Research Article

New Approach for Risk Estimation Algorithms of *BRCA1/2* Negativeness Detection with Modelling Supervised Machine Learning Techniques

Hulya Yazici [ID],[1] **Demet Akdeniz Odemis** [ID],[1] **Dogukan Aksu** [ID],[2] **Ozge Sukruoglu Erdogan** [ID],[1] **Seref Bugra Tuncer** [ID],[1] **Mukaddes Avsar** [ID],[1] **Seda Kilic** [ID],[1] **Gozde Kuru Turkcan** [ID],[1] **Betul Celik** [ID],[1] and **Muhammed Ali Aydin** [ID][2]

[1]*Istanbul University, Oncology Institute, Department of Basic Oncology, Division of Cancer Genetics, 34093 Fatih, Istanbul, Turkey*
[2]*Istanbul University-Cerrahpasa, Engineering Faculty, Computer Engineering Department, 34320 Avcilar, Istanbul, Turkey*

Correspondence should be addressed to Hulya Yazici; hulyayazici67@gmail.com

BRCA1/2 gene testing is a difficult, expensive, and time-consuming test which requires excessive work load. The identification of the *BRCA1/2* gene mutations is significantly important in the selection of treatment and the risk of secondary cancer. We aimed to develop an algorithm considering all the clinical, demographic, and genetic features of patients for identifying the *BRCA1/2* negativity in the present study. An experimental dataset was created with the collection of the all clinical, demographic, and genetic features of breast cancer patients for 20 years. This dataset consisted of 125 features of 2070 high-risk breast cancer patients. All data were numeralized and normalized for detection of the *BRCA1/2* negativity in the machine learning algorithm. The performance of the algorithm was identified by studying the machine learning model with the test data. k nearest neighbours (KNN) and decision tree (DT) accuracy rates of 9 features involving Dataset 2 were found to be the most effective. The removal of the unnecessary data in the dataset by reducing the number of features was shown to increase the accuracy rate of algorithm compared with the DT. *BRCA1/2* negativity was identified without performing the *BRCA1/2* gene test with 92.88% accuracy within minutes in high-risk breast cancer patients with this algorithm, and the test associated result waiting stress, time, and money loss were prevented. That algorithm is suggested be useful in fast performing of the treatment plans of patients and accurately in addition to speeding up the clinical practice.

1. Introduction

Machine learning is a computer-based predictive method or an estimation algorithm which makes an assumption of a hypothesis and uses this assumption for the estimation of the unknown condition using various mathematical and statistical methods. Thanks to this association between mathematics and computer science, the difficulties on creating and calculating the statistical models from large datasets including billions of data are resolved [1]. Machine learning is accepted as a branch of artificial intelligence owing to obtaining significant patterns from the samples such as human intelligence. Recently, machines were found to learn the duties that are regarded highly complex, and the

computer-aided diagnostic and decision support systems of machine learning algorithms were demonstrated to be ultimately useful. After discovering this, the question of how machine learning will enter in the practice particularly in medicine comes to mind. Up to present, various biological mechanisms were revealed using the conventional statistical modelling and analysis methods [2–8]. The biological and medical data produced in the area of biology and medicine have become more heterogeneous and complex with the rapid development of highly productive technologies in recent years, and the evaluation of these data becomes highly difficult with the known statistical analyses. More different methods such as advanced machine learning algorithms are required considering this rapid development in biology and

medicine and with the increase of the size of these data. Therefore, there is a need for the development of more effective and productive approaches such as advanced machine learning algorithms and network-based analyses for analysing such tremendous data. The investigation of the literature studies showed that machine learning is successfully practiced in resolving numerous problems such as classification, regression, and clustering [9, 10].

The studies on the use of machine learning algorithms in medicine are particularly based on the evaluations associated with the medical screening systems [11, 12]. These algorithms are used for various difficult tasks such as pulmonary embolism segmentation [13, 14] by performing angiography with computed tomography (BT), detection of polyp in colon cancer using virtual colonoscopy or CT [15, 16], detection of breast cancer diagnosis using mammography [17], brain tumor segmentation using magnetic resonance imaging (MRI) [18], and detection of the cognitive condition of brain (e.g., Alzheimer's disease) [19–21] using functional MR imaging.

