features.

In (K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, 2015)discussed numerous statistical models of recent approaches to machine learning used to detect cancer advancement. In this work, the author has reviewed many ML-relevant publications. Dependent on the dataset and its variables, each category and classification of the papers varies. These documents often consist of mammograms of up to 14 variables, with the precision of mammographic data as high as 83 percent, and other datasets as high as 71 percent.

In (M. Rana, P. Chandorkar, A. Dsouza, and N. Kazi, 2015) Comparative experiments included many machine learning approaches for estimation and diagnosis of breast cancer by these approaches, including SVM, Logistical Regression, Naïve Bayes, and the KNN. For the analysis, data collection is used in the Wisconsin prognostic breast cancer data repository with 32 variables, and 95.6% and 68% of the breast cancer recurrence and non-recurrence data were used.

In (E. Venkatesan and T. Velmurugan, 2015) Performed various classification algorithms, including Director tree (AD), decision tree (j48) algorithm, and Best First Tree (B+ tree). From the diagnostic center of Swami Vivekananda in Chennai dataset was taken. It contains 220 medical data and is used for the review of nine attributes. The outcome shows 99 percent out of 4 algorithms j48.

In (H. T. T. Thein and K. M. M. Tun, 2015) proposed a breast cancer solution that distinguishes between different types of breast cancer. The method focuses on breast cancer diagnosis and estimate in Wisconsin, as well as the identification of several forms of breast cancer. The suggested approach leverages and analyzes two separate migratory topologies in Iceland for a more precise and time-saving training procedure.

In (K. Sivakami and N. Saraswathi, 2015.) presented a forecast of disease status using a hybrid approach for forecasting improvements and their effects which are key to lethal infections. Their approach consists of two major components to alert the seriousness of diseases: 1. Treatment and Extraction of Options for Information, 2. Tree-Support Hybrid Model for Predictions (DT-SVM) Decision. To create precise predictive models using data mining techniques in breast cancer, they analyzed data from UCI machine learning datasets from Wisconsin. Three classification technologies in the Weka program are comparable findings indicate that the DT-SVM is more predictive Naïve based classification, and Sequential Minimal Optimization are all superior.

In (B. Gayathri and C. Sumathi, 2015) Fuzzy logic was used to detect the presence of breast cancer. For this study, data from the UCI learning repositories is collected. The aim is to find breast cancer by reducing the causes of the disease and shortening the diagnosis time. The LDA technique was used to choose the feature, and the Fuzzy Mamdani method was used to teach it, The LDA process. A fuzzy deduction is an inference tool. The fuzzy logic was used to test the results. 93 percent of the findings were made public.

(A. Bharathi and K. Anandakumar, 2015) created an effective cancer classification machine learning method that could increase the accuracy of cancer classification. The work is divided into two stages. The first step is to use the Ariance Analysis (ANOVA) rating scheme to choose the important genes. The classification task in the second stage necessitates the use of an appropriate classifier. Relevance Vector Machine (RVM) Learning and Fuzzy Support Vector Machine (FSVM) Learning were two of the most effective machine learning classifiers used. Three origin points, Lymphoma, Leukemia, and SRBCT data sets, have been defined to compute the testing values.

In (S. Kharya and S. Soni, 2016) The Naive Bayes Classifier investigated the machine learning method's performance criterion using a new weighted approach to breast cancer classification. Weighted concepts are implemented to expand standard Naive Bayes and enhance their performance. Exploration of UCI's machine learning library breast cancer dataset-based domain awareness weight assignment. The tests demonstrate that an approach to weighty naive berries is superior to naive.

In (Y. Tan, K. Sim, and F. Ting, 2017)a CNN-based technique was created to assist doctors in diagnosing defects in a quick diagnostic procedure. The classifier developed a model to detect the tumor in cancer by using improved mammogram images. Percent precision and a short diagnostic time we're given in the proposed procedure.

In (R. Alyami, J. Alhajjaj, B. Alnajrani, I. Elaalami, A. Alqahtani, N. Aldhafferi, et al, 2017) SVM and ANN are proposed templates for combined feature series. The classification tasks were carried out using a separate mix of feature subsets by deciding the best parameters and partitioning the results. The SVM demonstrated improved classification of samples with a precision of 97,1388% as opposed to the accuracy of 96,7096% obtained by ANN.

In (I. Routray and N. P. Rath, 2018) A proposal to use a Law's Texture Energy Measurement (ITEM) tool to diagnose breast cancer. The backpropagation Artificial Neural Network (BPANN)uses a technique for classifying malignant, normal, and natural tissue sections. Usual irregular grouping, 90.9 percent is responsive to the proposed procedure, with an accuracy of 94.4 percent. The accuracy for benign-malignant designation is 91.7 percent and 66.66 percent, respectively.

In (H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, 2018) SVM Weighted Area Unless the recipient's operating curve includes a variant of the breast cancer learning ensemble (AUC). Six kernel functions C-SVM and v-SVM can be added to the base model package. It was shown that the proposed model significantly improves the diagnosis of breast cancer. At 97.68 percent, the model was correct.

