Abant Sağlık Bilimleri ve Teknolojileri Dergisi

Abant Journal of Health Sciences and Technologies

2022;2(2):72–87

An Applied Analysis of Breast Cancer Diagnosis by Using Different Methods

Farklı Yöntemler Kullanarak Meme Kanseri Teşhisinin Uygulamalı Bir Analizi

İclal Çetin Taş

Geliş Tarihi (*Received*): 24.09.2022 **Kabul Tarihi (*Accepted*):** 30.11.2022 **Yayın Tarihi (*Published*):** 21.12.2022

Öz: Meme kanseri dünyanın her bölgesinde en sık görülen kanser türlerinden biridir. Meme kanserinden ölümler her yıl katlanarak artıyor. Tüm kanser türlerinde olduğu gibi meme kanserinde de erken teşhis önemlidir ve birçok kez hayat kurtarır. Bu nedenle erken tanıyı kolaylaştırmak veya hastalığı erken öngörmek için birçok çalışma yapılmaktadır. Tahmin uygulamalarında kullanılan yöntemlerin başında makine öğrenmesi yöntemleri gelmektedir. Bu çalışmada, genel regresyon sinir ağları (GRNN), radyal temel fonksiyon (RBF), karar ağacı ormanı (DTF) ve gen ekspresyon programlaması (GEP), destek vektör makineleri (SVM), çok katmanlı algılayıcılar (MLP) meme kanseri Wisconsin Diagnostic veriseti üzerinde analiz edilmiştir. Elde edilen sonuçlara göre makine öğrenmesi algoritmaları kullanılarak meme kanserinin erken teşhisine katkı sağlamak için sınıflandırıcılar arasında performans değerlendirmesi ve karşılaştırma yapılmıştır. Elde edilen en iyi doğruluk %98.8 oranı ile GRNN algoritmasından elde edilmiştir.

Anahtar Kelimeler: Makine Öğrenmesi, Meme Kanseri, Sınıflama, Tahminleme, Teşhis.

&

Abstract: Breast cancer is one of the most common cancer types in every region of the world. Deaths from breast cancer are increasing exponentially every year. As with all cancer types, early diagnosis is important in breast cancer and saves lives many times over. For this reason, many studies are carried out to facilitate early diagnosis or to predict the disease early. Machine learning methods are at the forefront of the methods used in prediction applications. In this study, general regression neural networks (GRNN), radial basis function (RBF), decision tree forest (DTF) and gene expression programming (GEP), support vector machines (SVM) and multilayer perceptron (MLP) were analyzed on the Wisconsin Diagnostic Breast Cancer dataset. According to the results obtained, a performance evaluation and comparison were made between the classifiers to contribute to the early diagnosis of breast cancer by using machine-learning algorithms. The best accuracy was obtained from the GRNN algorithm, it is 98.8%.

Keywords: Breast Cancer, Classification, Diagnosis, Machine Learning, Prediction.

Atıf/Cite as: Taş İÇ. An Applied Analysis of Breast Cancer Diagnosis by Using Different Methods. Abant Sağlık Bilimleri ve Teknolojileri Dergisi, 2022;2(2):72-87.

Introduction

The most common type of cancer is breast cancer which is seen in middle-aged women all over the world. Apart from this, there are cases where it is seen in different age groups or gender. It is an important disease that causes death for most individuals. Although it is especially seen in women, the same threat is present for men as well. It is the second most common type of cancer after lung cancer. It is the main cause of cancer deaths, especially in women. (1) In Figure 1, breast cancers incidence and mortality rates graph are shown.

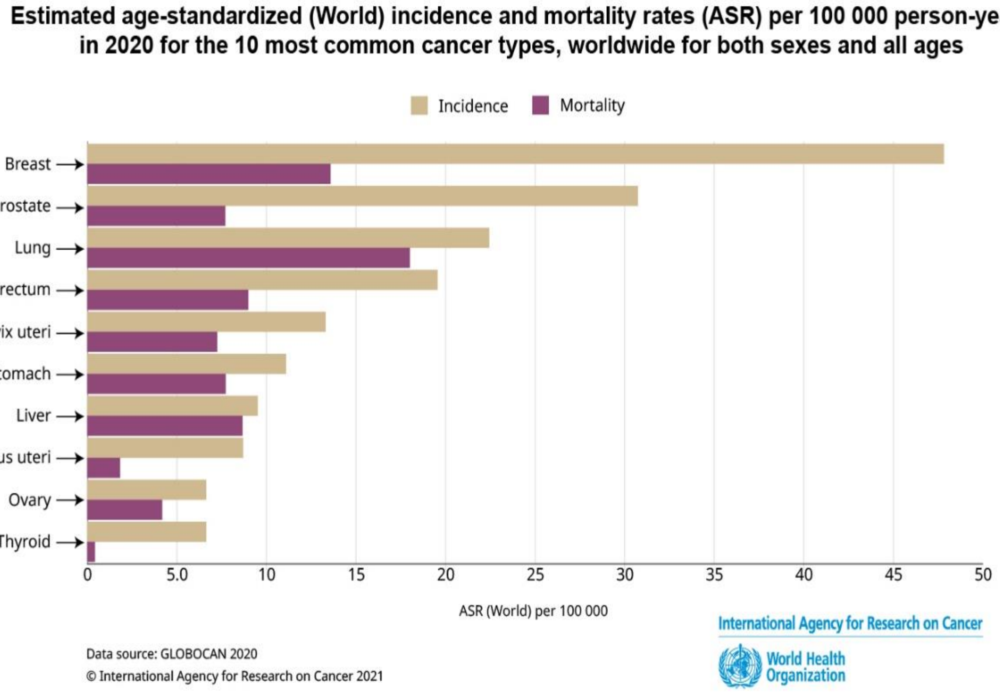

Figure 1. Breast Cancer incidence and mortality rates worldwide (2).