The cause of the breast cancer is still unknown, and there is no available efficient prevention method against breast cancer. Therefore, early diagnosis or the identification of the present risk are the most important steps in treatment, and in the prevention of breast cancer [22]. Important steps have been put forward with the developments of effective diagnostic techniques and treatment methodologies in breast cancer in recent years [23]. Breast cancer results in the death of many people each year and constitutes a significant health problem worldwide. Breast cancer is the second most prevalent cancer after lung cancer and is the main reason of mortality in women worldwide [24]. In accordance with the Globocan 2018 data, the most prevalent cancer among women and men with a rate of 10.6% was breast cancer after lung cancer [25]. Approximately 80% of hereditary breast cancer cases develop due to the mutations on BRCA1/2 genes [26]. BRCA1/2 genes are tumor suppressing genes and have a role in response to cellular stress through the activation of the DNA repair processes. The pathogenic mutations in BRCA1/2 genes show that breast cancer has hereditary characteristic and increases the risk of getting breast cancer diagnosis [27]. Although approximately 12% of women in general population develop breast cancer all through their life, 72% of women with pathogenic BRCA1 mutation and 69% of women with pathogenic BRCA2 mutation are diagnosed with breast cancer at a stage of life. In addition, researchers reported that approximately 40% of women with pathogenic BRCA1 mutation and approximately 26% of women with pathogenic BRCA2 mutation develop secondary breast cancer in the contralateral 20 years after the first diagnosis of breast cancer [28]. Therefore, the identification of the BRCA1/2 gene mutations, the fast practice, and concluding of the gene analysis are highly important in planning the treatment and identifying the secondary cancer risk. The completion of the study of this test which is difficult, expensive, and long-lasting and therefore stressful for patients is significantly important in the rapid selection of the treatment protocols and rapid change of treatment when required.

There is no algorithm which identifies the BRCA1/2 negativity in the literature; in addition, there is no available algorithm that describes the BRCA1/2 negativity by only evaluating the clinical and demographic data of patients without performing the BRCA1/2 gene test in high-risk breast cancer patients with a family history. There is a need for algorithms which identify the BRCA1/2 gene mutation negativity in a short time and enabling the planning treatment rapidly by reducing the health costs and workload of cancer genetics polyclinic and laboratory. Therefore, an algorithm which identifies the individuals with negative BRCA1/2 gene mutation in short time was developed in the study by considering all the clinical, demographic, and genetic features of the total of 2070 patients.

2. Materials and Methods

The study flow was arranged as follows in the study conducted to develop an algorithm in order to determine the negative cases for BRCA1/2 mutation without having a BRCA1/2 gene test. The experimental dataset with 2070 individuals were created and pre-processing step was started with the feature selection. Datasets obtained in the feature selection were independently performed in pre-processing procedure, and were normalized and numeralized for developing classification models. The training and testing dataset were generated for the supervised machine learning models for detection of the BRCA1/2 negativity. All the work flow diagram for machine learning algorithm is presented in Figure 1 and all steps are detailed below.

2.1. Creation of Experimental Dataset. A dataset, including all the clinical, demographic, and genetic features of 2070 breast cancer patients who presented to the Cancer Genetics polyclinic in the Cancer Genetics Division of the Oncology Institute in Istanbul University for BRCA1/2 genetic testing between 1999 and 2019, was created. Some part of this raw data is presented in Table 1. The study was approved by the Local and Clinical research Ethics Committee of Istanbul University according to the tenets of the Declaration of Helsinki. All patients were informed about the study, and the consents were granted in the scope of the cancer genetics polyclinic. The clinical parameters which were particularly important for breast cancer were identified in the study. These parameters were classified as the modifiable and nonmodifiable parameters. A data classification technique which might predict the hereditary breast cancer was used with the use of a function that included the modifiable and nonmodifiable factors in the study. The selected nonmodifiable factors were menopause, the result of the BRCA1/2 genes, family history, and receptor condition; however, the modifiable factors were the use of oral contraceptive, alcohol consumption, hormone therapy, breast feeding, and smoking.

