Ensemble of Machine Learning Fusion Models for Breast Cancer Detection Based on the Regression Model

Hamzah A. Alsayadi[1, *], Abdelaziz A. Abdelhamid[2], El-Sayed M. El-Kenawy[3], Abdelhameed Ibrahim[4], Marwa M. Eid[5]

[1]Computer Science Department, Faculty of Sciences, Ibb University, Yemen
[2]Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, 11566, Egypt
[3]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt
[4]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, 35516, Mansoura Egypt
[5]Faculty of Artifcial Intelligence, Delta University for Science and Technology, Mansoura, Egypt
Email: hamzah.sayadi@cis.asu.edu.eg; abdelaziz@cis.asu.edu.eg; skenawy@ieee.org; afai79@mans.edu.eg; mmm@ieee.org

Abstract

Breast cancer is one of the deadliest cancers among women worldwide and one of the main causes of mortality for women in the United States. Breast cancer can be detected earlier and with more accuracy, extending life expectancy at a lower cost. To do this, the efficiency and precision of early breast cancer detection can be increased by evaluating the large data that is currently available utilizing technologies like machine learning fusion-based decision support systems. In this paper, we investigate the prediction performance of various regression models and a decision support system based on these models that provided the predicted category along with a prediction confidence measure. The various machine learning (ML) algorithms applied include decision tree regressor, MLP regressor, SVR, random forest regressor, and K-Neighbors regressor. The models are enhanced by average ensemble and ensemble using K-Neighbors regressor. We used the Breast Cancer Wisconsin Dataset from Wisconsin Prognostic Breast Cancer (WPBC) with 569 digitized images of a fine needle aspirate (FNA) of breast mass and 10 real-valued feature information. Among all five machine learning methods, K-Neighbors regressor had the best performance and ensemble using K-Neighbors regressor gave the best accuracy. The results show that there is a decrease in RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI when compared to the traditional methods.

Keywords: Breast Cancer; Ensemble model; Machine learning Fusion; Regression model.

1 Introduction

Cancer is brought on by alterations, or mutations, in the genes that control cell proliferation. The cells' mutations enable uncontrolled cell division and proliferation. Cancer-related or causative agents fall into the following main categories: chemical or toxic substance exposures, ionizing radiation, certain infections, and human genetics. Anything that can cause a normal body cell to develop abnormally has the potential to cause cancer. (As of today's Medical News, 2021) Cancer signs and symptoms vary depending on the type and stage of the disease; however, although general signs and symptoms are not very specific, patients with various cancers may experience the

Doi : https://doi.org/10.54216/FPA.090202
Received: May 16, 2022 Accepted: October 12, 2022

following: fatigue, weight loss, pain, skin changes, changes in bowel or bladder function, unusual bleeding, a persistent cough or voice change, fever, lumps or tissue masses. Breast cancer is the most prevalent type of cancer in the US, followed by lung and prostate cancers, according to the National Cancer Institute, which excluded nonmelanoma skin cancers from these data (Medical News Today, 2021) [1].

Breast cancer is one of the most lethal cancers among women worldwide and one of the main causes of death in the United States. In the U.S., over 41,000 women passed away from breast cancer in 2015, while approximately 266,000 new cases were discovered in 2018. The most prevalent cancer in women worldwide and one of the main causes of cancer death in women is breast cancer [2]. Globally, 2.09 million women were diagnosed with breast cancer in 2018, according to Globocan data, making it the most common type of cancer in women. The incidence of breast cancer varies greatly by geography around the world, with industrialized nations experiencing a markedly greater prevalence than developing nations [3]. The chance of having breast cancer in the United States is one in eight women. The data from 1999 to 2015 indicated a consistent rise over the entire nation, and the number of people with diagnoses increased to 242,476. However, after declining for six straight years from 1999 to 2005, the annual rate of new cases continued to stabilize [3].

