İstanbul Med J 2022; 23(3): 210–5
DOI: 10.4274/imj.galenos.2022.62443

Prediction of Breast Cancer Distant Metastasis by Artificial Intelligence Methods from an Epidemiological Perspective

Sami Akbulut[1], Fatma Hilal Yagin[2], Cemil Colak[2]

[1]Inonu University Faculty of Medicine, Department of Surgery, Biostatistics and Medical Informatics and Public Health, Malatya, Turkey
[2]Inonu University Faculty of Medicine, Department of Biostatistics, and Medical Informatics Malatya, Turkey

ABSTRACT

Introduction: Despite significant advances in breast cancer (BC) management, the prognosis for most patients with distant metastasis remains poor. We predicted distant metastasis in BC patients with artificial intelligence (AI) methods based on genomic biomarkers.

Methods: The dataset used in the study included 97 patients with BC, of whom 46 (47%) developed distant metastases, and 51 (53%) did not develop distant metastases, and the expression level of 24,481 genes of these patients. An approach combining Boruta + LASSO methods was applied to identify biomarker genes associated with BC distant metastasis. Mann-Whitney U test was used to examine the difference between groups in terms of gene expression levels in statistical analyses, and Cohen d effect sizes and odds ratios were calculated. AdaBoost and XGBoost algorithms, which are tree-based methods, were used for BC distant metastasis prediction, and the results were compared by evaluating comprehensive performance criteria.

Results: After Boruta + LASSO methods, 14 biomarker candidate genes were identified. These predictive genes were *PIB5PA, SSX2, OR1F1, ALDH4A1, FGF18, WISP1, PRAME, CEGP1, AL080059, NMU, ATP5E, SMARCE1, FGD6,* and *SLC37A1*. In effect size results; in particular, show that the AL080059 (Cohen's D: 1.318) gene is clinically predictive of BC Metastasis. The accuracy, F1-score, positive predictive value, sensitivity, and area under the ROC Curve (AUC) values obtained with the AdaBoost algorithm for BC metastasis prediction was 95%, 96.3%, 100%, 92.6%, and 98.8%, respectively. The model created with the XGBoost algorithm, on the other hand, obtained 90%, 92.9%, 92.9%, 92.9%, 97.6% accuracy, F1-score, positive predictive value, sensitivity, and AUC values, respectively.

Conclusion: Identifying genes that successfully predict BC distant metastasis with AI methods in the study may be decisive for future therapeutic targets and help clinicians better adapt adjuvant chemotherapy to their patients. Additionally, the AdaBoost prediction model created can discriminate patients at risk of BC distant metastases.

Keywords: Breast cancer, distant metastasis, genetic risk factors, genomics, artificial intelligence

Introduction

Breast cancer (BC) is among the leading causes of mortality and morbidity among women. GLOBOCAN 2020 results reveal that BC is the most frequent cancer with an incidence of 11.7% and the fifth most frequent cause of death due to cancer with an incidence of 6.9% (1). Lifetime BC risk of a woman in a developed country is 12.5%, whereas the risk of mortality due to BC is 3.4% (2). A wide distrubition of incidence is also present between different ethnic groups and caucasian or afro-american populations of the same country.

BC is a significant public health problem for either developed or developing countries regarding financial and psychosocial issues. BC incidence is relatively higher in countries with higher income compared with BC incidence in middle and lower income countries. Epidemiological studies to analyse this difference revealed the impact of environmental factors, lifestyle, nutritional factors and sociocultural status as triggering factors of BC.

Predisposing factors for BC are considered in seven subgroups: demographic (age, female gender), reproductive (late age of menopause, pregnancy characteristics), hormonal (hormonal contraceptive methods, postmenopausal hormone therapy), breast-related factors (some benign breast disorders), lifestyle (obesity or overweight, alcohol consumption, smoking, diet), others (air pollution, night work, socioeconomic status, diabetes, radiation) and hereditary factors (genetic factors, positive family history of BC) (3,4). Genetic factors, which are among the hereditary risk factors, have been studied for many years. Mutations of either oncogenes or anti-oncogenes and abnormal amplification effects formation and progression (4). BC-associated genes revealed in previous studies are *BRCA1, BRCA2, c-erbB-2 (HER2), c-erbB-1 (HER1), TP53, PTEN, PALB2, STK11, CDH1, ATM, CHRK*, c-Myc ve Ras (4,5). However, there are many genes whose relationship with BC is still at the research level.