In (S. B. Sakri, N. B. A. Rashid, and Z. M. Zain, 2018) Five different WBCD data processing phase-based approaches were used. They published a paper relating classification to classification without the use of a feature selection scheme. The NB, RepTree, and K-NNs all produce 70%, 76.33%, and 66.33% of the time, respectively. they use the Weka platform to process their results. With the implementation of PSO, the four best features for this classification role were identified. The precision rate for NB, RepTree, and K-NN with PSO was 81.3 percent, 80 percent, and 75 percent, respectively.
In (E. Alickovic and A. Subasi, 2019) To identify breast cancer with extreme accuracy, researchers created a model employing a Normalized Multilayer Perceptron Neural Network. The findings obtained are excellent (accuracy is 99.27 percent). In comparison to past studies in which Artificial Neural Networks were employed, this is a highly encouraging outcome. Breast Cancer Wisconsin (Original) was utilized as a control test.

In (K. Vijayakumar, V. J. Kadam, and S. K. Sharma, 2021) 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. The study serves two functions. To begin, this research is a start toward a more in-depth knowledge of DNN with layer-wise distinct AFs. Second, research is being conducted to investigate better DNN-based systems for developing prediction models for breast cancer data with increased accuracy. As a result, the framework was validated using the benchmark UCI dataset WDBC. Multiple simulations and testing results demonstrate that the suggested method outperforms the Sigmoid, ReLU, LeakyReLU, and Swish activation DNN in terms of various parameters.
In (J. Wu and C. Hicks, 2021) As create the models for categorizing the two forms of breast cancer, researchers used four distinct classification models, including SVM, KNN, Nave Bayes, and DT, with characteristics picked at varying threshold levels. The suggested approaches were applied to separate gene expression datasets for performance evaluation and validation. Results: The Support Vector Machine

method was able to classify breast cancer more correctly into triple-negative and non-triple negative breast cancer and had fewer misclassification errors than the other three algorithms assessed.

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 (S. Chaurasia, P. Chakrabarti, and N. Chourasia, vol. 59, 2012.)	9
Random forest tree	99%	Random forest classification along with breast cancer detection and prognostic function collection. (C. Nguyen, Y. Wang, and H. N. Nguyen, 2013)	17
Neural Network SVM and DT	95%	Three machine learning models for predicting breast cancer recurrence (L. G. Ahmad, A. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. Razavi, 2013)	8
SVM	92.59%	Breast cancer identification using decision tree and support vector ensemble with reduced function subset (Hota, vol. 3, 2014.)	5
various predictive model	83%	Applications for machine learning for cancer prognosis (K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, 2015)	14
(SVM), Logistic Regression, Naïve Bayes and (KNN)	95.6%	ML models for identifying breast cancer and forecasting recurrence (M. Rana, P. Chandorkar, A. Dsouza, and N. Kazi, 2015)	----
j48, Best First Tree (B+ tree), and AD tree	99%	Performance analyzes of breast cancer decision tree algorithms (E. Venkatesan and T. Velmurugan, 2015)	4
ISLAND APPROACH TO NEUROL NETWORK	99.97%	A neural network approach to characterization and breast cancer diagnosis. (H. T. T. Thein and K. M. M. Tun, 2015)	9

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

DIFFERENTIAL			
Support Vector Machine and Decision tree	95%	Big-data extraction: DT-SVM hybrid platform for breast cancer prediction (K. Sivakami and N. Saraswathi, 2015.)	9
Fuzzy inference system	93%	Breast cancer risk identification Mamdani fuzzy inference method (B. Gayathri and C. Sumathi, 2015)	4
Support Vector Machine	97%	A significance vector machine learning technique is used to classify cancer. (A. Bharathi and K. Anandakumar, 2015)	4
Naïve Bayes	92%	Bayes Weighted Classification: a Breast Cancer Detection Predictive Plan (S. Kharya and S. Soni, 2016)	9
CNN	82.73%	In mammogram imaging applications, convolutional neural networks are used to detect breast cancer. (Y. Tan, K. Sim, and F. Ting, 2017)	
SVM and ANN	97.14%	An examination of the effect of artificial neural networks and vector-supporting devices on the diagnosis of breast cancer (R. Alyami, J. Alhajjaj, B. Alnajrani, I. Elaalami, A. Alqahtani, N. Aldhafferi, et al, 2017)	subset
backpropagation Artificial Neural Network	91.7%	Textural feature-based mammogram classification using ANN (I. Routray and N. P. Rath, 2018)	
c-SVM and v-SVM	97.68%	A vector-based ensemble algorithm supports the detection of breast cancer (H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, 2018)	
RepTree Naive Bayes k-NNs	81.3% 80% 75%	prediction Particle swarm Selection for breast cancer recurrence Optimization. (S. B. Sakri, N. B. A. Rashid, and Z. M. Zain, 2018)	
Normalized Multilayer Perceptron Neural Network	99%	Created a model employing a Normalized Multilayer Perceptron Neural Network. The findings obtained are excellent (E. Alickovic and A. Subasi, 2019)	14

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

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. (K. Vijayakumar, V. J. Kadam, and S. K. Sharma, 2021)	4
SVM KNN Naïve Bayes DT	90% 87% 85% 87%	four distinct classification models, including SVM, KNN, Naïve Bayes, and DT, with characteristics picked at varying threshold levels. The suggested approaches were applied to separate gene expression datasets for performance evaluation and validation (J. Wu and C. Hicks, 2021)	

3. Preprocess of Machine Learning

The learning process can be divided into two groups with ML methods: regulated and unattended. Various data instances are used and labeled to provide optimum performance to train the system for unmonitored instruction. However, no predetermined information sets are available in schooling, and thus the goal is difficult to achieve. The findings are not foreseen. Classification is one of the most popular types of controlled education. It uses historical data to create a standard for future forecasts. It uses historical data. Clinics and hospitals maintain vast databases in the field of medicine that include patient records and diagnoses of their symptoms. Researchers then use this experience to construct classification models based on historical events. Thus, medical inference with computer help, using the pure amount of medical data available today, has become a much easier process. It is worth noting that all methods used in this paper are classified as models (P. Misra and A. S. Yadav, 2019) .

