Early Breast Cancer Prediction Using Dermatoglyphics: Data Mining Pilot Study in a General Hospital in Iran

T M U

ARTICLE INFO

Authors
Ayyoubzadeh S.M.[1] *PhD,*
Almasizand A.[2] *BSc,*
Rostam Niakan Kalhori Sh.[1] *PhD,*
Baniasadi T.[3] *PhD,*
Abbasi S.*[2] *PhD*

ABSTRACT

Aims Dermatoglyphic is the study of skin patterns on hands and feet. It has been shown in some studies that specific finger patterns could be a risk factor for breast cancer. Thus, this study aims to evaluate fingerprint patterns and other easy-to-obtain features in the risk of breast cancer.

Instrument & Methods This descriptive study was conducted in 2020. A dataset containing 462 records included female patients in Imam Khomeini Hospital Complex, Tehran, Iran. The factors' weight was determined by the Information Gain index. Predictive models were built once without fingerprint features and once with fingerprint features using Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine, and Deep Learning classifiers. RapidMiner 9.7.1 Software was used.

Findings The most important factor determining breast cancer were age, having a child, menopause situation, and menopause age. The best performance was the Random Forest model with accuracy and Area under Curve of a Receiver operating characteristic of 84.43% and 0.923, respectively. The fingerprint patterns feature increased the RF accuracy from 79.44% to 84.43%.

Conclusion An early breast cancer screening model could be built with the use of data mining methods. The fingerprint patterns could increase the performance of these models. The Random Forest model could be used. The results of such models could be used in designing apps for self-screening breast cancer.

Keywords Data Mining; Risk Factor; Breast Cancer; Dermatoglyphics

[1]Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
[2]Department of Laboratory Science, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
[3]Department of Health Information Technology, Faculty of Para-Medicine, Hormozgan University of Medical Sciences, Bandar Abbas, Iran

Correspondence
Address: Department of Laboratory Science, 1st Floor, School of Allied Medical Sciences, Tehran University of Medical Sciences, No 17, Farredanesh Alley, Ghods St, Enghelab Ave, Tehran, Iran.
Phone: +98 (21) 88982886
Fax: +98 (21) 88983025
sakineh4612004@yahoo.com

CITATION LINKS

[1] Global cancer statistics 2018: Globocan ... [2] Estimating the global cancer incidence and ... [3] Association of quantitative ... [4] Early diagnosis of breast ... [5] Study of fingerprints pattern ... [6] Effect of three decades of ... [7] Mammography screening .... [8] Barriers related to ... [9] Barriers to mammography ... [10] Barriers to mammography ... [11] Breast cancer screening barriers ... [12] Perceived barriers to mammography ... [13] Perceived barriers to mammography ... [14] A case-control study of breast ... [15] Predictive analysis of palmar ... [16] Study of dermatoglyphic patterns ... [17] Qualitative and quantitative ... [18] Comparison of the dermatoglyphic ... [19] Dermatoglyphics: A plausible ... [20] International evaluation of an AI ... [21] Analysis of SEER dataset for breast ... [22] A GAs based approach for ... [23] Breast cancer diagnosis via data ... [24] A comparative study on breast ... [25] Data mining with decision ... [26] Predicting breast cancer survivability ... [27] Data mining techniques for predicting ... [28] Knowledge based analysis of ... [29] Irvine: UCI Machine Learning ... [30] Breast cancer data ... [31] Data mining: concepts and ... [32] Performance analysis of Naive Bayes and J48 ... [33] Analysis of various decision tree algorithms ... [34] Deep neural networks, gradient-boosted ... [35] Predicting the prolonged length of ... [36] Deep learning for healthcare: Review ... [37] Deep learning ... [38] A K-fold averaging cross-validation ... [39] Breast cancer recurrence prediction ... [40] Classification and prediction of breast ... [41] Random forest classifier combined ... [42] A deep learning mammography ... [43] A novel deep learning based framework ... [44] Predicting breast cancer by applying ... [45] Performance analysis of decision tree ... [46] Dermatoglyphics pattern on breast cancer patients ... [47] Predictive values of quantitative analysis of ... [48] Clinical significance of dermatoglyphics in patients ...