2.2. Preprocessing of Dataset: Feature Selection. Feature selection is a method which may be used for removing the unnecessary, irrelevant, and invalid features which have no contribution to the accuracy of a predictive model. The raw dataset of 2070 high-risk breast cancer patients included
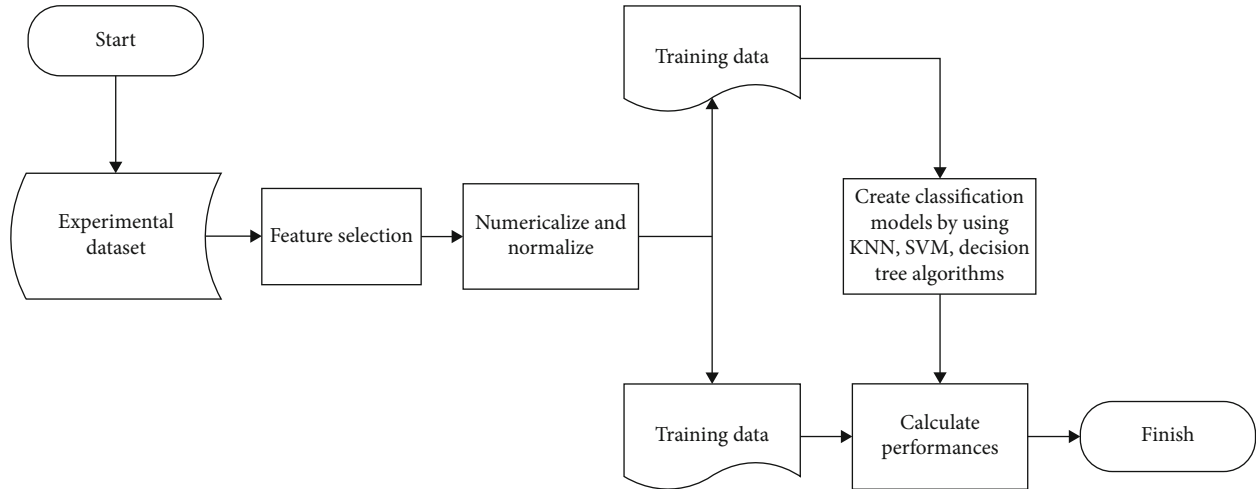
FIGURE 1: Flowchart of the proposed method.

TABLE 1: A part of raw dataset.

Data number	Diagnosis	Age	FIB	. . .	FIO	FIO > 40	BRCA
n0	1	38	0	. . .	0	0	0
n1	5	43	0	. . .	0	0	0
.	.	.	.	. . . .	.	.	.
.	.	.	.	. . . .	.	.	.
n2068	2	48	1	. . .	0	0	1
n2069	2	43	1	. . .	0	0	1

125 modifiable and unmodifiable parameters. First, the raw dataset was performed feature selection because there were numerous variables which did not affect the *BRCA1/2* negativity or there were unnecessary, irrelevant, and invalid features in the dataset. Then, the raw dataset consisting of 125 features was evaluated using Pearson's correlation method using the Statistical Package for the Social Sciences (SPSS) v.21, and statistically significant ($p < 0.05$) features were selected. After the feature selection stage, two different datasets were generated for 1460 and 1532 breast cancer patients and their 26 and 9 variables. The patient and feature data of the datasets are given in Table 2.

After Pearson's correlation analysis, the statistically most significant 9 features and their associations in the raw dataset are given in Figure 2. As shown in Figure 2, 0.75-1 was detected between FIO and FIO > 40, however, 0.50-0.75 between THB and THB < 40, 0.25-0.50 between the age and FIB, and -0.25-0.25 between other 5 variables.