3.1 Neural Network

The shortened sight of an artificial neural network is a knowledge-based computational algorithm. The meaning and function of ANN are the same as the human brain is conceived. Raw data show and observe associations and general patterns [8]. The network has connections to its weight and nodes. The neural network has four hidden layers: center, input, and output. Each of these layers is linked to the neural network by a connector known as the weight connector (M.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Sardar, K. Al-Jumur, and S. W. Kareem, Raghad z. yousif, 2021). A multilayered vision, which uses the back distribution, is the neural network in this paper. There are three different layers in the back spreading network (input, secret, and exit), where a signal crosses one path such that the signal does not revert to the source after transferring the neuronal output from the input neuron. Figure depicts the background neural network 1.

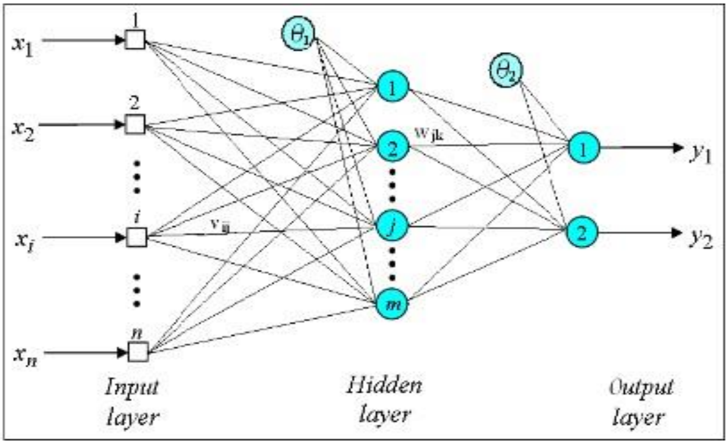

Fig 1: Back propagation neural network

3.2 Random Forest (Rf)

The back spreading network has three layers (input, secret, and exit), each of which allows a signal to travel one path without reverting to the source after transmitting the neuronal output from the input neuron. RF does not supply a single decision tree, but rather a set of trees that together produce a fundamental or distinctive pattern (Kononenko, 2001). RF has greater reliability compared to each decision-making tree. This shows that the noise in the input data does not affect RF. For cancer diagnosis, RF has the property of managing minority results. It also makes deviations possible, detects errors, and prevents outliers. It contains detailed guidelines for the cultivation and combination of trees, post-processing, and self-testing. The algorithm of RF is recursive; a sample is randomly selected and substituted from the data set whose dimensions are labeled with N; a new sample is randomly selected and not replaced from the predictors. Any time the data obtained is separated, the preceding process is done. The information from the

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

previous steps will then be removed from the bag and the prior steps will be repeated according to the number of trees needed. Finally, an account is taken of the number of trees in these classes. Following the remainder of the Treaties, the identification then takes place (Y. Yasui and X. Wang, 2009). Figure 2 shows an example of Random forest RF as a graph.

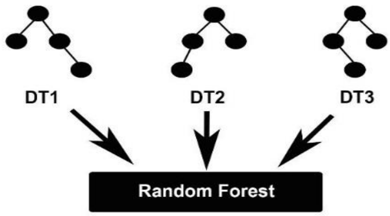

Fig 2: Graph for the algorithm of the Random Forest

3.3 Decision Tree (Dt)

It is a statistical model used to solve problems of classification and regression. The decision tree imitates human logic and makes knowledge easier to understand and analyze. It is a tree-like classifier that matches any possible data results into groups (W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, 2018). Each derived class subset is repetitively sent to the same process and again referred to as recursive partitioning. This strategy generally stops if a subset node has the same value as the goal attributes or cannot be added to the forecast. The root node is the tree's highest node. A feature, each decision reference, and each output leaf are included in every DT node. It can process digital and categorical data. Variety methods of DT are available, including C4.5 and Classification and Retrieval Tree (CART) (ID3). A decision tree example used to determine whether or not a person has cancer in order to operate on a Breast Cancer Diagnosis. The procedure of the DT is an example shown in figure 3.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

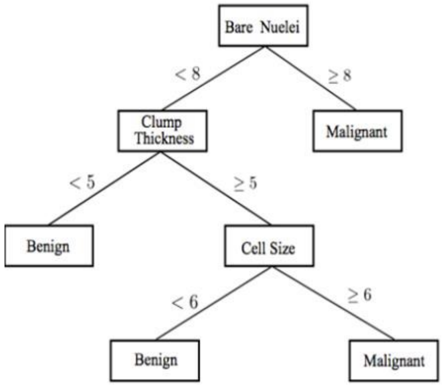

Fig 3: Graph for the algorithm of the decision tree

3.4 SUPPORT VECTOR MACHINE (SVM)

The supervised master learning approach divides the data set into groups with an adequate hyperplane margin. This method is commonly used for the diagnosis of the disease in the field of medicine. Because the data set can contain several hyper lines, the SVM algorithm maximizes the limit, thereby trying to establish a maximal difference between different groups. Observe the darkness of groups in this dataset. The Wisconsin breast cancer dataset cannot be divided into ideal groups on a single line. The darkness is seen in figure 4.