Introduction

Breast cancer is the most common type of cancer among women worldwide [1] and the main cause of their death [2]. In most cases, breast cancer is diagnosed too late [3]. Early diagnosis of breast cancer could prevent stage progress and the death of the patients [4].

After the formation of the tumor, breast cancer could be diagnosed through imaging techniques such as soft tissue mammography, breast ultrasound, and magnetic resonance imaging [5]. Techniques with the ability to predict the risk of breast cancer could be helpful for the prevention and successful treatment of this disease. Mammography is the gold standard of screening and could effectively detect cancer in this stage [6, 7]. It has barriers such as costs, fear, embarrassment, self-denial, lack of family support, issues in centers providing screening, culture, and religion [8-13].

The causes of breast cancer are a combination of genetic and environmental factors [3]. Thus, predicting breast cancer even before mass formation could be a solution by using data analysis from cancer-causing agents in other patients and creating models.

There are several factors affecting breast cancer addressed in the literature. These factors could be categorized in 1) Socio-demographic characteristics include age, ethnicity, and education 2) Clinical, exogenous hormonal, and menstrual and reproductive factors include breast surgery history, first-degree family history of breast cancer, hormonal replacement therapy history, menopausal status, breastfeeding history 3) Anthropometric and lifestyle factors include BMI, Alcohol intake, Soybean related products intake, and physical activity [14]. One of the less addressed factors in the literature that could be easily obtained is skin patterns of hands and feet.

Skin line patterns show congenital abnormalities and might be used for identifying genetic disorders [15] also. They could be helpful to identify women with breast cancer or at risk of breast cancer [16]. A family history of breast cancer may be associated with a specific pattern of fingerprints. However, skin line patterns are genetically controlled but may also be influenced by environmental factors in the first three months of pregnancy to reflect a person's genetics and susceptibility to breast cancer. The fingerprint can be an anatomical, noninvasive and cheap marker to detect susceptible individuals [5]. For women with breast cancer, compared to controls, most patients with breast cancer have a Whorl fingerprint pattern on their six or more fingers [5, 16, 17]. Also, some studies have not found a significant relationship between fingerprint patterns and breast cancer [18]. The fingerprint patterns could be categorized as Loop, Central Pocket Loop (Whorl), Lateral Pocket Loop (Whorl), Pain Arch, Tented Arch, True Whorl (Whorl), Accidental Loop (Whorl), Twinned Loop (Whorl), Double Loop (Whorl) [16]. In general, dermatoglyphics is defined as the study of skin patterns of hands and feet. It is affected by genetics; It could be a risk indicator for diseases [19]. Also, some studies showed they could be used to predict the risk of breast cancer [5, 16, 17].

Recently, Artificial Intelligence (AI) systems have attracted significant attention, especially in the healthcare field. A recent study showed that an AI system could predict breast cancer risk more accurately than average radiology experts [20]. Data mining as a subfield of Artificial Intelligence could be used for prediction purposes [21]; they deal with methods and tools discovering new and hidden knowledge from databases [22].

Due to the importance of breast cancer risk prognosis and detection in the early stages, many studies investigated data mining models in this regard. These studies used multiple data mining techniques, including decision trees, logistic regression, support vector machines, Naïve Bayes [21, 23-28]. Although several studies use AI systems, especially data mining methods, to evaluate the risk of breast cancer [23-25], few studies evaluate finger patterns with data mining for breast cancer risk prediction. Thus, this study aims to assess fingerprint patterns and other easy-to-obtain features in the risk of breast cancer.