According to the data published by the Ministry of Health, the incidence of breast cancer is 46.3 per 100 thousand in the world, while it is 92.6 for Northern European countries, 39.2 for East Asia, 38.3 for the United States, and 45.6 for Turkey. As in most types of cancer, these negative results can be prevented by early diagnosis. If it can be detected at a relatively early stage, patients' mortality rate from breast cancer can be reduced. Thanks to the newest, most effective, and most advanced monitoring methods with different types of biomedical equipment, most of these cancers are diagnosed while the disease is still in the localized stage.

Connection points with lymph are very critical. Breast cancer is a disease that occurs as a result of a tumor that occurs due to the change and uncontrolled proliferation of one of the cell groups that make up the breast tissue. Similar to other types of cancer, cancerous tissue tends to spread. In cases of breast cancer, the disease first spreads to the immediate surroundings and then to the lymph nodes close to the breast. With early diagnosis, this spread can be prevented to some extent and the patient can regain her health. Even if it does not result in death otherwise, breast loss in this process has a great psychological effect, especially on women. Even this situation may be the subject of a different study on its own.

We can briefly summarize the main causes of the disease as follows:

- Especially in women between the ages of 50-70 and in the post-menopausal period,
- Genetic factors (family history of the disorder)
- Having had breast cancer before,
- Early menstruation, late menopause,
- Having never given birth or having the first birth after the age of 30,
- Having given birth but not breastfed the baby,
- Taking long-term hormone therapy,
- Using cigarettes and tobacco products,
- Obesity,
- Lack of physical activity

We can list the common symptoms of the disease as a palpable hardness or mass in the breast, asymmetry or changes in shape & direction between the breasts, recession in the nipple or breast skin, redness in the breast, scar, eczema, etc., swelling in the armpit, and discharge in the breast. However, even the most common of these symptoms, stiffness and swelling, may not be seen. Compared with men, women have an approximately 100 times higher lifetime risk of developing breast cancer. Apart from being a female individual, increasing age is the most important risk factor (3,4).

The application of machine learning techniques in healthcare analytics is increasing day by day. Certainly, the analysis of the patient's clinical data and the judgment of the physician are the most important features of diagnosis (5). It has been observed that the results obtained from the classification methods are close to the actual ones in the studies carried out to solve or increase awareness of many medical problems and prevent them. Since cancers are long-term diseases, although this process has a direct impact on the life expectancy or quality of the patient, it also contributes economically to the public by increasing the quality of life of health personnel and reducing the burden on the health sector. Accurate and timely prediction of breast cancer enables doctors and healthcare providers to make the most appropriate decision about patient treatment.

According to studies, breast cancer is the most common type of cancer that causes death in women. For this reason, there are many studies conducted in this field. Especially, machine learning has serious success rates in predicting the future from existing disease data. In this study, various machine learning algorithms were applied to breast cancer disease data shared by the University of Wisconsin. Recently, it has provided prolongation of life expectancy in cancer patients with early diagnosis and treatment opportunities. The data and publications published by the World Health Organization support the ideas on this subject. The aim of the WHO Global Breast Cancer Initiative (GBCI) is to decrease global breast cancer deaths by 2.5% per year, thereby preventing 2.5 million breast cancer deaths globally between 2020 and 2040. Decreasing global breast cancer deaths by 2.5% per year will prevent this. 25% of breast cancer deaths in women under 70 by 2030 and 40% by 2040 (Breast Cancer, 2021). The primary way for this purpose is seen as early diagnosis. Some examples of benign and malignant tumors can be seen in Figure 2.