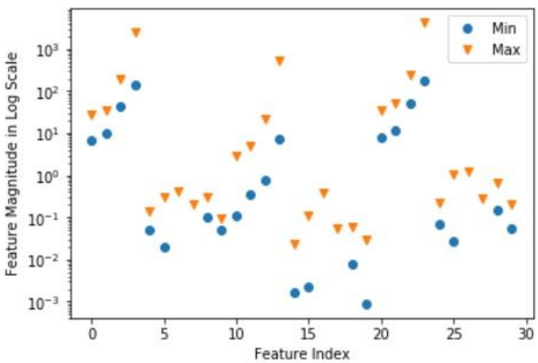

Fig 4: Wisconsin Dataset class distribution.

Using transformation to solve this difficulty and add another dimension as Z-axis. Now, the strong difference between groups is easily apparent when a dataset is plotted on the Z-axis. The process is performed utilizing kernels. Polynomial and exponential kernels measure a higher dimensional separation line. Figure 5 shows why kernels have played an important role in determining the collection of dark data.

Since the precision of SVM was not suited for 95, 1%. When further processing, precision was ultimately increased to 97.3 percent by the value of the regulation parameter (C)

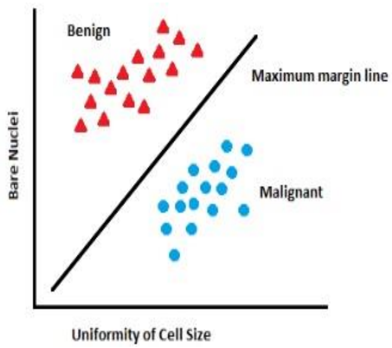

Fig 5: Using the kernel for classification

3.5 NAIVE BAYES CLASSIFIER

Is a fundamental tool for adding classes to input instances? By specifying the framework for modeling a decision the classification of Naïve Bayes can be defined. The variables used are dependent on the conditions in this classification (F. Cugnata, R. S. Kenett, and S. Salini, 2016) (KAREEM, 2020) (S. W. Kareem and M. C. Okur, 2020). Bayesian classifiers use the theorem of Bayes and are particularly favored if the input dimension is strong[5]. Bayes' theorem is given as The mathematical equation

$$P(y/X) = \frac{P(X/y)\,P(y)}{P(X)} \qquad (1)$$

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

P(y) is the per-case likelihood, i.e. the pre-evidence likelihood of the event, P(X/|y) is the predictor probability, P(y/X) is the late class probability, i.e. the probability of the occident after the proof, and P(X) is the pre-evidence probability predictor. As defined earlier, X indicates a function vector of the dependent size. (S. K. Sarkar and A. Nag, 2017) (S. Kareem and M. C. Okur, 2018). Equation 2 can now be defined as:

$$P\,(y|x_1 \ldots x_n) = \frac{P(x_1|y)P(x_2|y)\ldots P(x_n|y)P(y)}{P(x_1)P(x_2)\ldots P(x_n)} \qquad (2)$$

and can be explained further:

$$P\,(y|x_1 \ldots x_n) = \frac{P(y)\prod_i^n=1\,P(x_i|y)}{P(x_1)P(x_2)\ldots P(x_n)} \qquad (3)$$

The denominator can be seen for a given input in the above equation as constant. Once the calculation is simplified, the numbers can be deducted directly relative to each other.

Multinomial, Gaussian, and Bernoulli are the three different grouping groups of Naive Bayes. The indicator of a data set with a continuous value is the Gaussian Naive Bayes classification, for which the Wisconsin Breast Cancer dataset can be used for precision measurements.

3.6 K Nearest Neighborhood (KNN)

KNN is one of the most central classification techniques in machine language. This classification algorithm is a parametric, lazy algorithm that does not take into account the underlying data. For both classification and regression, KNN may be used. Each evaluation data point has the closest K training data point in classification. Both training information points are grouped into classes, and the evaluation data is allocated to the most common class. K is thus used to describe a set of nearby training data points. Below is an example of a KNN breast cancer screening algorithm.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

The KNN where K=6 is located is shown in Figure 6. The blue disk is in the center. Blue is the nucleus of a blue sample, the malignancy is a rosy rectangle, and the neutral is a gray triangle.

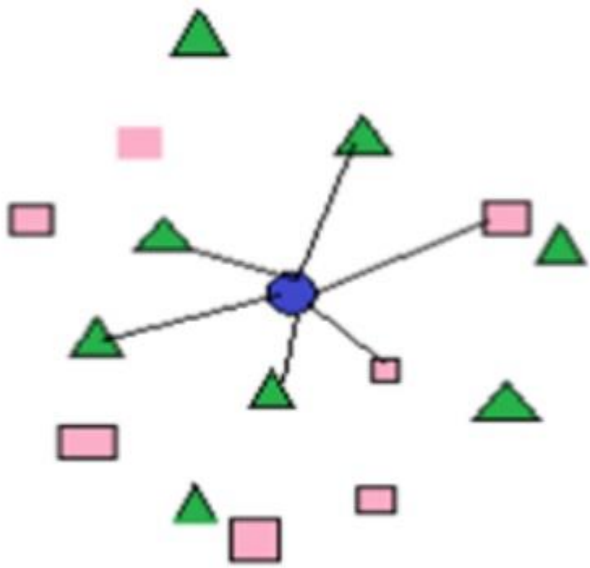

Fig 6: KNN algorithm for classification.