Instrument & Methods

This descriptive-retrospective study was conducted in 2020 and included female patients in Imam Khomeini Hospital Complex, Tehran, Iran. This case-control study contains 308 controls and 154 cases described in the previous study [16]. The sample size was chosen based on previous similar datasets [29, 30]. The samples were randomly selected. Case group patients were included in the study in case of pathological diagnosis of breast cancer in the Cancer Institute (wards 1 and 3 of women and central clinics 1 and 2) of Imam Khomeini Hospital and the control group included healthy women with no history of breast cancer and other neoplastic diseases, and none of their relatives had a history of breast cancer. Women with hysterectomy and artificial menopause or exposed to radiation and chemotherapy during their lifetime were excluded from the study.

This study was approved by the Research Ethics Committee of Tehran University of Medical Sciences. Informed consent for the research was obtained from all participants (or their parents / legal guardians) before entering the study. After obtaining informed consent and permission, both patient and control groups were asked to complete a form of demographic information about age, ethnicity, marital status, type of parental marriage, family history of cancer, and smoking.

After placing the subject's fingers or thumbs in purple ink, the fingerprint was taken on white paper. Patterns of all ten fingers were analyzed. Fingerprints were classified into nine patterns including Loop, Central Pocket Loop (Whorl), Lateral Pocket Loop (Whorl), Pain Arch, Tented Arch, True Whorl (Whorl), Accidental Loop (Whorl), Twinned Loop (Whorl), Double Loop (Whorl) patterns. A dataset contains 462 records were created. This dataset includes 18 input features and one target feature indicating the presence of breast cancer. The features are represented in Table 1.

Table 1) Features of the extracted dataset

Feature name	Values
Age	Integer
Menstruation age	Integer
Menopause situation	0=Is not menopause 1=Is menopause 2=Is menopause due to chemotherapy 3= Is menopause due to hysterectomy
Menopause age	Integer
Has a child	0=Has not child 1=Has child
Age at first live birth	Integer
Family history	0=No family cancer 1=Her sister has cancer 2=Her mother has cancer 3=Her daughter has cancer
Finger 1 dermatoglyphic pattern	1=Loop
Finger 2 dermatoglyphic pattern	2=Central Pocket Loop (Whorl) 3=Lateral Pocket Loop (Whorl)
Finger 3 dermatoglyphic pattern	4=Pain Arch 5=Tented Arch
Finger 4 dermatoglyphic pattern	6=True Whorl (Whorl) 7=Accidental Loop (Whorl)
Finger 5 dermatoglyphic pattern	8=Twinned Loop (Whorl) 9=Double Loop (Whorl)
Finger 6 dermatoglyphic pattern	
Finger 7 dermatoglyphic pattern	
Finger 8 dermatoglyphic pattern	
Finger 9 dermatoglyphic pattern	
Finger 10 dermatoglyphic pattern	
Has Breast Cancer (Label)	1=Has Cancer 2=Has not cancer

The value transformation was done from numerical to categorical values for menopause situation, has Child, operation history, family history, finger dermatoglyphic patterns, and has breast cancer features. Missing values were imputed using K-NN (K-Nearest neighbor) method. The age, Menstruation age, first live birth age values are normalized.

Information gain is a calculated based on Shannon's information theory [31] Naïve Bayes classifier uses Bayes' theorem with the assumption of features independence. It predicts the target feature by calculating the most probable value with given input features [32].

A decision tree (DT) is an easy-to-interpret and important classifier in medical research. The tree is shown in a flowchart-like format that each node except leaves indicates a test on a specific input feature. Each branch that exits from the node shows the possible values of the feature. At the bottom of the tree, the leaves show the target feature (class label) with that path in the tree. The decision tree classifier approach is converting the sample space to the highest purity in the leaves. Algorithms such as CHAID, CART, ID3, C4.5 could be used to build this classifier [33].

