Abstract: The early diagnosis of the breast cancer has become imperative in cancer research because it may facilitate the subsequent clinical treatment of patients. Separation of breast cancer patients into normal, low and high groups has become important in bioinformatics and biomedical fields. This has led to an increase in the practice of machine learning (ML) methods for early breast cancer diagnosis in the literature. Machine learning methods have been used to model the diagnosis and treatment of breast cancer. ML methods have been used to detect complex cell characteristics in breast cancer images. (ANN), Bayesian Networks (BNS), Random Forest (RF), Support Vector Machines (SVM), Decision Trees (DT), Linear Discriminant Analysis (LDA), Sammon mapping, Stochastic Neighbor New algorithms have been proposed using various machine learning techniques. Although machine learning techniques for breast cancer have been widely applied and ultimately yielded high classification performances, an appropriate level of validation is required to take these methods into account in daily clinical treatment and practice. In this study, the methods used in algorithms for early diagnosis of breast cancer and the classification ratios are described. In the advanced algorithms, various different features and image data are used. As a result, in this article, ML methods for breast cancer research are increasing. For this reason, published articles have been presented to model the risk of breast cancer.

Keywords: Breast Cancer, Machine Learning, Detection

Introduction
In the world, breast cancer is a rapidly developing disease that is often the cause of death and is more common in developed countries. Breast cancer is a type of cancer that develops rapidly in a very short time, starting from a cell of breast tissue. The most accurate and most important way of fighting with breast cancer is early diagnosis and diagnosis. The earlier the breast cancer can be diagnosed, the more chances the patient can heal. In many countries, the number of women who are sick due to breast cancer and die from late detection of breast cancer is increasing. Breast cancer is one of the most common types of cancer in women and is the second leading cause of death after lung cancer. Early detection of breast cancer reduces mortality by 40% or more (Jemal, 2011). 1 out of every 9 females can get breast cancer after middle age. When we look at the work done, 570,000 women are diagnosed with cancer every year on the earth. Breast cancer patients constitute 31% of this cancer rate. Of these breast cancer patients, 17% died. The frequency of breast cancer in North American and European countries is higher than in other parts of the world (Ferlay, 2007). The cause of breast cancer is still unknown. However, it is estimated that some of the factors are related to breast cancer. These factors include breed, breast tumor tumor, age, menses (menstruation) age, gestational age, menopause gray age, female sex, diet and horn. The risk of having breast cancer in people over 50 years old compared to other age groups is approaching 80%. If a mammal is cancerous, the risk of cancer is 5 times higher on the other mammal. Breast cancer has a slightly higher risk of developing breast cancer in women. The risk of developing breast cancer is 3-4 times higher than that seen before 12 years of age. For example, in a woman who entered the menopause at the age of 57, the risk of breast cancer is about 50% higher than that of a woman who entered the menopause at the age of 43. Do not give birth at the age of 30 or give birth at first, increase risk of breast cancer. 99% of breast cancers occur in males and females. Progesterone and estrogen hormone have been shown to increase the risk of breast cancer by an average of 36% (American Cancer Society, 2009). The risk of breast cancer increases if you are fed poor and overweight. Mothers who give milk to baby for less than 1 year carry more breast cancer risk. Protect your mother from breast cancer. The risk of breast cancer in women receiving radiotherapy in the chest region due to any disease increases the risk (American Cancer Society, 2009). Women in the postmenopausal period, hormone therapy containing estrogen for more than 5 years of continuous use of the breast cancer risk increases conditions. While obesity increases the risk of breast cancer after menopause, it does not have such a risk in the premenopausal period. In these people, even less breast cancer is seen. Another factor that increases breast cancer is that it is fed with foods rich in oil. Taking alcohol is not a proven finding today as it increases the risk of breast cancer. Regular sports have been reported to reduce the risk of breast cancer, especially in brisk walking (Karabulut, 2009). Breast cancer is caused by the cell renewal of the human body and the formation of cells that refuse to die. The cells that lose their lives and die must remain in human conscience without dying and soon begin to divide and eventually cover the entire tissue and thus make the human organism unusable. Breast cancer cells have the qualification to pass to other
organs and prevent other organs from functioning. If breast cancer is not detected early and is not treated, it can result in human death. Early diagnosis and treatment of breast cancer is crucial for reducing mortality rates. In order for the treatment of breast cancer to be easy, it needs to be diagnosed early. The early detection of breast cancer increases the chance of rescuing the patient from death. However, early detection of cancer may provide better treatment. Early detection, however, also requires accurate and reliable diagnosis (Turusbekova, 2012), which can differentiate between benign and malignant tumors. In many countries around the world, the Ministry of Health places emphasis on the early detection of breast cancer. If breast cancer is detected early before spreading, the patient has a 96% chance of survival. Every year in the world, 44,000 women die from breast cancer. Therefore, any breast mass that is noticed in the mammary should be evaluated in terms of breast cancer and the possibility of cancer should be determined definitively. This possibility may increase the chances of the patient having breast cancer during treatment (Aytaç Korkmaz, 2015). There are many studies on machine learning techniques on other types of cancer besides breast cancer in the literature (Korkmaz and Binol, 2017), (Korkmaz, S. A., Binol, H., Akçiçek, A., and Korkmaz, M. F. 2017), (Korkmaz, S. A., Akçiçek, A., Binol, H., and Korkmaz, M. F., 2017), (Korkmaz, 2018), (Korkmaz, S. A. and Esmeray, F. 2018), (Korkmaz, 2018). Machine learning techniques have also been used to classify more diverse data analyzes (KORKMAZ, 2017), (Korkmaz and Poyraz 2016). But, we will examine computer-aided studies of breast cancer in this article. In Literature, there are computer assisted studies in machine learning techniques in breast cancer diagnosis.