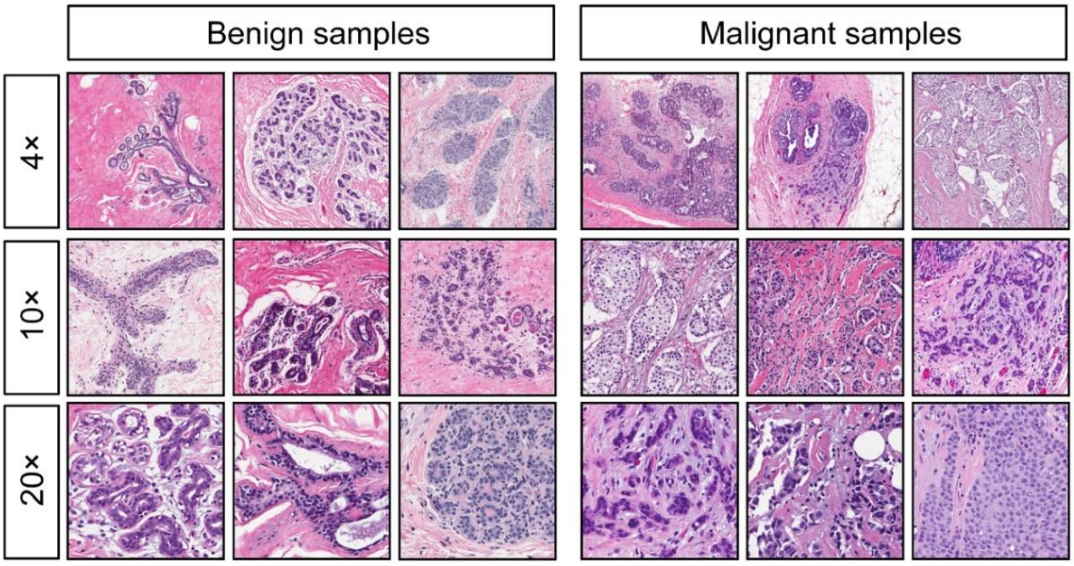

Figure 2. Breast cancer samples types images (6).

With the acceleration of development in the fields of biomedical engineering and computer technologies, various clinical factors have been recorded to support scientific studies on the early diagnosis and treatment processes of many diseases. The situation is similar in breast cancer disease. (7). In this study, GRNN, RBF, GEP and DTF methods are used.

The GRNN network needs less sampling than other networks for the learning procedure. One advantage of the GRNN network is that a single iteration is sufficient to achieve the desired outcome. Therefore, it is easy to implement and is frequently used for estimation in modeling and classification problems (8).

The RBF neural network also consists of a three-layer feed-forward architecture, but the hidden layer mostly uses the Gaussian function and is called the radial basis layer. Each neuron includes a point-centered RBF. Centers and spreads are calculated by training. To apply the RBF kernel function to the distance using the spread values, a traced neuron calculates the Euclidean distance of the input vector and the test case from the neuron's center point (9).

Decision tree ensembles (DTEs) are prediction methods that can more easily capture relevant features. Among different DTEs, Random forest (RF) is generally faster, but more resistant to noise and congestion. Several trees are generated that can classify each instance, and the prediction tag for a particular sample member is resolved by the voting mechanism base on all tree predictions. Each tree in RF is constructed using a subset feature and substituted random sampling.

DT models are clear and have the ability to make transparent models; Since it is built by combining in DTEs like RF, it also inherits the set of rules from the previous generation and therefore changes in size and takes on a complex shape. It is seen that the studies are generally based on improving the process of determining the accuracy of cancer diagnosis by researchers. And this approach is done through community learning and it is sometimes overlooked that the situation involves an interpretive process. (10).

A DT iteratively classifies each possible results of the data. It can basically be modeled as a network diagram of each non-leaf node having a specific attribute test, each branch a result of that test, and each leaf node a classification or decision on it. The node at the top label of the tree is called the root node, which corresponds to the best. It has applications for datasets in many fields. (11).

GEP is an improved way based on genetic algorithm and genetic programming (12). Parse trees of genetic programming like protein molecules in their complex and unique hierarchical structure. Parse trees can perform a wide variety of functions, but a weakness is that their breeding by modification is very limited in evolutionary terms. Only a limited range of changes is possible, as changes must occur in the parse tree itself. In the literature, special genetic operators have been developed that change the branching between trees by working at the tree level. Although this approach has advantages, it can sometimes cause some limitations. For example, in the crossover operation, branches are selected that are swapped between two parent trees to generate new offspring by developing mathematically dense blocks of content with smaller but complex structures to develop hierarchical solutions. Unlike natural point mutation, the mutation operator in GP selects a node in the detached tree and proceeds by replacing the underlying branch with a new branch created, which is usually applicable for tree structures that are not too large (13).

In its most basic form, the SVM method can be summarized as the use of a hyperplane to distinguish between two classes of data (14). By using WBCD dataset, cancer individuals or healthy individuals can be categorized for early detection of breast cancer. Since this situation is modeled as a classification problem, it can be easily solved with SVM.

The MLP algorithm is artificial neural network based and an artificial neural network is a mathematical or computational model based on a biological neural network. In other words, ANN is a simulation of the biological nervous system. One of the usage areas of ANN is classification and multilayer perceptron is often preferred in this area. In our study, MLP was preferred for classification. An MLP generally uses an input layer, an output layer, and at least one hidden layer (15,16) .

Our aim is to predict and diagnosis breast cancer with this study by using machine-learning algorithms. The content of this paper is organized with 4 section exclude abstract part. Section 1 introduction consist of information about previous studies and their results about breast cancer. In section 2 explain materials and methods. Experiments and results describe the proposed methodology and for all results in section 3. Section 4 concludes the paper.