2.3. Normalization and Numeralization of Dataset. Two different datasets obtained in the feature selection procedure were independently performed in a preprocessing procedure and were normalized and numeralized for developing classification models of the dataset for the machine learning algorithms. The nonnumerical data included in the dataset during this procedure were numeralized. The numerical data

were degraded between 0 and 1 for normalization.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}. \quad (1)$$

z_i indicates the normalized i data. x_i is normalized between 0 and 1 interval by division with the value between $\min(x)$ and $\max(x)$ after deduction from the minimum x value included in the dataset to be normalized.

All data were numaralized and normalized for obtaining more successful results in the machine learning algorithm.

2.4. Creation of the Supervised Machine Learning Models. After the preprocessing procedure, the first dataset with 1460 patients and 26 features and the second dataset with 1532 patients and 9 features were, respectively, evaluated using the k nearest neighbours (KNN), support vector machines (SVM), and decision tree (DT) machine learning algorithms, and classification models were generated for the detection of the *BRCA1/2* negativity. k nearest neighbour is a supervised machine learning technique. k performs the classification procedure in accordance with the nearest neighbour. SVM classifies the samples by generating the optimum hypersurface which differentiates the samples in the space. DT is a supervised machine learning technique which performs classification by generating a decision tree.

In the scope of the study, both 2 datasets that were normalized and numaralized were divided into two groups: 67% as the training dataset and 33% as the test dataset. Three different classification models were developed with training of KNN, SVM, and DT algorithms using the training dataset. The performances of the developed classification models were measured using the test data and were comparatively presented.

3. Results and Discussion

The raw dataset involving 125 features of 2070 high-risk breast cancer patients was used for the detection of the *BRCA1/2* negativity. The number of features was reduced to

TABLE 2: Patient and feature information of the datasets.

Datasets	Patient number (n)	Feature number (n)	Features
Raw dataset	2070	125	Dia, age, sex, CS, PS, HG, MS, NI, ER, PR, HER2, RS, NOBCIF, NOOCIF, NOOTIF, NOTCIF, GR, S, AI, MG, P, LS, M, OC, HI, BRCA etc.
Dataset 1	1460	26	Dia, age, sex, CS, PS, HG, MS, NI, ER, PR, HER2, RS, NOBCIF, NOOCIF, NOOTIF, NOTCIF, GR, S, AI, MG, P, LS, M, OC, HI, BRCA
Dataset 2	1532	9	Dia, age, FIB, SEB, THB, THB < 40, FIO, FIO > 40, BRCA

Dia: diagnosis; CS: clinical stage; PS: pathological stage; HG: histological grades; MS: metastasis status; NI: node involvement; ER: estrogen receptor; PR: progesterone receptor; RS: receptor status; NOBCIF: number of breast cancer in family; NOOCIF: number of ovarian cancer in family; NOOTIF: number of other tumors in family; NOTCIF: number of total cancer in family; FIB: number of first degree relative in breast; SEB: number of second degree relative in breast; THB: number of third degree relative in breast; THB < 40: number of third degree relative in breast under age of 40; FIO: number of first degree relative in ovarian; FIO > 40: number of first degree relative in ovarian above age of 40; GR: geographical regions; S: smoking; AI: alcohol intake; MG: mammography; P: pregnancy; LS: lactation status; M: menopause; OC: oral contraceptive; HI: hormone intake.
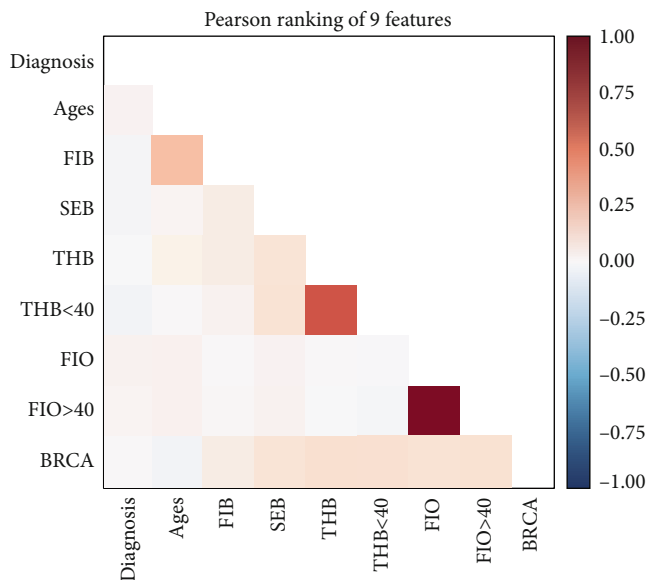