Related Work

Many studies in the literature have been reviewed for early diagnosis of breast cancer. According to these studies; In article (Korkmaz, 2015), light microscopy was used for early diagnosis of breast cancer. 180 images were obtained with light microscope. 23 features are calculated for 180 images. These images were rotated at 4 different angles and a 23x4 feature value was calculated for each 180 images. Minimum Redundancy Maximum Relavance method is used to reduce the size of the properties. Least Square Support Vector Machine method was used as classifier. Images were classified as normal, benign and malignant. Classification success rate is 100%.

In article (Korkmaz and Poyraz, 2014), normal, benign, and malignant histopathology cell images of breast cancer were used. In this article, these images were used to contribute to the early diagnosis of breast cancer. Discrete Wavelet Transform families were applied to breast cancer images. 16 Discrete Wavelet Transform families and 16 feature vectors are obtained. Jensen Shannon, Hellinger, Triangle classifiers were used as classifier. The average classification rate for Jensen Shannon classifier for 16 feature vectors was found to be 97.81%. It was found 97.75% with Hellinger Classifier. It was 97.87% with Triangle classifier. Then the averages of these classification ratios were taken. The classification success rate was 97.81% according to this average ratio.

In article (Korkmaz, S. A., Korkmaz, M. F., and Poyraz, 2016), both mammograms and histopathology images of the same patients were taken to be able to contribute to the early diagnosis of breast cancer. 150 mammograms were obtained from the radiology department of Fırat University Medical Faculty and 150 histopathology images from Fırat University Medical Faculty Pathology Department. These images have Gray level co-occurrence matrix (GLCM) properties. The dimensions of these features have been reduced by the method of minimum redundancy and maximum relevance. 10 properties were obtained. Suspicious probability values of selected properties were obtained with the help of exponential curve. These probabilistic values are used in weight calculations. Calculated weights are used in the Jensen Shannon, Triangle, and Hellinger classifiers. The classification ratios were analyzed by ROC curve. According to this; the highest classification success rate was found with the Jensen Shannon classifier. The accuracy rates obtained with the Jensen Shannon classifier are 99% for histopathology images and 98% for mammography images.

In article (Korkmaz, S. A and Korkmaz, M. F. 2015), mammographic images obtained from the Digital Database for Screening Mammography (DDSM) database were used to be useful for early detection of breast cancer. Some features of the 3x126 view from this database were found. The dimensions of these properties are reduced by the minimum-Redundancy-Maximal-Relevance m (RMR) method. These images were then classified as normal, benign, and malignant with the Kullback-Leibler (KL) classifier using these acquired features for the 3x126 image. An approximate performance estimate was made with the ROC curve. This performance estimate was found to be 98.3%.

In article (Korkmaz, S. A. and Eren, H. 2013), found probabilistic values of suspicious lesions with the aid of exponential curve fitting of 3x10-image from the DDSM database in order to benefit early detection of breast cancer. These probabilistic values were used to find the weight value for analysis with the Kullback Leibler
classifiers. After finding the weight values, the mammographic images were classified as normal, benign, and malignant with the kullback leibler classifier.