Material and Methods

Dataset

In our study, Breast Cancer Wisconsin Diagnostic dataset is used which is obtained from University of Wisconsin Hospitals Madison Breast Cancer Database (17). 3 different data sets were created with these data. These are Wisconsin Breast Cancer (Original), Wisconsin Diagnosis Breast Cancer (WDBC), Wisconsin Prognosis Breast Cancer (WPBC)(18). We collected a scientific dataset of breast cancer patients from Kaggle the features of dataset are computed from a digitized image of a breast cancer sample obtained from fine-needle aspirate (FNA) (19).Dataset has 569 samples (Benign: 357 Malignant: 212) and 11 features. They are listed below:

i. Id
ii. Diagnosis
iii. Radius
iv. Texture
v. Area
vi. Perimeter
vii. Smoothness
viii. Compactness
ix. Concavity
x. Concave points
xi. Symmetry
xii. Fractal dimension

Methods

All experimental procedures were performed using Matlab 2020b version in this study. Matlab machine learning toolbox is used for this study. GRNN, RBF, DTF, GEP, SVM and MLP models are applied the dataset.

GRNN is a single pass learning algorithm. It has been shown that the algorithm provides smooth transitions from one observed value to another, even when working with sparse data in a multidimensional measurement space (20). Basic formulas of GRNN are shown

$$D_i = (X - X_i)^T (X - X_i)$$

$$Y(X) = \frac{\sum_{i=1}^{n} Y_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} \left(-\frac{D_i^2}{2\sigma^2}\right)}$$

X_i refers to normal distribution. Y is predicted value. D_i represents Euclidean distance and exponential part is corresponded to Gaussian kernel function. Basic GRNN structure can be find in Figure 3. GRNN parameters and kernel function details are shown in Table 1.

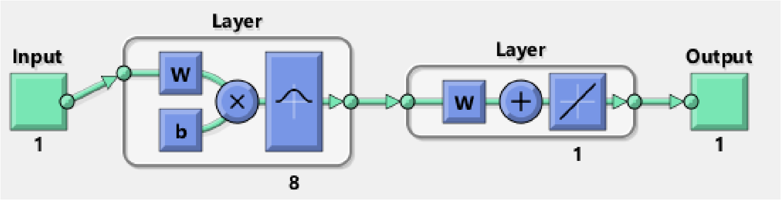

Figure 3. A sample of GRNN.

Table 1. GRNN methods parameter values.

GRNN	
Parameter type	Type, Value or Range
Kernel Function	Gaussian
Neuron Number	2
σ	0.001-10
Validation	10 cross-fold validation

RBF has three main parts: input, hidden and output layers. Radially symmetric basis functions prefer in hidden layer (21) (RBF architecture in Figure 4). This model performance is based on spread and maximum number of neurons. The hidden neurons consist Gaussian transfer functions because these function outputs are reversely proportional to the space from the center of the neuron. Parameters and kernel information is shown in Table 2.

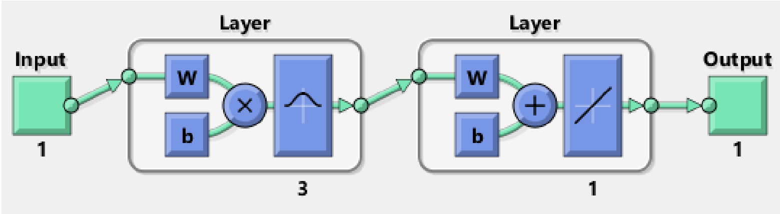

Figure 4. A sample of RBF.

DTF is one of the machine learning algorithms used in both classification and regression problems with a predefined target variable. It has a top-down flow. Decision trees are often developed with a design goal of human-like thinking. In this respect, it is simpler to understand, interpret and visualize the data.

Table 2. RBF methods parameter values.

RBF	
Parameter type	Type, Value or Range
Kernel Function	Gaussian
Neuron Number	56
λ	0.001-10
Regularization Lambda for final weights	1.6400e-007 after 4 iterations.
r	0.001-513.421
Validation	10 cross-fold validation

Decision trees are recursive structures. They start from a basic node, divide into new branches and form the tree structure. When the algorithm runs by using parameters in Table 3, the entered value proceeds in a certain way by looking at the nodes and gives results.

Table 3. DTF methods parameter values.

DTF	
Parameter Type	Type, Value or Range
Number of Trees in Forest	200
Max. Tree Levels	50

Genetic programming is built with tree-like individuals that can represent mathematical expressions (22). The GEP algorithm is also an approach that simulates biological evolutionary processes to mine a function that has the best fit for representing data relationships. GEP consists of five basic components. These are function set, terminal set, fitness function, control parameters, and termination condition (12) . The algorithm uses processes such as selection, crossover and mutation to generate offspring in the evolution process. Each offspring is evaluated with a fitness function. Like natural selection, an individual with good fitness has a better chance of being selected to produce the next generation. The process and development continue until a satisfactory function is found that can describe the data relationships. Population size, maximum generations, head length and genes per chromosome are effective values for GEP algorithm. Selection of fitness function is one of the basic criteria for algorithm performance. The hits with penalty are basic fitness function type but it is very efficient, and it uses in literature commonly. It is based on TP and TN values.

f_{i} is the fitness function. It refers an individual program corresponds to the number of hits. It is shown below formula:

$$If\ (\mathrm{TP} = 0 \text{ or } TN_{\mathrm{i}} = 0)$$
$$\text{then } f_{\mathrm{i}} = \mathrm{TP} + \mathrm{TN}$$
$$\text{else } f_{\mathrm{i}} = 0$$

So, for this fitness function, maximum value of fitness is given by $f_{max} = \mathrm{n}$. In here n refers to the number of fitness cases. GEP architecture is used by using parameters and function in Table 4.