Random Forest is an ensemble classifier. This classifier is comprised of multiple decision trees built on a subset of features. This model has two main hyperparameters: features count in the subset and the decision tree count in the forest [34, 35].

In the SVM classifier, boundary samples that could separate the samples classes' space, namely support vectors, are determined, and an optimal linear decision boundary is assumed between these boundaries. This boundary separates the samples' space with a hypersurface with a maximum margin [31].

In contrast with traditional data mining classifiers, deep learning can automatically extract features from raw data datasets by using multiple layers [36, 37]. Deep learning could be seen as a subset of artificial neural networks [37].

As the previous study shows, the whorl pattern count could be a predictor factor in breast cancer prediction [16]. The models were built triple times; once without finger pattern features and once with finger pattern features, and again with calculating and adding whorl count feature to the dataset.

The performances of the models are measured by calculating the following indices: Accuracy (Equation 1), sensitivity (Equation 2), specificity (Equation 3), and The Area under the Curve of a Receiver operating characteristic (AUC). These indices are measured with the use of test data categorization by a classifier. These categories for a binominal classifier are True Positives (TP), the number of positive instances that the classifier predicts to be positive; False Positives (FP), the number of instances that are not positive but the classifier predicts them to be positive; False Negative (FN), the number of positive instances but the classifier predicts them to be negative; and True Negative (TN), the number of negative instances and the classifier predicts them as negative. The evaluation of these models is calculated using the K-fold cross-validation method (K=10). Cross-validation is a method widely used to facilitate model estimation and variable selection [38].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \qquad (2)$$

$$Specificity = \frac{TN}{FP+TN} \qquad (3)$$

RapidMiner 9.7.1 Software was used to build and evaluate models. Naïve Bayes, Decision Tree, Random Forest, and Support Vector Machine classifiers are built once without fingerprint features and one with fingerprint features.

Findings

The most important factors determining breast cancer were age, pregnancy experience (or has a child), menopause situation, and menopause age (Diagram 1).
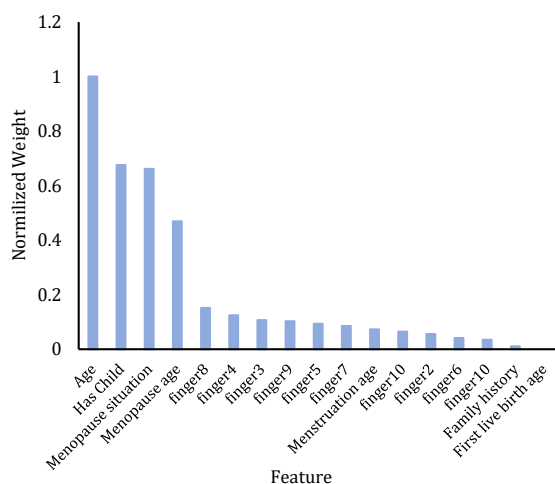

Figure 1) Factors' weight affecting breast cancer classification by Information Gain

The best AUC in the first experiment belongs to the deep learning model, and the best AUC in two other experiments belongs to the Random Forest classifier (Diagram 2; Table 2). The fingerprint patterns feature increased the RF accuracy from 79.44% to 84.43%. Also, the highest accuracy and sensitivity belong to this classifier in these two experiments. The Naïve Bayes classifier shows the best specificity (82.85%).
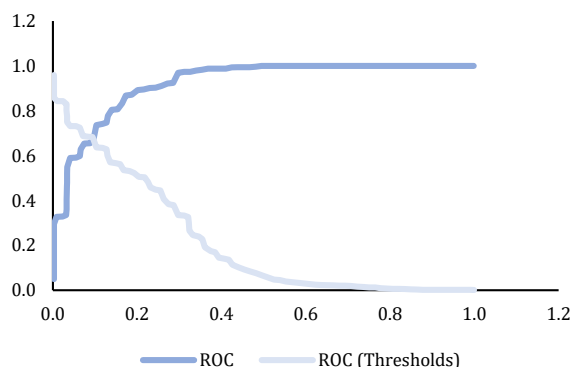