Regular breast cancer screenings allow for early detection and treatment of breast cancer. Artificial intelligence and machine learning techniques are now frequently used to enhance cancer detection. Data on cancer has been successfully classified in the medical field using statistical machine learning fusion techniques [4-6]. Many machine learning techniques, such as a convolutional neural network (CNN) or deep neural network (DNN), have outperformed conventional picture identification performance in medical imaging by using data volume and computation as the driving force [4]. Other cutting-edge ML models, including Logistic Regression, Support Vector Machine, Decision Trees, and Nave Bayes, have been found to yield good prediction performance for medical data consisting of numerical, ordinal, and nominal attributes, whereas Artificial Neural Network (ANN) models have been found to yield good prediction performance for image data [7, 8]. One of the earlier studies utilizing ANN in cancer detection, which used demographic data and mammographic findings, had demonstrated that it can achieve an area under the curve (AUC) value of 0.965 with good accuracy even for a big dataset [7]. However, other studies using ANN applications in a small dataset of lung cancer patients revealed poorer accuracy [8].

In this work, breast cancer prediction has been done using different regression methods. We proposed five different models that include regression models such as decision tree regressor, multi-Layer perceptron Model (MLP) regressor, support vector regression (SVR), random forest regressor, and K-Neighbors regressor. In addition, the ensemble model is proposed to optimize the parameters of the regression models. Also, the proposed have been evaluated using statistical metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and R2. Results show the achievement of better performance with decreased error rate when compared to traditional prediction models.

The rest of the paper is organized as follows. In Section 2, we present the literature review. In Section 3, the methodology is presented. Section 4 includes the experimental setup. Finally, Section 5 includes the conclusion and suggestions.

2 Literature Review

Many studies reported in the literature have utilized several data mining techniques in breast cancer. Many researchers have conducted the research using the regression methodology. This article includes some research articles and shows the trends of prediction using the regression analysis method.

In [9], the author discussed a study on the use of several regression approaches to predict disease outbreaks. The importance of regression approaches in several facets of the healthcare industry was covered by the authors in [10]. Based on the data in the work [11], which compares various machine learning approaches, they included the prediction of breast cancer incidence using the linear regression method and the UCI repository data set, and with that, they discovered the accuracy of linear regression to be 99.06%. Authors in [12] employed CUDA parallel programming assistance and the logistic regression approach to detect breast cancer with good results.

In [13], authors investigated a logistic regression approach to create a prediction model for clinical application. They discussed the value of prediction models utilizing logistic regression in the clinical area by considering a real-time

Doi : https://doi.org/10.54216/FPA.090202
Received: May 16, 2022 Accepted: October 12, 2022

dataset and ultimately concluded that regression is crucial to delivering better treatment at a lower cost. Another study [14] discussed how to anticipate the increase of diabetes using the regression approach while considering actual hospital data from Africa using the Weka tool and the logistic regression method. They also employed certain categorization algorithms. In [15], the author used logistic regression to predict heart illnesses and reached an accuracy of the result of 86.89%; they also compared their findings to those of knn and naive Bayes algorithms, but the regression approach produced good results. In [16], the authors used a variety of machine learning techniques to predict the variables that affect a breast cancer patient's chance of survival; in this study, they also used logistic regression as one of the machine learning fusion techniques for prediction.

In [17], the authors demonstrated the prediction of breast cancer utilizing supervised approaches such as logistic regression, KNN, and support vector machines. The logistic regression method produced good results. For the purpose of predicting heart disease, the study in [18] employed a multiple linear regression model, and the level of accuracy attained was satisfactory. In addition, several writers demonstrated how we can forecast the proportion of breast cancer patients who would survive by comparing decision trees and logistic regression in [19] and discussing the various aspects influencing the survival rate. In the research conducted by the authors of this publication [20], a logistic regression model and a Bayesian strategy for gene classification and selection using microarray data are provided.

The research presented in this publication [21] focuses on identifying drug-related physiological effects using HER records. To investigate the methodological changes, a regression model is created. Authors in [22] examined the use of data-mining technology on clinical big data to encourage the generation of research findings that are helpful to physicians and patients. Using machine-learning algorithms, the author's investigations in [23] concerning the prediction and diagnosis of breast cancer have determined which methods are the most efficient in terms of confusion matrices, accuracy, and precision. Additionally, some authors discussed statistical models for the examination of a group of medical data in [24].