Table 4. GEP methods parameter values.

GEP	
Parameter Type	Type, Value or Range
Population Size	50
Maximum Generations	2000
Head Length	6
Fitness Function	Number of hits with penalty

Multilayer perceptron (MLP) is well known machine learning method. MLP architecture is consisted of 3 layer one of them is hidden layer. Activation function is selected logistic function for hidden and output layers. The back propagation algorithm, which is widely used in the training of MLP type ANNs, aims to optimize the weighted connections between the input layer hidden layer and the hidden layer-output layers depending on the error occurring at the network output. The input and associated output vector are used to train the network. After the input vector that determines, the pattern is given to the first layer of the network, it passes through the hidden layers in between and reaches the last output layer with the help of weights. After each neuron in the network receives the arithmetic sum of the weight vectors terminating in it, the result is transferred to all neurons of the next layer depending on the activation function used (23–25).

Support vector machine (SVM) was suggested by Vapnik et al (26). It is a method used differently in many breast cancer classification applications containing images (27,28) and numerical data. SVM is a frequently used method in regression and classification problems. In some cases, feature extraction is powered by different approaches such as the F score. In this study, it was used for the diagnosis of breast cancer, which was reduced to a classification problem. The basic structure in the SVM algorithm is to determine a boundary line between groups. One of the important criteria of the determined border line is a place far from the data of both groups. However, the place where this border line will be found should be the farthest place from the members of the two groups. One of the most important criteria in the SVM algorithm is how to draw this boundary. Besides, kernel function selection is important. In this study, linear kernel function is used for SVM method. (29–34).

Experiments and Results

We followed flowchart for our study in Figure 5. Firstly, we started pre-processing applications. This step includes determination of train- test data rate, and selection of training type. Then, determining method and application of selected method are followed. Obtaining results evaluated calculated comparison criteria and the last step analysis and discuss the performance criteria.

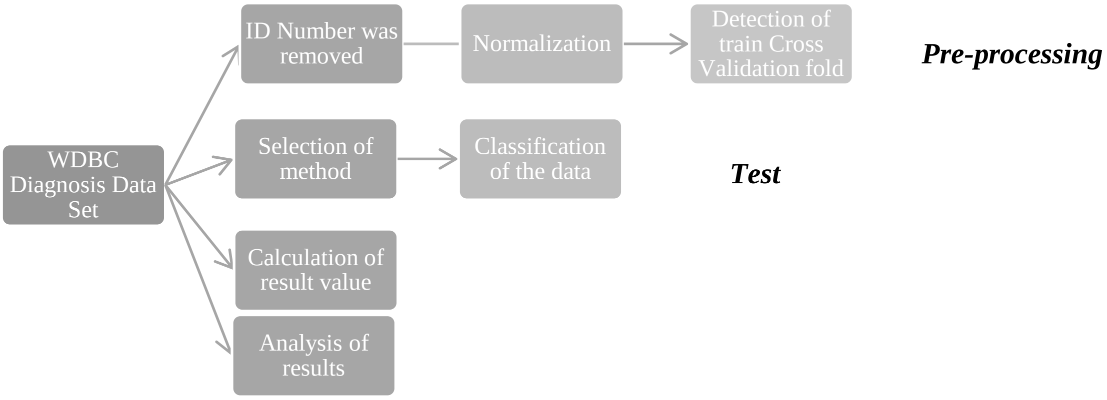

Figure 5. Flowchart of study's steps.

The confusion matrix, the number of correct and incorrect predictions, is obtained by dividing the number values into classes and summarizes the prediction results by giving information about the model applied. The confusion matrix applied in this study is shown in Table 5.

Table 5. Confusion matrix.

Actual Value	**Predicted Value**	
	Positive	Negative
Positive	True positive (TP)	True negative (TN)
Negative	False positive (FP)	False negative (FN)

Our performance criteria are shown in below formulas(35–38):

$$Prediction\ accuracy\ (\%) = \frac{TP + TN}{TP + TN + FN + FN}$$

$$Sensitivity(\%) = \frac{TP}{TP + FN}\ x\ 100$$

$$Specificity(\%) = \frac{TN}{TN + FP}\ x\ 100$$

$$Positive\ predictive\ value\ (\%) = \frac{TP}{TP + FP}\ x\ 100$$

$$Negative\ predictive\ value\ (\%) = \frac{TN}{TN + FN}\ x\ 100$$

In this study, WDBC data were classified by using four different methods. The accuracy, sensitivity, specificity, positive predictive and negative predictive values are obtained. It has been observed that

GRNN has the highest learning performance. The results of these classifications are shown in Table 6 (training) and Table 7 (test).