Despite the advances in BC treatment in the last 20-30 years, patients with metastatic disease still have a poor prognosis with a survival of five

Address for Correspondence: Sami Akbulut MD, Inonu University Faculty of Medicine, Department of Surgery, Biostatistics and Medical Informatics and Public Health, Malatya, Turkey
Phone: +90 532 325 12 12 **E-mail:** akbulutsami@gmail.com **ORCID ID:** orcid.org/0000-0002-6864-7711

Cite this article as: Akbulut S, Yagin FH, Colak C. Prediction of Breast Cancer Distant Metastasis by Artificial Intelligence Methods from an Epidemiological Perspective. İstanbul Med J 2022; 23(3): 210-5.

Received: 07.06.2022
Accepted: 25.07.2022

to ten years (6,7). Recent studies determined locoregional recurrence and distant metastasis rates as 5-15% to 11-18.7% respectively (8-10). Hence, distant metastasis and recurrence have a negative impact on recurrence-free and overall survival.

The vast majority of studies for the molecular structure of BC are focused on primary cancers.

Gene expression profiles divide BC into different subgroups and clinical trial point out these transcriptional signatures to impact making therapeutic decisions (7,11). Recent large scaled genomic analyses ease revealing complicated mutational configurations. Despite the largely defined genomic configuration of BC, the same success is not actually present for genetic configurations of locoregional or distant-metastasis BC. Studies for metastatic disease up to date clonally determined relationship between metastases and primary tumor, presence of various common mutations and presence of typically additional mutations, which are not present in primary tumors (12,13).

Microarray technology, which provides simultaneous quantitative monitoring of expression levels of thousands of genes, is an important research topic in the early diagnosis of primary BC and its metastases (locoregional recurrence, distance metastasis) with artificial intelligence (AI)/machine learning (ML) methods. However, the predictive performance of AI/ML models may be adversely affected by many genes unrelated to the disease(s) and may not contribute to the classification. A technology that can be used to eliminate this problem is ML. ML is widely recognized as the choice approach in BC pattern classification and forecast modeling due to its unique benefits in detecting essential characteristics/genes from complicated BC datasets. Recently, ML approaches have played an essential role in the diagnosis and prognosis of BC by using classification techniques to identify persons with BC, differentiate benign from malignant tumors and predict prognosis. Accurate categorization can also help clinicians prescribe the best treatment regimen (14,15). Considering these data, this study intended to identify biomarker candidate genes in predicting BC recurrence with AI modeling.

Methods

Dataset

Gene expression and clinical data used in the study were obtained from the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) database. The dataset included 97 patients with lymph node-negative (pN0) BC, of which 46 (47%) had developed distant metastasis within 5 years and 51 (53%) had not developed distant metastasis. Clinical data included information on patients' age, pathological tumor size and grade, ER and V-ERB-B2 avian erythroblast leukemia viral oncogene homolog 2 (ERBB2) statutes, and follow-up results. In the gene expression data set, 97 patients had expression levels of 24,481 genes (16).

Data Preprocessing and Modeling

At baseline, there were 24,481 gene expression levels in the BC metastasis dataset. The Boruta + LASSO method was used to select candidate gene biomarkers associated with metastasis. Boruta is a method that iteratively removes from the dataset variables that have been statistically proven to be less relevant to the response (here, metastasis). LASSO for variable selection obtains a sparse regression model. For a given dataset (X,y), X is the explanatory variable and y is the variable to be explained. The LASSO method estimates the β parameters of the model. It then selects the important variables by applying a λ constraint to the predicted parameters. Variables with β parameters shrinking to zero are considered unimportant. Of the variable selected data set, 80% is randomly split to train the model and the remaining 20% to test the model. This split was repeated 100 times and average scores were calculated 100 times in the evaluation of the models. Two different models, AdaBoost and XGBoost, were created for BC metastasis prediction based on genomic biomarkers. The performance of the generated models was evaluated by accuracy, F1-score, positive predictive value, sensitivity, and the area under the ROC curve (AUC), and the results were compared.

Study Protocol and Ethics Committee Approval

This study, which was prepared using the NCBI GEO open-access dataset, involving human participants, was in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical approval was obtained from the İnönü University Institutional Review Board in Non-Interventional Clinical Research (approval number: 2022/3645, date: 07.06.2022). Strengthening the reporting of observational studies in epidemiology guideline was used to assess the likelihood of bias and overall quality of this study (17).