3 Methodology

In this work, we use regression models, that are presented in [25, 26], for breast cancer prediction. We employ five regression models: K-Nearest Neighbors, Decision Tree Regressor, MLP Regressor, SVR Regressor, and Random Forest Regressor. Figure 1 shows a summary of the regression model for predicting breast cancer. It is divided into three levels: Data pre-prospecting and feature extraction are included in the first level; five different regression models are contained in the second level's second layer and are used to predict breast cancer along with their underlying principles; and finally, the prediction process is completed in the third level, which includes training, testing, and evaluating the models.

3.1 Data Pre-processing

Data preparation is the process by which we load our data into an appropriate location and make it ready for use in our machine learning training. After gathering all of our data, we'll randomly order them.

3.2 Feature extraction

The process of choosing a subset of pertinent characteristics to be used in the creation of a model is known as feature selection in machine learning and statistics, as well as the variable selection and attribute selection. Selection of the Data File and Feature Wisconsin Breast Cancer (Diagnostic) - Data out of 31 parameters, we chose roughly 8–9 parameters from the machine learning repository. Breast cancer diagnosis, whether it is malignant or benign, is our target parameter. For Feature Selection, the Wrapper Method was utilized. The study's key findings include: Concave points worse, Area worst, Area se, the worst texture, the average texture, the average smoothness, the average radius, and the average symmetry.

Doi : https://doi.org/10.54216/FPA.090202
Received: May 16, 2022 Accepted: October 12, 2022

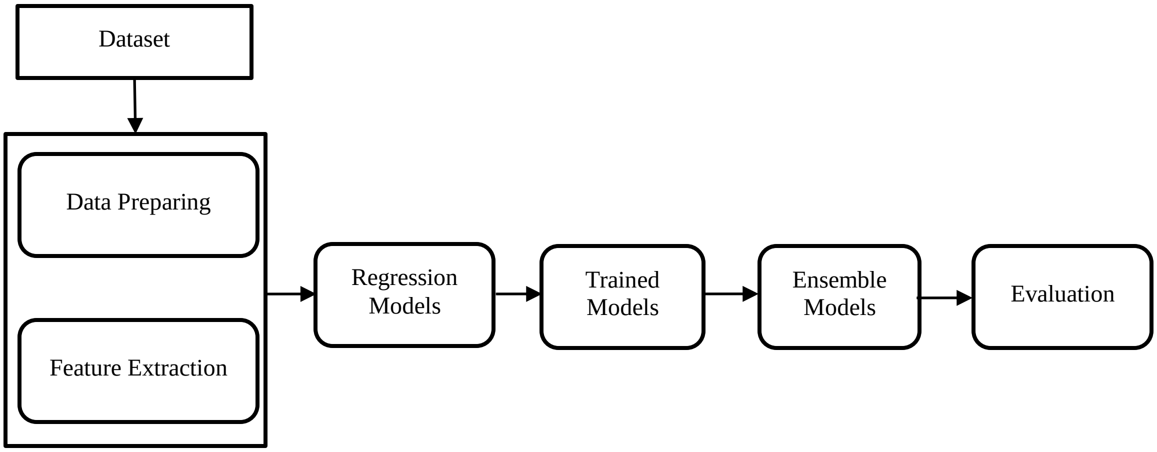

Figure 1: Regression model for the breast cancer prediction

3.3 The proposed models

We proposed a decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor to build five models as presented in [25, 26]. We give a brief overview of these methods as follows.

Decision Trees. The decision tree is a fundamental classification approach, and the classification of instances based on features is represented by the tree structure of the decision tree. In fact, this classification resembles a set of if-then statements quite a bit. Feature selection, decision tree generation, and decision tree pruning are the typical three processes in decision tree learning. The nodes and directed edges of a classification decision tree model. Internal nodes, which represent a feature or attribute, and leaf nodes, which represent a class, are the two different types of nodes in a tree structure [27, 28, 29].