FIGURE 2: Association between the features. Dia; diagnosis; FIB; number of first degree relative in breast; SEB; number of second degree relative in breast; THB; number of third degree relative in breast; THB<40: number of third degree relative in breast under age of 40; FIO: number of first degree relative in ovarian; FIO>40: number of first degree relative in ovarian above age of 40.

two different datasets as 26 and 9 using the experimental dataset dimension reduction procedure. All the learning algorithms were separately performed to both 2 datasets, and the success rates were measured. The success rates of algorithms used in developing the classification models for Dataset 1 and Dataset 2 are given in Table 3 and Figure 3.

The performance of the algorithm was identified by studying the machine learning model with the test dataset. Accordingly, 9 features involving dataset 2 was found to be the most effective for KNN and DT accuracy rates compared with the 26 features involving Dataset 1. However, no difference was detected between both datasets for SVM rates. We found that the rem006Fval of the unnecessary data by decreasing the number of features in the dataset significantly

increased the accuracy rate of the algorithm compared with the decision tree. As shown in Table 3 and Figure 3, Dataset 2 was the algorithm with the highest accuracy rate of 92.88% compared to the DT algorithm. KNN and SVM were the other algorithms with higher accuracy rates of 87.94% and 86.56%.

The risk of being diagnosed with breast cancer and *BRCA1/2* gene positivity have been studied to be measured with various statistical methods in high-risk individuals for breast cancer. However, the developed statistical models could only identify the possibility of *BRCA1/2* positivity and could not estimate the possibility of *BRCA1/2* positivity with high accuracy [29, 30]. Some of these statistical methods could identify the patient groups from specific ethnic groups and cannot be used in patient groups from different ethnicities. Some of the methods could only estimate the *BRCA1* positivity, and some could only identify the risk for specific age intervals [31–33].

Doncescu and Kabbaj developed a model which could estimate the breast cancer development in accordance with the age of the individual using the machine learning techniques. The age of being diagnosed with breast cancer of the individual is estimated in this modelling based on the *BRCA1/2* gene polymorphisms and familial cancer history. The clusters corresponding to different profiles were identified using an accurate mathematical model for estimating the cluster of the individual in this analysis method. The individuals were evaluated in accordance with this clustering method, and the possibility of getting breast cancer of the individual was found to be higher in the second decade of life using this mathematical method if the healthy individual had 2 relatives diagnosed with breast cancer in their 20s [34].

KhajePasha et al. developed an algorithm which predicted the clinical importance of the *BRCA1/2* single-nucleotide substitution variants with unknown clinical significance using the probabilistic neural network and deep neural network-stacked autoencoder [35]. As a conclusion, intelligent methods were found to have significantly higher estimation rate with no human intervention and were found to be appropriate and applicable methods in resolving various problems when the modelling is performed well and tested with adequate dataset.

TABLE 3: Results of the supervised machine learning algorithms.

Method	Dataset	Number of features (n)	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
SVM	Dataset 1	26	86.72	75	87	81
	Dataset 2	9	86.56	75	87	80
KNN	Dataset 1	26	85.68	78	86	81
	Dataset 2	9	87.94	86	88	85
DT	Dataset 1	26	78.83	80	79	79
	Dataset 2	9	92.88	93	93	92