In article (Korkmaz, 2016), Atomic force microscopy images were used to identify early breast cancer lesions. Breast cancer pathology cells were visualized from atomic force microscope. For this process, atomic force microscopy was used in the nanotechnology laboratory in the physics department of Firat University. 23 Gray level co-occurrence matrix (GLCM) properties of nanobiomechanical images obtained in this article were obtained. These 23 property values are obtained for 4 different angle values of the images. Thus, a total of 92 feature values were obtained. 92 property values have been reduced to lower dimensions with Minimum Redundancy Maximum Relevance (MRMR) and Principal Component Analysis (PCA) methods. Classifiers such as Least Square Support Vector Machine (LSSVM), Maximums of Statistical Values (MSMMR), and fuzzy k-nearest neighbor (KNN) classifiers are used. The lowest classification rate was 75.56% with PCA_KNN. The highest classification rate is 100% with mRMR_LSSVM, mRMR_KNN methods.

In article (Korkmaz, 2015), developed a computer-aided study for the early detection of breast cancer. Both light microscope images and mammography images were used for this study. For this study, 23 feature values were obtained. By rotating these images at different angles, 92 feature values were obtained for both microscopy and mammography views. This work was carried out in 2 steps. In the first step, the 92 property values obtained are reduced to a smaller size with minimum redundancy and maximum relevance (mRMR). In the second step, these optimum properties are classified by Least Square Support Vector Machine (LSSVM) and fuzzy k-nearest neighbor (KNN). It has been suggested that when combined with mammography images and microscope images, that some patients have higher classification success rates when mammograms and microscope images are analyzed at the same time.

Wolfe examined the relationship between breast tissue and breast cancer and reported that the breast parenchyma was divided into 4 groups (Wolfe, J.N., 1976).

In 2006, Cheng, Shi, Min, Hu, Cai and Du using textual features for mammogram images, it was aimed to identify and classify the kits by rotating images at different angles using the gray level co-occurrence matrix. 100% accuracy was found with Roc curves (Cheng, H., Shi, H., Min, R., Hu, L., Cai, X., Du, H., 2006).

In another article, using the Weka program and Data Mining Methods, Breast Cancer Cells were estimated and diagnosed. In this study, decision tree algorithm is applied in malign and benign mass classification process. The C4.5 decision tree from data mining methods contributed to the early diagnosis of an important disease such as breast cancer with an accuracy of 97.43% in disease diagnosis and diagnosis (Danacı, M., Çelik, M., Akkaya, A.E., 2009).

In one article, the screening of breast cancer is addressed. For this purpose, since the support vector machine known in the world does not give good results, this probing adaptive support vector machine and the composition of Fuzzy C-means algorithm are used. In this study, the United States of Wisconsin State Breast Cancer Database was used. In this study, the accuracy of classification was analyzed according to sensitivity, specificity, positive and negative values. Classification accuracy was 99.87% (Palanivel J., Kumaravel N., 2011).

Zwiggelaar segments the mammograms according to their intensity using re-appearance matrices. For the density classification, the shape features of the density region, such as the size, have been used (Zwiggelaar, R., Blot, L., Raba, D. and Denton, E.R.E., 2003).

In another article, they applied the Fuzzy-Genetic Algorithm method for the diagnosis of breast cancer and the classification accuracy was 97.36% (Pena-Reyes C. A., and Sipper M., 1999).

In another article, artificial neural networks were used to diagnose breast cancer, with an accuracy of 98.10% (Setiono R., 2000).

In the article published in 2008, the diagnosis of the breast cancer cell was made only with images taken from the atomic force microscope. According to the results obtained, a malignant breast cell is softer than a benign breast cell, resulting in a benign cell being stiffer than a malignant cell (Li, Q.S., Lee, G.Y.H., Ong, C.N., Lim, C.T., 2008).
In another article published in 2010, biomechanical and biochemical properties of microspectroscopy and atomic force microscopy were investigated (Yangzhe, W., et al., 2010).


In another article, for mammography images, artificial neural networks were used for the early detection of breast cancer and the accuracy was found to be 98.10% (Setiono, R., 2000).

In another model, 98.53% accuracy was achieved using the support vector method (Polat, K., and Gunes, S., 2007).

In another article, 99.51% correct results were obtained using the support vector machine and feature selection technique (Akay, M.F., 2009).

In a study conducted in 2011, the problem of classification of breast masses was addressed and it was suggested that the method achieved 84.6% success rate (Fraschini M., 2011).