4. Performance Evaluation

This paper used the training data collection obtained from the Breast Cancer Data of Wisconsin. This data set can be found in the UCI Machine Learning repository (N. Louridi, M. Amar, and B. El Ouahidi, 2019). Its data collection has more than 569 instances and is multivariate. a digital image cell nuclei are categorized as malignant or benign based on more than ten characteristics (R. S. Ismael, R. S. Youail, and S. W. Kareem, 2014). The ten characteristics are as follows: (Area, Compactness: (p*p/a-1) where p is the perimeter and area, Concave points: the number of concave contour sections, A coastline approximation is a kind of fractal. Concavity is the tip of a concave contour component., Perimeter, Radius: mean of d, where d is a center-perimeter radius, Smoothness: local differences in radius length, Symmetry, and

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Texture: grayscale S.D where the standard deviation is standard S.D.). The Anaconda program was included in this paper as a method for learning machines. Anaconda is an open-source tool based on pythons, published under the New BSD License for the first time in 2012. It contains various algorithms and techniques for machine learning including the algorithms studied in this article. The Python and Programing languages for scientific computation are distributed free and open-source. In addition, this software provides data science, large-scale data handling, and predictive analytics. After comparing the outcomes, the parameters and displays the results of six different machine learning models. The authors found that the highest accuracy of Support Vector Machine (SVM) and Random Forest was 97.3%, the best precision of Naive Bayes, and a 97.3% & 97 % recall. In addition, we review the number of researches that provide machine learning algorithms for detecting breast cancer. The identification rate is also known as accuracy. The accuracy metric is defined by the number of instances correctly detected divided by total of instances in the data set. The precision for various sets may be changed and depends very much on the threshold used in the classification (Kareem S. O., 2021). The accuracy can be determined as:

$$\text{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{T+P}} \qquad (4)$$

Where TN is the True Negative, TP is the True Positive. The shape entire of P is also positive, indicating cancer cells, while N represents negatives and the non-cancer cells that are benign. Precision is often referred to as confidence. The rates of True Positive instances and True Positive instances are used to define precision. Precision reveals the classifier's capacity to cope with positive situations; it does little of negative cases. Precision and recall have a mutually proportional relation (Raghad Z. Yousif1, Shahab W. Kareem, Shadan M. Abdalwahid, 2020). The equation can be used to determine this parameter:

$$\text{Precision}= \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (5)$$

The recall is represented as very optimistic and false-negative instances. This measurement is used in the medical field as it gives details on the proper classification of the number of malignant and benign cases. The model will find all relevant cases in the dataset. With the equation, the recall can be calculated.

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (6)$$

The following table shows the analysis and contrast of recall, precise use of the Wisconsin dataset for six machine learning techniques:

Table 2: Comparison of Wisconsin breast cancer classification algorithms

Method	Accuracy	Precision	Recall
Support Vector machine	97.3 %	97.2%	96.9%
NAIVE BAYES	95.7%	97.3%	97%
NEURAL NETWORK	95.7%	97.2%	96.8%
Random forest	97.3%	97.2%	96.9%
K Nearest Neighborhood	93.8%	96.2%	96.2%
Decision tree	93%	94%	93.3%

5. Conclusion

The most common type of cancer worldwide is breast cancer. A randomly chosen woman has a 12% risk of being diagnosed. Early diagnosis of breast cancer will save a lot of life as well. This article presents six machine-learning techniques for breast cancer prediction: vector support equipment, Nave Bayes Classifier, artificial neural networks, K-neighbors, random forests, and DT. All these six ML techniques were compared to be accurate, recall, and precise. The effectiveness of the above algorithms has also been compared using the Wisconsin dataset. We found that the best accuracy of Support Vector Machine (SVM) and Random Forest was 97.3%, the best precision of Naive Bayes, and a 97.3% & 97 % recall

References

A. Bharathi and K. Anandakumar. (2015). Cancer classification using relevance vector machine learning approach. *Journal of medical imaging and health informatics, vol. 5*, 630-634.

A. S. Mohammed, S. W. Kareem, A. Al Azzawi, and M. Sivaram. (2018). Time series prediction using SRE-NAR and SRE-ADALINE. *Journal of Advanced Research in Dynamical and Control Systems* , 1716-1726.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

A. T. Azar and S. A. El-Said. (2013). Probabilistic neural network for breast cancer classification. *Neural Computing and Applications, vol. 23*, 1737-1751.

Al-Alkawi, H. J. M., Hanfesh, A. O., & Mohammed Rauof, S. M. N. (2019). Development of High Performance (Mechanical and Wear Properties) of AA 6061-Hybrid Nano Composites Via Liquid Metallurgy Route. *Al-Nahrain Journal for Engineering Sciences, 22(2), 143*.

Alyousuf, F. Q. A., Din, R., & Qasim, A. J. (2020). Analysis review on spatial and transform domain technique in digital steganography. *Bulletin of Electrical Engineering and Informatics*, *9*(2), 573-581.

Ashish Sharma, A. M.-1.-6. (2022). The 3D Facemask Recognition: Minimization for Spreading COVID-19 and Enhance Security, . *ICT Analysis and Applications, Springer Singapore 2022 P 619-637* (pp. 619-637). Singapore : Springer .