Table 6. Classification results for training data.

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
GRNN	100%	100%	100%	100%	100%
RBF	98.07%	95.75%	99.44%	99.02%	97.53%
GEP	95.96%	92.92%	97.76%	96.10%	95.88%
SVM	98.77%	97.64%	99.44%	99.04%	98.61%
MLP	98.07%	96.23%	99.16%	98.55%	97.79%

Table 7. Classification results for test data by four methods

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
GRNN	98.77%	96.70%	100.00%	100.00%	98.08%
DTF	96.31%	95.75%	96.64%	94.42%	97.46%
RBF	96.13%	91.98%	98.60%	97.50%	95.39%
GEP	94.55%	91.51%	96.36%	93.72%	95.03%
SVM	97.19%	95. 57%	98.04%	96.67%	97.49%
MLP	97.54%	95.75%	98.60%	97.60%	97.51%

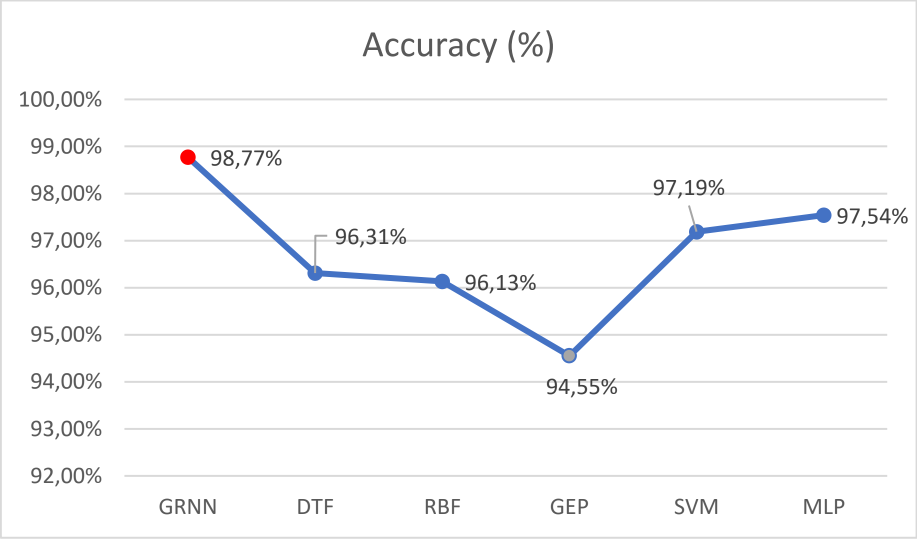

Figure 6. Comparison of the methods accuracies.

Among the applied methods, the GRNN estimator was validated and tested, and comparisons were made with experimental datasets. Graphical representation of percent error during GRNN validation in Figure 6 is made in comparison with other methods. It is clear from the figure that 1.3% error was found during maximum GRNN validation. It requires less training data compared to other ANN models. F-Measure, AUC are 0.98, 0.99 respectively. The GRNN approach can be used to predict the diagnosis of breast cancer with an accuracy of 98.77%. When all metrics are examined, it is seen that the model that provides the highest success is GRNN (in Table 8). This is followed by MLP and SVM respectively. It was observed that the lowest accuracy was obtained with the GEP model. It has been seen that the GEP algorithm has a lower performance than other algorithms. In addition, a detailed comparative performance analysis of the method used in the study with the studies in the literature was made. Experimental results show that the proposed method is effective and can be used by doctors for early diagnosis.

Table 8. GRNN methods classification result.

Actual Value	Predicted Value	
B	TN=357	FP=0
M	TP=205	FN=7

Conclusion and Discussion

Breast cancer is one of the leading causes of death among women worldwide, and the development of new approaches to screening, diagnosis and treatment of breast cancer has an important role in reducing the death rate. This article proposes a comparative approach for the prediction, diagnosis and classification of breast cancer using different machine learning algorithms. Using machine learning approaches, we observed that the 569 patients with benign and malignant tumors included in this study had a success rate of over 90% in all methods in predicting their condition.

In the literature, besides the studies in which the methods are used separately, there are many studies in which hybrid methods are recommended. Information on the results obtained in these studies is shown in Table 9.

The use of 569 pieces of data is the most important limiting factor of this study. For this reason, it would be beneficial to carry out the necessary studies to spread the study and to form an opinion by supporting it with similar datasets.

Although there are many studies in the literature with WBCD dataset, there are very limited studies using GRNN and 10-fold cross validation. This study is thought to be an example for future GRNN-based hybrid studies.

In future studies, the features of the cell image such as radius, texture, perimeter, area, compactness, concavity and concave points, which are among the features that make up the data set, can be examined more deeply and converted into tools for future clinics. Treatment decisions with machine learning methods in the diagnosis and treatment of breast cancer. When the results are evaluated, it is seen that the GRNN method is more successful than other methods in terms of accuracy on the data in question.

Table 9. Literature review.