Figure 2) Random Forest classifier ROC curve (with Whorl count feature)

Table 2) results of comparison of classifiers' performance by the effect of finger pattern related features (numbers are in percentage)

Model	Accuracy	Sensitivity	Specificity	AUC
Without finger pattern				
Naïve Bayes	80.75±6.46	80.46±8.84	80.87±7.00	0.876±0.076
Decision Tree	78.35±5.74	80.67±16.26	77.26±6.40	0.800±0.091
Deep Learning	79.87±4.91	90.96±9.70	74.29±8.97	0.907±0.039
Random Forest	79.44±6.73	88.96±8.02	74.63±9.25	0.896±0.049
Support Vector Machine	77.27±6.19	88.96±9.17	71.42±6.85	0.885±0.052
With finger pattern				
Naïve Bayes	81.80±4.42	77.71±12.45	83.74±4.16	0.868±0.042
Decision Tree	79.65±3.73	85.79±7.15	76.66±6.31	0.838±0.040
Deep Learning	81.60±5.16	83.79±8.36	80.51±7.04	0.896±0.054
Random Forest	83.56±5.46	87.00±6.11	81.76±7.23	0.921±0.042
Support Vector Machine	77.72±6.08	80.54±9.59	76.32±7.42	0.872±0.045
With whorl pattern count feature				
Naïve Bayes	80.56±6.13	75.96±6.01	82.85±6.73	0.874±0.056
Decision Tree	77.95±7.35	85.79±8.98	74.09±12.23	0.835±0.058
Deep Learning	79.66±6.70	79.79±13.73	79.47±12.83	0.907±0.044
Random Forest	84.43±4.73	88.92±9.73	82.13±5.58	0.923±0.032
Support Vector Machine	80.93±5.07	84.38±7.04	79.25±6.42	0.878±0.035

Discussion

This study used data mining techniques to predict the early risk of breast cancer in Iranian women. This prediction is made by using easy-to-obtain features, and thus, it could be performed anywhere and inexpensively. The most important factors predicting breast cancer in our dataset were respectively: age, pregnancy experience, menopause situation, menopause age.

Some literature that used data mining methods for predicting breast cancer are reviewed in this section:

Senturk *et al.* used seven different algorithms that were used for the prediction of other patients. Discriminant analysis, multi-layer perceptron, decision trees, logistic regression, support vector machines, Naïve Bayes, K-NN are seven methods for contributing to the early diagnosis of breast cancer. In summary, the best algorithm for breast cancer prediction is the Support Vector Machines algorithm [23].

Padmavathi used RBFNN, MLP, and Logistic regression methods to study the performance of a neural network using breast cancer data. The accuracy of RBFNN was 97.0%, MLP was 91.3%, and logistic regression was 73.7%. The specificity of them was 96.8%, 91.1%, and 72.6%, respectively. Also the sensitivity were 97.3% , 92.1% and 75.5% respectively [24].

Rajesh & Anand have attempted to classify breast cancer data using the C4.5 algorithm to diagnose cancer and distinguish between the different stages. Five hundred random records were used. They found an accuracy of ~94% in the training phase and ~93% in the testing phase [21].

Delen *et al.* compared three methods to find a better prediction model for breast cancer survivability, and the dataset was more than 200,000 cases. The results indicated that the decision tree (C5) is the best predictor with 93.6% accuracy on the holdout

sample. 90.66% and 96.02% were shown to be specificity and the sensitivity of the model, respectively. Artificial neural networks came out to be the second high-performance classifier with 91.2% accuracy, 87.48% specificity, and 94.37% sensitivity, and the logistic regression models came out to be the next one with 89.2% accuracy 87.86% specificity, and 90.17% sensitivity [26].