Multi-Layer Perceptron Model. The Multilayer Perceptron (MLP) is a feed forward neural network type that is used for air pollution prediction because it has the capacity to build incredibly complex nonlinear models. The network receives the necessary input parameters, causing the MLP to fire. The input signals produced by these input parameters are transmitted across the network, first from the input layer to the hidden layer and then from the hidden layer to the output layer. The weights, which are real numerical values, are multiplied by the scaled input vector that is introduced by the neurons in the input layer [30, 31, 32].

Support Vector Regression. One use for the 1954-invented support vector machines (SVM) is support vector regression (SVR). The foundations of SVM are structural minimization and statistical learning theory. By transferring the data from a low-dimensional to a high-dimensional space, the kernel function's main goal is to convert the data from a nonlinear to a linear feature space. Radial basis functions, linear basis functions, and polynomial basis functions are examples of common kernel functions that can be utilized in SVR [33, 34, 35].

Random Forest. To produce an ensemble of trees, the Random Forest (RF) categorization aggregates the results of many decision-making trees. A single decision tree's ability to build either a very specific or basic model is used to support this claim [36].

k Nearest Neighbors. The k-Nearest Neighbor algorithm is another fundamental classification and regression analysis technique that presupposes that every instance has been given a class label based on the training data. A new instance is categorized by the class of its k closest training instances when it is being categorized. As a result, the kNN method lacks an explicit learning process. It divides the feature space in reality in accordance with the type of training data, and then employs this procedure as the classification model. The distance between test data and training data, the choice of the value of k, and the classification statement all play fundamental roles in kNN models. Since the feature space in kNN is an n-dimensional real vector space, there are two possible methods for calculating the distance between two instance points: Euclidean distance and distance. In this study, the kNN analysis uses the Euclidean distance. Additionally, the model's categorization output will be considerably impacted by the choice of k

Doi : https://doi.org/10.54216/FPA.090202
Received: May 16, 2022 Accepted: October 12, 2022

value .

The Proposed Ensemble Model. Each regression model in the suggested ensemble has its parameters optimized before being combined with other models to form a single ensemble. This ensemble includes the Decision Tree Regressor, MLP Regressor, SVR, Random Forest Regressor, and KNeighbors Regressor. The best-predicted value that most closely resembles the air quality input attributes is the ensemble model's output.

3.4 Regression Model Training

After the feature extraction step, the features represent the data training for models. Five separate models were subsequently created for testing and training. In order to deploy the models, you must first confirm the prediction value range. If necessary, you should then forecast if the air quality levels are satisfactory or good; if not, the models and datasets need to be improved once more.

3.5 Evaluation

The most popular evaluation criteria are the correlation coefficient (R2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), MBE, RRMSE, NSE, and WI. The R2 value shows the fitting degree of regression. MAE represents the difference between predicted and actual values. RMSE focuses on the impact of extreme values based on MAE, while RAE calculates the variance of a model when comparing the performance of different models. As MAE and RMSE depend on the scale of the data that’s why RAE can be extremely helpful when comparing different data with different scales.

4 Experimental Results

4.1 Dataset

We use the Breast Cancer Wisconsin (Prognostic) Data Set. Each entry contains the follow-up information for a single case of breast cancer. Since 1984, Dr. Wolberg has seen a steady stream of patients, and only those cases with invasive breast cancer and no signs of distant metastases at the time of diagnosis are included in this list. From a digital image of a fine needle aspirate (FNA) of a breast lump, the first 30 features are calculated. They characterize the traits of the visible cell nuclei in the image. Multisurface Method-Tree (MSM-T), a classification method that builds a decision tree using linear programming, was used to achieve the separation mentioned above. An exhaustive search in the domain of 1-4 features and 1-3 separation planes was used to select relevant features.

4.2 Results and discussion

We evaluate the proposed models based on the communities and crime dataset. The metrics, that are presented in subsection 3.5, are used to evaluate the proposed model. The results of the proposed model are being shown in Table 1.