Author	Method	Efficiency (%)
Sumbaly et al.(38)	J48	94,36%
Akdundogdu (22)	GEP	96,66%
Akay (32)	F-Score- SVM	99,51%
Yavuz and Eyupoglu (8)	GRNN	93,75%
Agarap (30)	MLP	99,03%
Hamsagayathri and Sampath (39)	Random Forest	96.66%
Kiyan and Yildirim(40)	RBF	96.18%
Elgedawy(41)	Random Forest	99.42 %
Chaurasia et al.(42)	NB	97.36%
Asri et al.(34)	SVM	97.13 %
Alzubaidi et al.	SVM	97.00%
Islam et al.(43)	SVM	97.00%
Chaurasia and Pal(44)	SVM	96.20%

In future studies, it is recommended to apply GRNN-based hybrid methods to increase this rate. First, a hybrid of the methods used in this study can be considered. In addition, the results obtained not only with the data, but also with the output of images taken from many biomedical devices can be supported (45). It is hoped that the integration of imaging methods will have a positive effect when examined in current studies. The effect on the performance of the methods can be analyzed by using different data preprocessing steps. In addition, the proposed method can be compared with different machine learning algorithms that were not used in the study, or studies can be carried out using deep learning methods.

References

1. Eyüpoğlu C, Yavuz E. Kanser Teşhisi için Makine Öğrenmesi Tekniklerine Dayalı Yeni Bir Sınıflandırma Metodu. BSEU Journal of Science Araştırma Makalesi-Research Article [Internet]. 2020 [cited 2022 Oct 25];7(2):1106–23. Available from: http://dergipark.gov.tr/bseufbd

2. Breast Cancer Awareness Month 2021 – IARC [Internet]. [cited 2022 Feb 14]. Available from: https://www.iarc.who.int/featured-news/breast-cancer-awareness-month-2021/

3. Açikgöz A, Akal YE.Etiology and risk factors of breast cancer. [cited 2022 Sep 19]; Available from: http://www.ncbi.nlm.nih.gov/

4. Bayraktar NG. Determination of women's awareness and practices on early diagnosis of breast cancer (Araştırma). Üniversitesi, Hacettepe Dergisi, Hemşirelik Fakültesi Güzel. 6(2):101–10.

5. Kumari M, Singh V. Breast cancer prediction system. Procedia Computer Science. 2018 Jan 1;132:371–6.

6. Levenson RM, Krupinski EA, Navarro VM, Wasserman EA. Pigeons (Columba livia) as trainable observers of pathology and radiology breast cancer images. Plos One. 2015 Nov 1;10(11).

7. Rani KU. Parallel approach for diagnosis of breast cancer using neural network technique. International Journal of Computer Applications. 2010 Sep 10;10(3):1–5.

8. Yavuz E, Eyupoglu C, Sanver U, Yazici R. An ensemble of neural networks for breast cancer diagnosis. 2nd International Conference on Computer Science and Engineering, UBMK 2017. 2017 Oct 31;538–43.

9. Mert A, Kiliç N, Bilgili E, Akan A. Breast cancer detection with reduced feature set. Computational and Mathematical Methods in Medicine [Internet]. 2015 [cited 2022 Mar 25];2015. Available from: /pmc/articles/PMC4452509/

10. Wang S, Wang Y, Wang D, Yin Y, Wang Y, Jin Y. An improved random forest-based rule extraction method for breast cancer diagnosis. Applied Soft Computing. 2020 Jan 1;86:105941.

11. Yue W, Wang Z, Chen H, Payne A, Liu X. Machine learning with applications in breast cancer diagnosis and prognosis. Designs 2018, Vol 2, Page 13 [Internet]. 2018 May 9 [cited 2022 Apr 28];2(2):13. Available from: https://www.mdpi.com/2411-9660/2/2/13/htm

12. Zor K, Çelik Ö, Timur O, Teke A. Short-term building electrical energy consumption forecasting by employing gene expression programming and GMDH networks. Energies. 2020 Mar 1;13(5).

13. Ferreira C, Leandro I, De Castro N, Zuben FJ Von. GEP and the evolutıon of computer programs, 2004;82–103.

14. Atrey K, Sharma Y, Bodhey NK, Singh BK. Breast cancer prediction using dominance-based feature filtering approach: a comparative ınvestigation in machine learning archetype. Brazilian Archives of Biology and Technology [Internet]. 2019 Nov 25 [cited 2022 Oct 25];62:1–15. Available from: http://www.scielo.br/j/babt/a/NyXbttjqTXKxpCpJqQN3fdr/?lang=en

15. Takcı,H., Centroid sınıflayıcılar yardımıyla meme kanseri teşhisi. Cumhuriyet Üniv Müh Mim Fak Der G 2016;2(2):323–30.