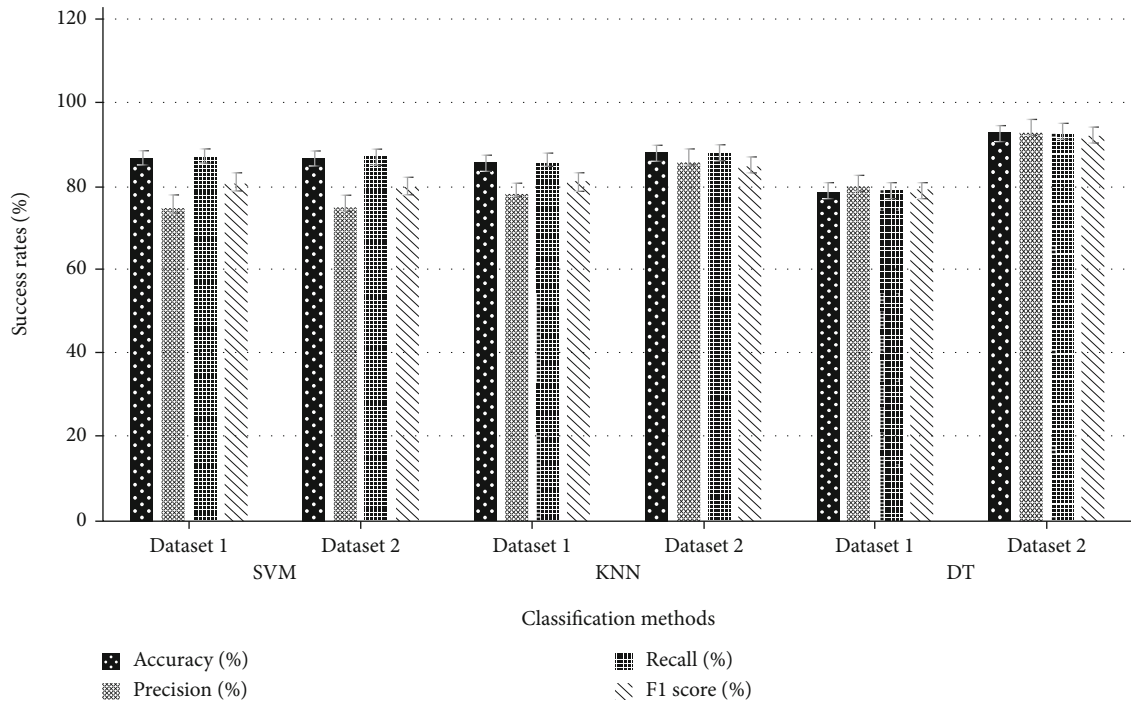SVM: support vector machines; KNN: k nearest neighbours; DT: decision tree.

FIGURE 3: The comparative demonstration of the performances of the algorithms. The x-axis shows the classification models for Dataset 1 and Dataset 2. The y-axis shows the success rates (%) of the algorithms. SVM: support vector machines; KNN: k nearest neighbours; DT: decision tree.

Sumitha investigated whether breast cancer could be diagnosed with *BRCA1/2* gene expression analysis using the machine learning algorithms in their study. Researchers investigated the *BRCA1/2* gene expression level using the machine learning algorithms in that study. The algorithms used in the study were the sequential algorithm based on confusion matrix, DCKSVM, and HRBFNN and the prediction algorithms. The HRBFNN algorithm was proven to be effective in analysing the *BRCA1/2* gene expression in the breast cancer dataset in that investigation. Researchers emphasized that it could be made stronger using the other computer-aided algorithms for improving the productivity [22].

The *BRCA1/2* gene test used in the clinical practice is difficult and requires excessive workload and is more expensive and time-consuming compared with the other tests. Longer time of testing causes stress for the patient. In addition, the identification of *BRCA1/2* gene mutation is significantly important for selection of treatment and identification of the risk of the possibility of the secondary cancer. All the above-mentioned algorithms can identify the breast cancer risk in healthy individuals in specific categories and can provide the possibility of carrying *BRCA1/2* gene mutation. However, the rate of the *BRCA1/2* gene mutation carriers is approximately 15%, and 85% of the high-risk patients are unnecessarily tested. Enabling to obtain the clinical results of this expensive, difficult, and long-term taking test and selection of the accurately correct individuals for testing in clinical practice will facilitate the routine practice and reduce the health costs. In addition, selection of *BRCA1/2* gene mutation-negative individuals among the high-risk breast cancer patients is significantly important for reducing the workload and health costs and reducing the stress of the patient and for fast selection of the treatment.