Alshammari *et al.* compared C4.5, Naïve Bayes, Multilayer Perceptron, and RBF Network for predicting breast cancer survivability. They revealed that the decision tree (C4.5) algorithm with an accuracy of 89.3%, the sensitivity of 89.1%, and specificity of 98.5% is showed the best performance among the others, and in this study, high accuracy was very important [27].

In another study by Kuo *et al.*, the decision tree model was used to increase the level of diagnostic confidence to classify breast tumors and has been shown to perform with 96% accuracy, 96.67% specificity, and 93.33% sensitivity [25].

Aruna *et al.* compared data mining methods to find the best classification according to accuracy, sensitivity, and specificity in breast cancer diagnosis. They used methods such as Naïve Bayes, SVM-RBF kernel, RBF neural networks, Decision trees (J48), and simple CART; it was found that the SVM-RBF kernel is a good classifier for this task [28].

The Random Forest model showed the best performance with an accuracy of 84.43% and a sensitivity of 88.92%. This model showed that it might be useful for breast cancer screening purposes. Random Forest showed to be useful in other studies as well [39-41]. For instance, Nguyen *et al.* [41] showed that a method using Random Forest and feature selection technique could have more than 99% accuracy. Comparing the performance of other studies' proposed models with this study could be misleading; most of the mentioned studies have employed a public dataset (Wisconsin breast cancer dataset [29]) and evaluate their models on that dataset. In this public dataset, the features are based on breast and breast mass features, while in this study, we didn't use any features directly pertaining to the breast.

The deep learning method didn't show a great performance compared to other studies [42-44]. This issue could be justified for three reasons 1) as the studies mentioned, the deep learning techniques should be applied to raw data (e.g., images) 2) deep learning is suitable when the dataset size contains many records. The traditional classifiers' performance with a small training dataset could be better compared to deep learning methods. 3) As mentioned earlier, the classifiers are applied to breast images and contains direct information about the breast that could lead to imaging costs and other barriers mentioned in the introduction section.

The Decision Tree classifier has shown a great performance (from 96% to 99%) [45]. This study uses relatively easy-to-obtain features, including gender, age, benign or malignant, past disease history, breast change, occupation, vegetarianism, height, and weight. A reason that might affect the performance not to be generalized is the small sample size (220 records) used in this research.

We experiment with the use of finger pattern-related features in the classifier's performance. As a result, it showed the finger pattern features could enhance the performance of the classifiers. And defining the whorl count in the features could slightly improve the performance of the classifiers again.

Some previous studies suggest that Whorl and loop finger skin patterns could risk breast cancer. Sani *et al.*, research on 100 women with breast cancer in Dhrmais Cancer Hospital, India, showed that most women with breast cancer had Whorl finger skin patterns [46]. In another study in Bosnian-Herzegovinian on 100 women with breast cancer, the authors showed that the loop patterns count is related to breast cancer [47]. Manhas *et al.*, in their review, concluded that the breast could be a low-cost factor for predicting breast cancer [48] that is aligned with the findings of this study.

It is recommended to conduct similar studies in a larger sample size and use deep learning methods in order to automatically detect skin patterns.

Conclusion

An early breast cancer screening model could be built with the use of data mining methods. The fingerprint patterns could increase the performance of these models. The Random Forest model could be used. The results of such models could be used in designing apps for self-screening breast cancer.

References

1- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.

2- Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin D, Pineros M, et al. Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods. Int J Cancer. 2019;144(8):1941-53.

3- Prathap L. Association of quantitative and qualitative dermatoglyphic variable and DNA polymorphism in female breast cancer population. Online J Health Allied Sci. 2017;16(2):1-5.

4- Wang L. Early diagnosis of breast cancer. Sensors. 2017;17(7):1572.

5- Haggag AA Khalil AMI, Elzahed ES, Abdelhamid MI. Study of fingerprints pattern in breast cancer patients insharkia governorate, a case–control retrospective clinical study. Zagazig Univ Med J. 2018;24(1):66-71.

6- Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med. 2012;367(21):1998-2005.

7- Autier P, Boniol M. Mammography screening: A major issue in medicine. Eur J Cancer. 2018;90:34-62.

8- Alexandraki I, Mooradian AD. Barriers related to mammography use for breast cancer screening among minority women. J Natl Med Assoc. 2010;102(3):206-18.

9- Stoll CRT, Roberts S, Cheng MR, Crayton EV, Jackson S, Politi MC. Barriers to mammography among inadequately screened women. Health Educ Behav. 2015;42(1):8-15.

10- Hassoun Y, Dbouk H, Saad Aldin E, Nasser Z, Abou abbas L, Nahleh Z, et al. Barriers to mammography screening: how to overcome them. Middle East J Cancer. 2015;6(4):243-51.

11- Azami-Aghdash S, Ghojazadeh M, Gareh Sheyklo S, Daemi A, Kolahdouzan K, Mohseni M, et al. Breast cancer screening barriers from the womans perspective: A meta-synthesis. Asian Pac J Cancer Prev. 2015;16(8):3463-71.

12- Shirzadi S, Allahverdipour H, Sharma M, Hasankhani H. Perceived barriers to mammography adoption among women in iran: A qualitative study. Korean J Acad Fam Med. 2020;41(1):20-7.

13- Fayanju OM, Kraenzle S, Drake BF, Oka M, Goodman MS. Perceived barriers to mammography among underserved women in a breast health center outreach program. Am J Surg. 2014;208(3):425-34.

14- Tan MM, Ho WK, Yoon SY, Mariapun S, Hasan SN, Lee DS, et al. A case-control study of breast cancer risk factors in 7,663 women in Malaysia. Plos One. 2018;13(9):0203469.

15- Metovic A, Musanovic J, Alicelebic S, Pepic E, Sljuka S, Mulic M. Predictive analysis of palmar dermatoglyphics in patients with breast cancer for small Bosnian-Herzegovinian population. Med Arch. 2018;72(5):357-61.

16- Abbasi S, Einollahi N, Dashti N, Vaez-Zadeh F. Study of dermatoglyphic patterns of hands in women with breast cancer. Pak J Medical Sci. 2006;22(1):18-22.

17- Chintamani, Khandelwal R, Mittal A, Saijanani S, Tuteja A, Bansal A, et al. Qualitative and quantitative dermatoglyphic traits in patients with breast cancer: A prospective clinical study. BMC Cancer. 2007;7:44.

18- Sariri E, Kashanian M, Vahdat M, Yari S. Comparison of the dermatoglyphic characteristics of women with and without breast cancer. Eur J Obstet Gynecol Reprod Biol. 2012;160(2):201-4.

19- Mukherjee A, Madhushree KN, Nawab A, Ravindra S, Vivekananda MR, Shivaprasad D. Dermatoglyphics: A plausible risk indicator. Indian J Forensic Odontol. 2018;11(1):9-12.

20- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577:89-94.

21- Rajesh K, Anand S. Analysis of SEER dataset for breast cancer diagnosis using C4. 5 classification algorithm. Int J Adv Res Comput Commun Eng. 2012;1(2):72-7.

22- Chen TC, Hsu TC. A GAs based approach for mining breast cancer pattern. Expert Syst Appl. 2006;30(4):674-81.

23- Senturk ZK, Kara R. Breast cancer diagnosis via data mining: Performance analysis of seven different algorithms. Comput Sci Eng Int J. 2014;4(1):35-46.

24- Padmavati J. A comparative study on breast cancer prediction using RBF and MLP. Int J Sci Eng Res. 2011;2(1):1-5.

25- Kuo WJ, Chang RF, Chen DR, Lee CC. Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. Breast Cancer Res Treat. 2001;66(1):51-7.