16. Kelek MM, Cengiz E, Oğuz Y, Yönetken A. RLBP Metodu ile Mamografi Görüntülerinin İncelenmesi ve Sınıflandırılması. Afyon Kocatepe Üniversitesi Uluslararası Mühendislik Teknolojileri ve Uygulamalı Bilimler Dergisi [Internet]. 2021 Dec 15 [cited 2022 Oct 26];4(2):59–64. Available from: https://dergipark.org.tr/tr/pub/akuumubd/issue/64247/978181

17. UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set [Internet]. [cited 2022 Feb 14]. Available from: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29

18. Salama G, Abdelhalim MB, Zeid MA. Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers. 2012;

19. Breast Cancer Wisconsin (Diagnostic) Data Set | Kaggle [Internet]. [cited 2022 Feb 17]. Available from: https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

20. Specht DF. A General Regression Neural Network. IEEE Transactions on Neural Networks. 1991;2(6):568–76.

21. Hong H, Zhang Z, Guo A, Shen L, Sun H, Liang Y, et al. Radial basis function artificial neural network (RBF ANN) as well as the hybrid method of RBF ANN and grey relational analysis able to well predict trihalomethanes levels in tap water. Journal of Hydrology. 2020 Dec 1;591:125574.

22. Akgündogdu A. Breast cancer classıfıcatıon wıth genetıc programmıng. International Journal Of Electronics, Mechanical and Mechatronics Engineering. 2:72–8.

23. Altun H, Eminoğlu U, Tezekici BS. MLP yapay sinir ağlarında öğrenme sürecinin aktivasyon fonksiyonu ve istatiksel değişim gösteren giriş verilerine bağımlılığı.

24. Baum EB. On the capabilities of multilayer perceptrons. Journal of Complexity. 1988 Sep 1;4(3):193–215.

25. Desai M, Shah M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). Clinical eHealth. 2021 Jan 1;4:1–11.

26. Boser BE, Guyon IM, Vapnik VN. Training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory [Internet]. New York, New York, USA: Publ by ACM; 1992 [cited 2021 Mar 6]. p. 144–52. Available from: http://portal.acm.org/citation.cfm?doid=130385.130401

27. Houssein EH, Emam MM, Ali AA, Suganthan PN. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications. 2021 Apr 1;167:114161.

28. Sevli O. Göğüs kanseri teşhisinde farklı makine öğrenmesi tekniklerinin performans karşılaştırması. European Journal of Science and Technology [Internet]. 2019 [cited 2022 Oct 25];(16):176–85. Available from: http://dergipark.gov.tr/ejosat

29. Saygılı A. Issue: 2 Date. Isvos Journal. 2018;2018(2):48–56.

30. Agarap AF. On Breast Cancer Detection: an application of machine learning algorithms on the Wisconsin diagnostic dataset. 2018 [cited 2022 Feb 8]; Available from: https://doi.org/10.1145/3184066.3184080

31. Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings. 2005 May 1;37(5):545–53.

32. Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Systems With Applications [Internet]. [cited 2022 Mar 25];36:3240–7. Available from: www.imaginis.com/breasthealth/breast_cancer.asp,

33. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proceedings of the National Academy of Sciences of the United States of America [Internet]. 2003 Sep 2 [cited 2022 Feb 15];100(18):10393–8. Available from: https://pubmed.ncbi.nlm.nih.gov/12917485/

34. Asri H, Mousannif H, Al Moatassime H, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science. 2016 Jan 1;83:1064–9.

35. Osuna E, Freund R, Girosi F. Training support vector machines: An application to face detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1997;130–6.

36. Pontil M, Verri A. Support vector machines for 3D object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998;20(6):637–46.

37. Schölkopf B, Sung KK, Burges CJC, Girosi F, Niyogi P, Poggio T, et al. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Transactions on Signal Processing. 1997;45(11):2758–65.

38. Sumbaly R, Vishnusri N, Jeyalatha S. Diagnosis of breast cancer using decision tree data mining technique. International Journal of Computer Applications. 2014 Jul 18;98(10):16–24.

39. Hamsagayathri P, Sampath P. Performance analysis of breast cancer classıfıcation using decision tree classifıers. International Journal of Current Pharmaceutical Research. 2017 Mar 1;9(2):19.

40. Kiyan T, Yildirim T. Breast Cancer Dıagnosıs Usıng Statıstıcal Neural Networks.

41. Elgedawy MN. Prediction of Breast Cancer using Random Forest, Support Vector Machines and Naïve Bayes. International Journal Of Engineering And Computer Science. 2017 Jan 1;

42. Chaurasia V, Pal S. A Novel Approach for Breast Cancer Detection Using Data Mining Techniques [Internet]. 2017 [cited 2022 Oct 25]. Available from: https://papers.ssrn.com/abstract=2994932

43. Islam MM, Iqbal H, Haque MR, Hasan MK. Prediction of breast cancer using support vector machine and K-Nearest neighbors. 5th IEEE Region 10 Humanitarian Technology Conference 2017, R10-HTC 2017. 2018 Feb 9;2018-January:226–9.

44. Chaurasia V, Pal S, Tiwari BB. Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms and Computational Technology [Internet]. 2018 Jun 1 [cited 2022 Oct 25];12(2):119–26. Available from: https://journals.sagepub.com/doi/10.1177/1748301818756225

45. Korkut O, Fen AÜ, Enstitüsü B, Kuşcu AC, Erol H, Tarihçesi M. Diagnosis of breast cancer by k-mean clustering and otsu thresholding segmentation methods. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi [Internet]. 2022 Mar 8 [cited 2022 May 13];5(1):258–81. Available from: https://dergipark.org.tr/tr/pub/okufbed/issue/68798/994481