4. Conclusions

We evaluated all the clinical, demographic, and genetic features of 2070 patients and developed an algorithm which identifies the *BRCA1/2* gene mutation in a very short time in our study targeting to solve these problems. *BRCA1/2* negativity was identified in 92.88% accuracy with no need for *BRCA1/2* gene testing within minutes in high-risk breast cancer patients using this algorithm and stress of waiting for the test result; time and money loss were prevented. We suggest that this algorithm will be useful in the rapid and accurate performing of the treatment planning of the patients in addition to accelerating the clinical practice.

Data Availability

The data used for the findings of this study are restricted by the Local and Clinical Research Ethics Committee of Istanbul University in order to protect patient privacy.

Ethical Approval

This study complies with ethical standards. The study was approved by the Local and Clinical research Ethics Committee of Istanbul University according to the tenets of the Declaration of Helsinki.

Consent

All patients were informed about the study, and the consents were granted in the scope of the cancer genetics polyclinic. Individual consent was signed by each individual.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[2] B. Wang, F. Yuan, X. Kong, L. D. Hu, and Y. D. Cai, "Identifying novel candidate genes related to apoptosis from a protein-protein interaction network," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 715639, 11 pages, 2015.

[3] F. Yuan, Y. Zhou, M. Wang et al., "Identifying new candidate genes and chemicals related to prostate cancer using a hybrid network and shortest path approach," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 462363, 12 pages, 2015.

[4] B. Sun, M. Zhang, P. Cui et al., "Nonsynonymous single-nucleotide variations on some posttranslational modifications of human proteins and the association with diseases," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 124630, 12 pages, 2015.

[5] Y. Zhan, W. Guo, Y. Zhang, Q. Wang, X. J. Xu, and L. Zhu, "A five-gene signature predicts prognosis in patients with kidney renal clear cell carcinoma," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 842784, 7 pages, 2015.

[6] Z. Ji, G. Meng, D. Huang, X. Yue, and B. Wang, "NMFBFS: a NMF-based feature selection method in identifying pivotal clinical symptoms of hepatocellular carcinoma," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 846942, 12 pages, 2015.

[7] L. Yuan, X. Tang, B. Zhang, and G. Ding, "Cell pluripotency levels associated with imprinted genes in human," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 471076, 8 pages, 2015.

[8] L. Chen, T. Huang, C. Lu, L. Lu, and D. Li, "Machine learning and network methods for biology and medicine," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 915124, 2 pages, 2015.

[9] D. Aksu, A. Abdulwakil, and M. A. Aydin, "Detecting phishing websites using support vector machine algorithm," *Press Academia Procedia*, vol. 5, pp. 139–142, 2017.

[10] D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, *Intrusion detection with comparative analysis of supervised learning techniques and Fisher score feature selection algorithm*, pp. 141–149, 2018.

[11] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.

[12] U. J. Schoepf and P. Costello, "CT angiography for diagnosis of pulmonary embolism: state of the art," *Radiology*, vol. 230, no. 2, pp. 329–337, 2004.

[13] U. J. Schoepf, A. C. Schneider, M. Das, S. A. Wood, J. I. Cheema, and P. Costello, "Pulmonary embolism: computer-aided detection at multidetector row spiral computed tomography," *Journal of Thoracic Imaging*, vol. 22, no. 4, pp. 319–323, 2007.

[14] M. M. Dundar, G. Fung, B. Krishnapuram, and R. B. Rao, "Multiple-instance learning algorithms for computer-aided detection," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1015–1021, 2008.

[15] R. M. Summers, "Improving the accuracy of CTC interpretation: computer-aided detection," *Gastrointestinal Endoscopy Clinics of North America*, vol. 20, no. 2, pp. 245–257, 2010.

[16] H. Yoshida and J. Nappi, "CAD in CT colonography without and with oral contrast agents: progress and challenges," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 267–284, 2007.