In another study, Neural Network Algortum was used for the breast mass classification process. According to the results obtained, the k-means method was 86.6%, the ANFIS-based SOFM algorithm was 86.2%, and the ANFIS-based learning method (BP) had 91% success rate (In-sung, J., Devinder, T., and Wang, G.N., 2009).

Mencattini used 13 Haralic tissue attributes extracted from images reinforced for classification due to the diagnosis of breast cancer (Mencattini, A., Rabottino, G., Salmeri, M., Caselli, F. and Lojacono, R., 2008).

Brake, Karssemeijer and Hendriks have proposed a method to distinguish malignant tissue from normal tissues in digital mammograms. In this article, a pixel-level method is used to identify spikes and masses. For each speckle there are 2 masses and 5 spe- cies related to 3 speckles (Brake, G.M., Karssemeijer, N. and Hendriks, J.H.C.L., 2000).

Sheshadri and Kandaswamy have attempted to determine whether the mammal is normal, benign, and malignant using mean, standard deviation, smoothness, third moment, uniformity, and entropy attributes (Sheshadri, H.S. and Kandaswamy, A., 2006).

Petroudi and Brady have attempted to differentiate breast types using frame-level matching in another study. Geometric and topographic structures of the images are investigated with this method (Petroudi, S., Kadir, T. and Brady, M., 2003).

Groshong and Kegelmayer used Hough transformations to find the masses in mammogram images (Groshong, B.R. and Kegelmeyer, W.P., 1976).

Li and colleagues used Markov Random fields for mass detection (Li, H.D., Kallergi, M., Clarke, L.P., Jain, V.K. and Clark, R.A., 1995).

Lefebvre et al. used the fractal approach in segmentation of microcalcifications (Lefebvre, F., Benali, H., Gilles, R., Kahn, E. and Di Paola, R., 1995).

Petrick and colleagues used contrast-enhanced filters based on density of images (Patrick, N., Chan, H.P., Sahiner, B. and Wei, D., 1996).

ErtAŞ and Gülçür have established a relative processing capability curve (ROC) to determine whether the whole mammographic image is intense, the Youden criterion is calculated and the optimum asymmetry value is found (ErtAŞ, G. ve Gülçür, H.Ö., 2001).

Zhang and colleagues used genetic algorithms to extract features (Zhang, P., Verma, B. and Kumar, K., 2005).

Arnoldi and colleagues conducted a screening of breast cancer cells from AFM images. Here, the elasticity and force diagrams of the cancerous cell were taken with AFM microscope, and the elasticity of the cancerous cell was found to be less than that of benign cells (Arnoldi, M., Kacher, C.M., Bauerlein, E., Radmacher, M., Fritz, M., 1998).

Petushi and colleagues, histopathology, using normal, benign, malign images, have used the microstructural features of these images. In the feature selection, Linear Discriminant Analysis (LDA) and Forward / Backward search methods are used. As a classifier, linear, quadratic, neural network, decision tree is used. The most accurate classification gave a quadratic classifier. The accuracy of the quadratic classifier was found to be 95.6% (Petushi S et al., 2006).

Conclusions

In the world, breast cancer is a rapidly developing disease that is often the cause of death and is more common in developed countries. Breast cancer is a type of cancer that develops rapidly in a very short time, starting from a cell of breast tissue. The most accurate and most important way of fighting with breast cancer is early diagnosis and diagnosis. The earlier the breast cancer can be diagnosed, the more chances the patient can heal. The early diagnosis of the breast cancer has become imperative in cancer research because it may facilitate the subsequent clinical treatment of patients. Separation of breast cancer patients into normal, low and high groups has become important in bioinformatics and biomedical fields. This has led to an increase in the practice of machine learning (ML) methods for early breast cancer diagnosis in the literature. Machine learning methods have been used to model the diagnosis and treatment of breast cancer. ML methods have been used to detect complex cell characteristics in breast cancer images. (ANN), Bayesian Networks (BNS), Random Forest (RF), Support Vector Machines (SVM), Decision Trees (DT), Linear Discriminant Analysis (LDA), Sammon mapping, Stochastic Neighbor New algorithms have been proposed using various machine learning techniques. Although machine learning techniques for breast cancer have been widely applied and ultimately yielded high classification performances, an appropriate level of validation is required to take these methods into account in daily clinical treatment and practice. In this study, the methods used in algorithms for early diagnosis of breast cancer and the classification ratios are described. In the advanced algorithms, various different features and image data are used. As a result, in this article, ML methods for breast cancer research are increasing. For this reason, published articles have been presented to model the risk of breast cancer.

References