B. Gayathri and C. Sumathi. (2015). Mamdani fuzzy inference system for breast cancer risk detection . *B. Gayathri and C. Sumathi, "Mamdani fuzzy inference system for bre IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1-6). IEEE.

C. Nguyen, Y. Wang, and H. N. Nguyen. (2013). Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic . *Journal of Biomedical Science and Engineering*, 551-560.

C. S. Park, S. H. Kim, N. Y. Jung, J. J. Choi, B. J. Kang, and H. S. Jung. (2015). Interobserver variability of ultrasound elastography and the ultrasound BI-RADS lexicon of breast lesions . *Breast Cancer, vol. 22*, 153-160.

D. Bazazeh and R. Shubair. (2016). Comparative study of machine learning algorithms for breast cancer detection and diagnosis. *D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detec5th international conference on electronic devices, systems and applications (ICEDSA)* , (pp. 1-4).

E. Alickovic and A. Subasi. (2019). Normalized neural networks for breast cancer classification. *International conference on medical and biological engineering*, (pp. 519-524).

E. Venkatesan and T. Velmurugan. (2015). Performance analysis of decision tree algorithms for breast cancer classification. *Indian Journal of Science and Technology, vol. 8*, 1-8.

F. Cugnata, R. S. Kenett, and S. Salini. (2016). Bayesian networks in survey data: Robustness and sensitivity issues. *Journal of Quality Technology, vol. 48*, 253-264.

H. T. T. Thein and K. M. M. Tun. (2015). An approach for breast cancer diagnosis classification using neural network. *Advanced Computing, vol. 6*, 1-8.

H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research, vol. 267*, 687-699.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Hota, H. (vol. 3, 2014.). Identification of breast cancer using ensemble of support vector machine and decision tree with reduced feature subset. *H. Hota, "Identification of breast cancer using ensemble of support vector machine and decision tree International Journal of Innovative Technology and Exploring Engineering* , 99-102.

I. Routray and N. P. Rath. (2018). Textural feature based classification of mammogram images using ANN. *9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, (pp. 1-6).

J. Wu and C. Hicks. (2021). Breast Cancer Type Classification Using Machine Learning. *Journal of personalized medicine, vol. 11*, 61-70.

K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis. (2015). K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. KaramMachine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal, vol. 13* , 8-17.

K. Sivakami and N. Saraswathi. (2015.). Mining big data: breast cancer prediction using DT-SVM hybrid model. *K. Sivakami and N. Saraswathi, "Mining big data: International Journal of Scientific Engineering and Applied Science (IJSEAS), vol. 1*, K. Sivakami and N. Saraswathi, "Mining big data: breast cancer prediction using DT-SVM418-429.

K. Vijayakumar, V. J. Kadam, and S. K. Sharma. (2021). Breast cancer diagnosis using multiple activation deep neural network. *Concurrent Engineering, vol. 29*, 275-284.

Kareem, S. O. (2021). Falcon Optimization Algorithm For Bayesian Network Structure Learning. *Computer Science Journal, Vol 22, No.4.* , 553-569.

Kareem, S. W. (2009). *Hybrid Public Key Encryption Algorithms For E-Commerce.* Erbil: University of Salahaddin–Hawler.

KAREEM, S. W. (2020). *NOVEL SWARM INTELLIGENCE ALGORITHMS FOR STRUCTURE LEARNING OF BAYESIAN NETWORKS AND A COMPARATIVE EVALUATION.* IZMIR: YAŞAR UNIVERSITY.

Karplus, A. (2012). Machine learning algorithms for cancer diagnosis . *Santa Cruz County Science Fair*.

Kh., T. & Hamarash, I. (2022). Model-Based Quality Assessment of Internet of Things Software Applications: A Systematic Mapping Study. *International Association of Online Engineering. Retrieved March*.

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine, vol. 23*, 89-109.

L. G. Ahmad, A. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. Razavi. (2013). L. G. Ahmad, A. Eshlaghy, A. PoorebrahUsing three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform, vol. 4, p. 3, 2013.*, 3.

M. K. Hasan, M. M. Islam, and M. Hashem. (2016). Mathematical model development to detect breast cancer using multigene genetic programming. *5th International Conference on Informatics, Electronics and Vision (ICIEV)*, (pp. 574-579).

M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir. (vol. 1,2020.). Breast cancer prediction: a comparative study using machine learning techniques. *SN Computer Science, vol. 1*, 1-14.

M. Mori, S. Akashi-Tanaka, S. Suzuki, M. I. Daniels, C. Watanabe, M. Hirose, et al. (2017). M. Mori, S. Akashi-Tanaka, S. SDiagnostic accuracy of contrast-enhanced spectral mammography in comparison to conventional full-field digital mammography in a population of women with dense breasts," . *reast Cancer, vol. 24* , 104-110 .

M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza, and M. K. Hasan. (2018). Performance evaluation of random forests and artificial neural networks for the classification of liver disorder. *international conference on computer, communication, chemical, material and electronic engineering (IC4ME2)*, (pp. 1-5).

M. Rana, P. Chandorkar, A. Dsouza, and N. Kazi. (2015). Breast cancer diagnosis and recurrence prediction using machine learning techniques. *IJRET: International Journal of Research in Engineering and Technology*, 2319-1163.