26- Delen D, Walker G, Kadam A. Predicting breast cancer survivability: A comparison of three data mining methods. Artif Intell Med. 2005;34(2):113-27.

27- Alshammari SM, Shah TM, Huang Y. Data mining techniques for predicting breast cancer survivability among women in the United States [Internet]. Denton: University of North Texas Libraries; 2014 [cited 2021 Aug 21]. Available from: https://digital.library.unt.edu /ark:/67531/metadc277303/.

28- Aruna S, Rajagopalan S, Nandakishore L. Knowledge based analysis of various statistical tools in detecting breast cancer. Comput Sci Inf Technol. 2011;2:37-45.

29- archive.ics.uci.edu [Internet]. Irvine: UCI Machine Learning Repository; 2017 [cited 2021 Aug 21]. Available from: https://archive.ics.uci.edu/ml/index.php. https://www.sciencedirect.com/science/article/pii/B978 0123814791000083?via%3Dihub

30- Zwitte M, Soklic M. Breast cancer data set [Internet]. Irvine: UCI Machine Learning Repository; 1988 [cited 2021 Aug 21]. Available from: https://archive.ics.uci.edu/ml/datasets/breast+cancer.

31- Han J, Pei J, Kamber M. Data mining: concepts and techniques. 3rd Edition. Amsterdam: Elsevier; 2011.

32- Patil TR, Sherekar SS. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int J Comput Sci Appl. 2013;6(2):256-61.

33- Gupta B, Rawat A, Jain A, Arora A, Dhami N. Analysis of various decision tree algorithms for classification in data mining. Int J Comput Appl. 2017;163(8):15-9.

34- Krauss C, Do XA, Huck N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. Eur J Oper Res. 2017;259(2):689-702.

35- Chuang MT, Hu Yh, Lo CL. Predicting the prolonged length of stay of general surgery patients: A supervised learning approach. Int Trans Oper Res. 2018;25(1):75-90.

36- Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: Review, opportunities and challenges. Brief Bioinform. 2018;19(6):1236-46.

37- Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-44.

38- Jung Y, Hu J. A K-fold averaging cross-validation procedure. J Nonparametr Stat. 2015;27(2):167-79.

39- Al-Quraishi T, Abawajy JH, Chowdhury MU, Rajasegarar S, Abdalrada AS. Breast cancer recurrence prediction using random forest model. In: Ghazali R, Deris M, Nawi N, Abawajy J. Advances in intelligent systems and computing. Cham: Springer; 2018.

40- Murugan S, Kumar BM, Amudha S, editors. Classification and prediction of breast cancer using linear regression, decision tree and random forest. 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC); 8-9 Sept 2017, Mysore, India. Picataway: IEEE; 2017.

41- Nguyen C, Wang Y, Nguyen HN. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. J Biomed Sci Eng. 2013;6(5):551-60.

42- Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology. 2019;292(1):60-6.

43- Khan S, Islam N, Jan Z, Ud Din I, Rodrigues JJPC. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognit Lett. 2019;125:1-6.

44- Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. Radiology. 2019;292(2):331-42.

45- Venkatesan E, Velmurugan T. Performance analysis of decision tree algorithms for breast cancer classification. Indian J Sci Technol. 2015;8(29):1-8.

46- Sani FQ, Mirfat M, Iskandar I. Dermatoglyphics pattern on breast cancer patients in dharmais cancer hospital. Glob Med Health Commun. 2020;8(1):47-52.

47- Musanovic J, Metovic A, Pepic E, Kapic D, Cosovic E, Rebic D, et al. Predictive values of quantitative analysis of finger and palmar dermatoglyphics in patients with breast cancer for Bosnian-Herzegovinian population. J Evolution Med Dent Sci. 2018;7(24):2855-60.

48- Manhas S, Singh U. Clinical significance of dermatoglyphics in patients with breast cancer-A review article. Int J Inf Comput Sci. 2018;5(8):152-6.