Table 1: The results of the regression model

Model	RMSE	MAE	MBE	R	R^2	RRMSE	NSE	WI
Decision Tree	0.0678	0.0504	-0.0074	0.95	0.91	22.42	0.90	0.87
Multi-layer Perceptron	0.0455	0.0333	-0.0014	0.98	0.96	15.05	0.95	0.91
support vector regression	0.0417	0.0312	0.0028	0.98	0.96	13.80	0.96	0.92
Random forest	0.0905	0.0703	-0.0210	0.91	0.83	29.94	0.82	0.81
K-nearest neighbors	**0.0181**	**0.0126**	**-0.0010**	**0.996**	**0.99**	**5.99**	**0.99**	**0.97**

As shown in Table 1, we note that the experimental results enhance the results. It is further observed all the five models produce good results.

In regression results of decision tree, the RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI terms are reported by 0.0678, 0.0504, -0.0074, 0.95, 0.91, 22.42, 0.90, and 0.87, respectively.

In regression results of multi-layer perceptron, the RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI terms are

Doi : https://doi.org/10.54216/FPA.090202
Received: May 16, 2022 Accepted: October 12, 2022

reported by 0.0455, 0.0333, -0.0014, 0.98, 0.96, 15.05, 0.95, and 0.91, respectively.

In regression results of SVR, the RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI terms are reported by 0.0417, 0.0312, 0.0028, 0.98, 0.96, 13.80, 0.96, and 0.92, respectively.

In regression results of random forest, the RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI terms are reported by 0.0905, 0.0703, -0.0210, 0.91, 0.83, 29.94, 0.82, and 0.81, respectively.

In regression results of K-nearest neighbors, the RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI terms are reported by 0.0181, 0.0126, -0.0010, 0.996, 0.99, 5.99, 0.99, and 0.97, respectively.

We note that the K-nearest neighbors achieved the best results for RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI terms.

Two experiments of ensembles are conducted in this work including average ensemble and ensemble using K-Neighbors Regressor. These models are evaluated based on the RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI metrics as shown in Table 2.

Table 2: The results of the ensemble models

Model	RMSE	MAE	MBE	R	R^2	RRMSE	NSE	WI
Average Ensemble	0.0399	0.0272	-0.0056	0.98	0.97	55.99	0.97	0.93
Ensemble using KNN regressor	**0.0010**	**0.0070**	**-0.0004**	**0.999**	**0.998**	**3.29**	**0.998**	**0.98**

The average ensemble yielded the results of RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI by 0.0399, 0.0272, -0.0056, 0.98, 0.97, 55.99, 0.97, and 0.93, respectively. While ensemble using K-Neighbors regressor achieved the results of RMSE, MAE, MBE, R, *R*2, RRMSE, NSE, and WI by 0.0010, 0.0070, -0.0004, 0.999, 0.998, 3.29, 0.998, and 0.98.

The above results illustrate that our ensemble model is more accurate than the other five regression models for the prediction. The regression results allow to believe that our proposed ensemble model is efficient for handling these data. In addition, the ensemble using KNN regressor achieved the best results for breast cancer prediction.

5 Conclusions

In this paper, we proposed different regression models for breast cancer prediction. In particular, we aimed to predict the number of breast cancer prediction cases receive. The proposed models include decision tree regressor, MLP regressor, SVR, random forest regressor, and K-Neighbors regressor. The average ensemble and ensemble using K-Neighbors regressor are used for enhancing the proposed models. The MSE, MAE, MBE, R, R2, RRMSE, NSE, and WI metrics are used to evaluate the proposed models. The results show that the regression models perform well, which are the competitive models. Also, the ensemble using KNN regressor achieved the best results for breast cancer prediction.

References

[1] Kaustubh Chakradeo, Sanyog Vyawahare, and Pranav Pawar. Breast cancer recurrence prediction using machine learning. In *2019 IEEE Conference on Information and Communication Technology*, pp. 1-7. IEEE, 2019.

[2] Mochen Li, Gaurav Nanda, Santosh Chhajedss, and Raji Sundararajan. Machine Learning-Based Decision Support System for Early Detection of Breast Cancer. *Indian Journal of Pharmaceutical Education and Research* 54, no. 3 (2020): S705-S715.

[3] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians* 69, no. 1 (2019): 7-34.