M. Sardar, K. Al-Jumur, and S. W. Kareem, Raghad z. yousif. (2021). Predicting temperature of Erbil city applying deep learning and neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 944-952.

Miran, A., & Kadir, G. (2019c). Enhancing AODV routing protocol to support QoS. *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE)*, *8*(5), 1824–1830.

M.V. Prakash, V. Porkodi, S. Rajanarayanan, M. Khan, B.F. Ibrahim, M. Sivaram. (2020). Improved conservation of energy in fog IoT services using machine learning model. *International Conference on Computing and Information Technology (ICCIT-1441)*, (pp. 1-4).

Mohammed, A. G. (2021). A Study of Scheduling Algorithms in LTE -Advanced HetNet. *QZJ, vol. 6, no. 3,* , 945–968

N. Louridi, M. Amar, and B. El Ouahidi. (2019). Identification of Cardiovascular Diseases Using Machine Learning. *7th Mediterranean Congress of Telecommunications (CMT)* (pp. 1-6). : CMT.

N. Mallios, E. Papageorgiou, and M. Samarinas. (2011). N. Mallios, E. PapageComparison of machine learning techniques using the WEKA environment for prostate cancer therapy plan. *IEEE 20th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enerprises* (pp. 151-155). IEEE .

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

P. Misra and A. S. Yadav. (2019). Impact of Preprocessing Methods on Healthcare Predictions. *2nd International Conference on Advanced Computing and Software Engineering (ICACSE).*

P. Suryachandra and P. V. S. Reddy. (2016). Comparison of machine learning algorithms for breast cance. *nternational Conference on Inventive Computation Technologies (ICICT)*, (pp. 1-6).

Qasim, A.J., Din, R.E., Alyousuf, F.Q.A. (2020). Review on techniques and file formats of image compression. *Bull. Electr. Eng. Inf. 9(2)*, 602–610.

R. Alyami, J. Alhajjaj, B. Alnajrani, I. Elaalami, A. Alqahtani, N. Aldhafferi, et al. (2017). R. Alyami, J. Alhajjaj, B. Alnajrani, I. Elaalami, A. Alqahtani, N. Aldhafferi, et al., "Investigating the effect of correlation based feature selection on breast cancer diagnosis using artificial neural network and support vector machines. *International Conference on Informatics, Health & Technology (ICIHT)*, (pp. 1-7).

R. S. Ismael, R. S. Youail, and S. W. Kareem. (2014). Image encryption by using RC4 algorithm . *European Academic Research, II, vol. 4* , 5833-5839 .

Raghad Z. Yousif1, Shahab W. Kareem, Shadan M. Abdalwahid. (2020). Enhancing Approach for Information Security in Hadoop. *Polytechnic Journal. 2020., 10*(1), 808.

Syah, R., Davarpanah, A., Nasution, M. K., Wali, Q., Ramdan, D., Albaqami, M. D., ... & Noori, S. M. (2021). The Effect of Structural Phase Transitions on Electronic and Optical Properties of CsPbI3 Pure Inorganic Perovskites. *Coatings*, *11*(10), 1173.

S. B. Sakri, N. B. A. Rashid, and Z. M. Zain. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access, vol. 6*, 29637-29647.

S. Chaurasia, P. Chakrabarti, and N. Chourasia. (vol. 59, 2012.). An application of classification techniques on breast cancer prognosis. *International Journal of Computer Applications* .

S. H. Ismael, S. W. Kareem, and F. H. Almukhtar. (2020). Medical Image Classification Using Different Machine Learning Algorithms. *AL-Rafidain Journal of Computer Sciences and Mathematics, vol. 14*, 135-147.

S. I. Ayon and M. Islam . (2019). Diabetes Prediction: A Deep Learning Approach . *nternational Journal of Information Engineering & Electronic Business, vol. 11*.

S. K. Sarkar and A. Nag. (2017). Identifying patients at risk of breast cancer through decision trees . *International Journal of Advanced Research in Computer Science, vol. 8* , 88-91.

S. Kareem and M. C. Okur. (2017). Evaluation Of Bayesian Network Structure Learning . *2nd International Mediterranean Science and Engineering Congress (IMSEC 2017).* Adana.

S. Kareem and M. C. Okur. (2018). Bayesian Network Structure Learning Using Hybrid Bee Optimization and Greedy Search . *Çukurova University, 2018* (p.). Adana: Çukurova University.

S. Kharya and S. Soni. (2016). Weighted naive bayes classifier: a predictive model for breast cancer detection. *S. Kharya and S. Soni, "Weighted naive bayes classifier: a predictInternational Journal of Computer Applications, vol. 133*, 32-37.

S. Kharya, D. Dubey, and S. Soni. (2013). Predictive machine learning techniques for breast cancer detection. *International Journal of Computer Science and Information Technologies, vol. 4,*, 1023-1028.

S. W. Kareem and M. C. Okur. (2019). Pigeon inspired optimization of bayesian network structure learning and a comparative evaluation. *Journal of Cognitive Science, vol. 20*, 535-552.

S. W. Kareem and M. C. Okur. (2020). Evaluation of Bayesian Network Structure Learning Using Elephant Swarm Water Search Algorithm. In *S. W. Kareem and M. C. Okur, "Evaluation of BayesiHandbook of Research on Advancements of Swarm Intelligence Algorithms for Solving Real-World Problems.* IGI.