[17] H. P. Chan, S. C. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network," *Medical Physics*, vol. 22, no. 10, pp. 1555–1567, 1995.

[18] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine and Biology*, vol. 58, no. 13, pp. R97–129, 2013.

[19] T. M. Mitchell, S. V. Shinkareva, A. Carlson et al., "Predicting human brain activity associated with the meanings of nouns," *Science*, vol. 320, no. 5880, pp. 1191–1195, 2008.

[20] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal Alzheimer's disease via pattern

classification of magnetic resonance imaging," *Neurobiology of Aging*, vol. 29, no. 4, pp. 514–523, 2008.

[21] D. Kim, J. Burge, T. Lane, G. D. Pearlson, K. A. Kiehl, and V. D. Calhoun, "Hybrid ICA-Bayesian network approach reveals distinct effective connectivity differences in schizophrenia," *NeuroImage*, vol. 42, no. 4, pp. 1560–1568, 2008.

[22] J. D. Sumitha and T., "Breast cancer diagonosis through analysis of brca gene using machine learning algorithms," *Pakistan Journal of Biotechnology*, vol. 13, no. 4, pp. 231–235, 2016.

[23] S. G. Kemal Polat, "Breast cancer diagnosis using least square support vector machine," *Digital Signal Processing*, vol. 17, no. 4, pp. 694–701, 2007.

[24] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *International Journal of Cancer*, vol. 127, no. 12, pp. 2893–2917, 2010.

[25] Globocan 2018, 2018, Available from: https://gco.iarc.fr/today/data/factsheets/populations/792-turkey-fact-sheets.pdf.

[26] E. M. Rosen, S. Fan, R. G. Pestell, and I. D. Goldberg, "BRCA1 gene in breast cancer," *Journal of Cellular Physiology*, vol. 196, no. 1, pp. 19–41, 2003.

[27] A. Paul and S. Paul, "The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers," *Frontiers in Bioscience*, vol. 19, no. 4, p. 605, 2014.

[28] K. B. Kuchenbaecker, J. L. Hopper, D. R. Barnes et al., "Risks of breast, ovarian, and contralateral breast cancer for-BRCA1andBRCA2mutation carriers," *JAMA*, vol. 317, no. 23, pp. 2402–2416, 2017.

[29] F. J. Couch, M. L. DeShano, M. A. Blackwood et al., "BRCA1-mutations in women attending clinics that evaluate the risk of breast cancer," *The New England Journal of Medicine*, vol. 336, no. 20, pp. 1409–1415, 1997.

[30] D. Shattuck-Eidens, A. Oliphant, M. McClure et al., "BRCA1 sequence analysis in women at high risk for susceptibility mutations," *JAMA*, vol. 278, no. 15, pp. 1242–1250, 1997.

[31] M. H. Gail, L. A. Brinton, D. P. Byar et al., "Projecting individualized probabilities of developing breast cancer for white females who are being examined annually," *Journal of the National Cancer Institute*, vol. 81, no. 24, pp. 1879–1886, 1989.

[32] G. Parmigiani, D. Berry, and O. Aguilar, "Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2," *American Journal of Human Genetics*, vol. 62, no. 1, pp. 145–158, 1998.

[33] T. S. Frank, S. A. Manley, O. I. Olopade et al., "Sequence analysis of BRCA1 and BRCA2: correlation of mutations with family history and ovarian cancer risk," *Journal of Clinical Oncology*, vol. 16, no. 7, pp. 2417–2425, 1998.

[34] B. T. Andrei Doncescu and N. Kabbaj, "Machine Learning applied to BRCA1 hereditary breast cancer data," in *2009 International Conference on Advanced Information Networking and Applications Workshops*, Bradford, UK, 2009.

[35] M. B. E. R. KhajePasha, H. Kheiri, R. M. Manjili, and R. Amandi, "Predicting clinical significance of BRCA1 and BRCA2 single nucleotide substitution variants with unknown clinical significance using probabilistic neural network and deep neural network-stacked autoencoder," *Bioinformatics*, vol. 34, 2018.