[4] L. Ghasem Ahmad, A. T. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. R. Razavi. Using three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform* 4, no. 124 (2013): 3.

[5] Anusha Bharat, N. Pooja, and R. Anishka Reddy. Using machine learning algorithms for breast cancer risk prediction and diagnosis. In *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, pp. 1-4. IEEE, 2018.

Doi : https://doi.org/10.54216/FPA.090202

Received: May 16, 2022 Accepted: October 12, 2022

[6] Dursun Delen, Glenn Walker, and Amit Kadam. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine* 34, no. 2 (2005): 113-127.

[7] Turgay Ayer, Oguzhan Alagoz, Jagpreet Chhatwal, Jude W. Shavlik, Charles E. Kahn Jr, and Elizabeth S. Burnside. Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration. *Cancer* 116, no. 14 (2010): 3310-3321.

[8] El-Kenawy, El-Sayed M., Marwa Eid, and Alshimaa H. Ismail. "A New Model for Measuring Customer Utility Trust in Online Auctions." International Journal of Computer Applications 975: 8887.

[9] Yen-Chen Chen, Wan-Chi Ke, and Hung-Wen Chiu. Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. *Computers in biology and medicine* 48 (2014): 1-7.

[10] Jithendra PR Nayak, B. D. Parameshachari, KM Sunjiv Soyjaudah, Reshma Banu, and A. C. Nuthan. Identification of PCB faults using image processing. In *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, pp. 1-4. IEEE, 2017.

[11] Hakizimana Leopord, W. Kipruto Cheruiyot, and Stephen Kimani. A survey and analysis on classification and regression data mining techniques for diseases outbreak prediction in datasets. *Int. J. Eng. Sci* 5, no. 9 (2016): 1-11.

[12] Liang Hong, Mengqi Luo, Ruixue Wang, Peixin Lu, Wei Lu, and Long Lu. Big data in health care: Applications and challenges. *Data and information management* 2, no. 3 (2018): 175-197.

[13] J Sivapriya, V Aravind Kumar, S Siddarth Sai, and S Sriram. Breast cancer prediction using machine learning. Int. J. Recent Technol. Eng. 8 (4) (2019).

[14] A. Peretti, and F. Amenta. Breast Cancer Prediction by Logistic Regression with CUDA Parallel Programming Support. *Breast Can Curr Res* 1, no. 111 (2016): 2..

[15] Maren E. Shipe, Stephen A. Deppen, Farhood Farjah, and Eric L. Grogan. Developing prediction models for clinical use using logistic regression: an overview. *Journal of thoracic disease* 11, no. Suppl 4 (2019): S574.

[16] M. N. Sohail, Ren Jiadong, M. M. Uba, M. Irshad, Musavir Bilal, Usman Akbar, and Tahir Rizwan. Forecast Regression analysis for Diabetes Growth: An inclusive data mining approach. *Int. J. Adv. Res. Comput. Eng. Technol.(IJARCET)* 7, no. 9 (2018): 715-721.

[17] Reddy Prasad, Pidaparthi Anjali, S. Adil, and N. Deepa. Heart disease prediction using logistic regression algorithm using machine learning. *International journal of Engineering and Advanced Technology* 8, no. 3S (2019): 659-662.

[18] Mogana Darshini Ganggayah, Nur Aishah Taib, Yip Cheng Har, Pietro Lio, and Sarinder Kaur Dhillon. Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC medical informatics and decision making* 19, no. 1 (2019): 1-17.

[19] Ch Shravya, K. Pravalika, and Shaik Subhani. Prediction of breast cancer using supervised machine learning techniques. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* 8, no. 6 (2019): 1106-1110.

[20] K. Polaraju, and D. Durga Prasad. Prediction of heart disease using multiple linear regression model. *International Journal of Engineering Development and Research Development* 5, no. 4 (2017): 1419-1425.

[21] El-kenawy, El-Sayed M., Marwa M. Eid, and Abdelhameed Ibrahim. "Anemia estimation for covid-19 patients using a machine learning model." Journal of Computer Science and Information Systems 17, no. 11 (2021): 2535-1451.