T. Nagashima, M. Suzuki, H. Yagata, H. Hashimoto, T. Shishikura, N. Imanaka, et al. (2002). T. Nagashima, M. Suzuki, H. Yagata, H. Hashimoto, T. ShishDynamic-enhanced MRI predicts metastatic potential of invasive ductal breast cancer. *Breast Cancer, vol. 9*, 226.

W. Kareem, R. Z. Yousif, and S. M. J. Abdalwahid. (2020). An approach for enhancing data confidentiality in Hadoop. *W. Kareem, R. Z. Yousif, and S. M. J. Abdalwahid, "An approach for enhancing data conIndonesian Journal of Electrical Engineering and Computer Science* , 1547-1555.

W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. *Designs, vol. 2*, 13-20.

Y. Tan, K. Sim, and F. Ting. (2017). Breast cancer detection using convolutional neural networks for mammogram imaging system. *International Conference on Robotics, Automation and Sciences (ICORAS)*, (pp. 1-5).

Y. Yasui and X. Wang. (2009). *Statistical Learning from a Regression Perspective by BERK, RA.* Wiley Online Library.

هەڵسەنگاندنی بەراوردی لۆگاریتمەکانی فێربوونی ئامێر لە شێرپەنجەی مەمک

پـوختـە:

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (1), Winter 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

شێرپەنجەی مەمک یەکێکە لە هۆکارە سەرەکیەکانی مردن لە ئافرەتدا و بەهۆی گەشەکردنی خانەی سنگی کۆنترۆڵ نەکراوە. دۆزینەوەی زوو و چاودێری کردنی گونجاو تاکە ئامرازی خۆ بەدوورگرتنە لە مردنەکان لە شێرپەنجەی مەمک. خەسڵەتی ووردی توورەکان کارێکی گرنگن لە پیشەی پزیشکیدا. بەهۆی وردی و وردی زۆرەوە، شێوازی فێربوونی ئامێرەکان بە شێوەیەکی گشتی بەکار دەهێنرێت لە ناساندن و پۆلێنکردنی شێوە جۆراوجۆرەکانی شێرپەنجە. لەم وتارە پێداچوونەوەیەدا، نووسەران ئەلگاریتمی جیاوازی فێربوونی ئامێریان تاقیکردۆتەوە و جێبەجێیان کردوون، کە دەکرێت لەلایەن پزیشکەکانەوە بەکاربهێنرێت بۆ ناساندنی خانە شێرپەنجەییەکان بە شێوەیەکی زوو و ورد. ئەم وتارە چەندین لۆگاریتم دەناسێنێت، لەوانە ئامێری بریکاری پشتگیری (SVM)، پۆلێنکەری ناڤ بایس (ئێن بی سی)، تۆڕی دەماری دەستکرد (ANN)، دارستانی هەڕەمەکی (RF)، داری بڕیار (DT)، و k-نزیکترین دراوسێ (KNN). ئەم لۆگاریتمانە ڕاهێنانیان پێکراوە لەگەڵ کۆمەڵێک زانیاری کە پارامیتەرەکانی تووری تێدایە بۆ کەسێک کە شێرپەنجەی مەمکیان هەیە. دوای بەراورد کردنی ئەنجامەکان، ئێمە بەرزترین وردی ئامێری بریکار و دارستانی هەڕەمەکی و بەرزترین ووردەکاری پۆلێنکەری نایڤ بایز (ئێن بی سی)مان دۆزیەوە. لەگەڵ ئەوەشدا، ئێمە چاو بە ژمارەی توێژینەوەکان دەکەینەوە کە لۆگاریتمی فێربوونی ئامێر دابین دەکەن بۆ دۆزینەوەی شێرپەنجەی مەمک.

التقييم المقارن لخوارزميات التعلم الآلي في سرطان الثدي

الملخص:

سرطان الثدي هو واحد من الأسباب الرئيسية في العالم للوفيات بين النساء ويرجع ذلك إلى نمو خلايا الثدي التي لا يمكن السيطرة عليها. الكشف المبكر والرعاية المناسبة هي الوسيلة الوحيدة لتجنب الوفيات الناجمة عن سرطان الثدي. التوصيف الدقيق للأورام مهمة حاسمة في مهنة الطب. نظرا لدقة ودقة عالية، وتستخدم عادة أساليب التعلم الآلي في تحديد وتصنيف أشكال مختلفة من السرطان. في هذه المقالة الاستعراضية ، اختبر المؤلفون خوارزميات التعلم الآلي المختلفة ونفذوها ، والتي يمكن استخدامها من قبل الأطباء لتحديد الخلايا السرطانية بطريقة مبكرة ودقيقة. يقدم هذا المقال عدة خوارزميات، بما في ذلك دعم ناقل آلة (SVM)، ناف بايز المصنف (NBC)، شبكة العصبية الاصطناعية (ANN)، عشوائية الغابات (RF)، شجرة القرار (DT)، وك-الأقرب الجار (KNN). يتم تدريب هذه الخوارزميات مع مجموعة من البيانات التي تحتوي على معلمات الورم للشخص المصاب بسرطان الثدي. بعد مقارنة النتائج، وجدنا أعلى دقة من آلة ناقلات الدعم والغابات العشوائية وأعلى دقة من مصنف الخلجان السذاجة (NBC). بالإضافة إلى ذلك، نستعرض عدد الأبحاث التي توفر خوارزميات التعلم الآلي للكشف عن سرطان الثدي.