[22] Somayeh Momenyan, Ahmad Reza Baghestani, Narges Momenyan, Parisa Naseri, and Mohammad Esmaeil Akbari. Survival prediction of patients with breast cancer: comparisons of decision tree and logistic regression analysis. *International Journal of Cancer Management* 11, no. 7 (2018).

[23] Xiaobo Zhou, Kuang-Yu Liu, and Stephen TC Wong. Cancer classification and prediction using logistic regression with Bayesian gene selection. *Journal of Biomedical Informatics* 37, no. 4 (2004): 249-259.

[24] Matthew E. Levine, David J. Albers, and George Hripcsak. Methodological variations in lagged regression for detecting physiologic drug effects in EHR data. *Journal of biomedical informatics* 86 (2018): 149-159..

[25] Ibrahim, Abdelhameed, and El-Sayed M. El-kenawy. "Applications and datasets for superpixel

Doi : https://doi.org/10.54216/FPA.090202

Received: May 16, 2022 Accepted: October 12, 2022

techniques: A survey." Journal of Computer Science and Information Systems 15, no. 3 (2020): 1-6.

[26] Wen-Tao Wu, Yuan-Jie Li, Ao-Zi Feng, Li Li, Tao Huang, An-Ding Xu, and Jun Lyu. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Military Medical Research* 8, no. 1 (2021): 1-12.

[27] Mohammed Amine Naji, Sanaa El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida Ait Abdelouhahid, and Olivier Debauche. Machine learning algorithms for breast cancer prediction and diagnosis. *Procedia Computer Science* 191 (2021): 487-492..

[28] Ibrahim, Abdelhameed, Seyedali Mirjalili, Mohammed El-Said, Sherif SM Ghoneim, Mosleh M. Al-Harthi, Tarek F. Ibrahim, and El-Sayed M. El-Kenawy. "Wind speed ensemble forecasting based on deep learning using adaptive dynamic optimization algorithm." IEEE Access 9 (2021): 125787-125804.

[29] El-Sayed Towfek, M., and M. Saber El-kenawy. "Reham Arnous. An Integrated Framework to Ensure Information Security Over the Internet." International Journal of Computer Applications 178, no. 29 (2019): 13-15.

[30] Hamzah A. Alsayadi, Nima Khodadadi, and Sunil Kumar. Improving the Regression of Communities and Crime Using Ensemble of Machine Learning Models. *Journal of Artificial Intelligence and Metaheuristics* 1.1 (2022): 27-34.

[31] Wenjuan Wei, Olivier Ramalho, Laeticia Malingre, Sutharsini Sivanantham, John C Little, and Corinne Mandin. Machine learning and statistical models for predicting indoor air quality. *Indoor Air*, 29(5):704– 726, 2019.

[32] S Abdullah, M Ismail, and AN Ahmed. Multi-layer perceptron model for air quality prediction. *Malaysian Journal of Mathematical Sciences*, 13:85–95, 2019.

[33] Ruizhi Zhong, Raymond L Johnson Jr, and Zhongwei Chen. Using machine learning methods to identify coals from drilling and logging-while-drilling lwd data. In *Asia Pacifc Unconventional Resources Technology Conference, Brisbane, Australia, 18-19 November 2019*, pages 970–994. Unconventional Resources Technology Conference, 2020.

[34] Zhi-Hua Zhou. *Machine learning*. Springer Nature, 2021.

[35] Eid, Marwa M., El-Sayed M. El-kenawy, and Abdelhameed Ibrahim. "A binary sine cosine-modified whale optimization algorithm for feature selection." In 2021 National Computing Colleges Conference (NCCC), pp. 1-6. IEEE, 2021.

[36] Mochen Li, Gaurav Nanda, Santosh Chhajedss, and Raji Sundararajan. Machine Learning-Based Decision Support System for Early Detection of Breast Cancer. Indian Journal of Pharmaceutical Education and Research 54, no. 3 (2020): S705-S715.

Doi : https://doi.org/10.54216/FPA.090202

Received: May 16, 2022 Accepted: October 12, 2022