Statistical Analysis

Qualitative variables are summarized as numbers and percentages. Quantitative variables were digested with the median and inter-quantile range. Two groups were compared with the Mann-Whitney U test. Statistical tests with a p-value of less than 5% were considered significant. The Cohen's D effect size was calculated for variables with a significant p-value. For the Mann-Whitney U test, the effect size (Cohen's D) was interpreted as a small effect between 0.20-0.50, a medium impact between 0.50-0.80, and a large impact above 0.80 (18). Additionally, odds ratio estimates for quantitative biomarker genes with significant p-value were obtained by logistic regression analysis. All statistical analysis were performed in IBM SPSS Statistics for Windows version 26.0 (New York; USA) and Python 3.9.

Results

Descriptive statistics on patient information in the clinical dataset are given in Table 1. In the study, 74 (76%) patients were older than 40 years and the remaining 23 (24%) patients were younger than 40 years old. The tumor size was smaller than 20 mm in 44 (45%) patients and larger than 20 mm in 53 (55%) patients. 37 (38%) of the patients were stage I-II and 60 (62%) were stage III BC. While 72 (74%) had a positive ER status, 25 (26%) were negative. While 15 (16%) were HER2 positive, 82 (84%) were negative. Fourty-six (47%) of 97 patients had a recurrence of metastasis within 5 years, and 51 (53%) had no metastasis.

Descriptive statistics, effect sizes, and odds ratios (95% confidence interval) of selected genes after Boruta + LASSO feature selection methods are given in Table 2. There was a statistically significant difference between the metastasis and no-metastasis patient groups in terms of expression levels of all 14 genes selected as biomarker candidates that may be effective in the diagnosis and treatment of BC distant metastasis. Effect size results; in particular, show that the *AL080059* (Cohen's D: 1.318) gene is clinically predictive of BC Metastasis (Table 2). When the odds ratio estimations are examined; a one-unit decrease in the expression levels of the *PIB5PA, OR1F1, ALDH4A1, FGF18, WISP1, CEGP1,* and SMARCE1 genes increases the risk of metastasis by 10.75, 125, 166.66, 43.47, 100, 5.52, 83.33 times, respectively. In contrast, a one-unit increase in the expression levels of the *PRAME, AL080059, NMU,* and *ATP5E* genes increases the risk of metastasis by 4.454, 57.248, 35.396, and 728.461 times, respectively. Table 3 shows the results of the performance criteria of the AdaBoost and XGBoost models created for BC metastasis estimation. When Table 3 is examined, the accuracy, F1-score, positive predictive value, sensitivity, and AUC values obtained in the test data set for the AdaBoost algorithm for BC metastasis prediction are 95%, 96.3%, 100%, 92.6%, 98.8%, respectively. The model created with the XGBoost algorithm, on the other hand, obtained the accuracy, F1-score, positive predictive value, sensitivity, and AUC values of 90%, 92.9%, 92.9%, 92.9%, 97.6%, respectively, in the test data set.

Table 1. Descriptive statistics on clinical information in the breast cancer dataset

Patient's clinical information		(n=97)
Age	≥40 years	74 (76%)
	<40 years	23 (24%)
Tumor size	<20 mm	44 (45%)
	≥20 mm	53 (55%)
Grade	1-2	37 (38%)
	3	60 (62%)
ER status	Positive	72 (74%)
	Negative	25 (26%)
HER2 status	Positive	15 (16%)
	Negative	82 (84%)
Metastatic relapse within 5 years	Yes	46 (47%)
	No	51 (53%)

Discussion

Despite significant advances in BC treatment recently, the prognosis for most patients with distant metastasis remains poor. BC patients with the same disease stage may have markedly different treatment responses and overall outcomes. The strongest predictors of metastasis (eg, lymph node status and histological grade) cannot accurately classify BCs based on their clinical behavior. Additionally, an in-depth understanding of the molecular phenotype of distant metastasis is critical to pave the way for earlier detection of metastasis and more effective treatments. Therefore, in this study, we predicted distant metastases in patients BC using AI methods based on genomic biomarkers (18,19).

Microarray data of 24,481 genes of 97 patients with and without distant metastasis were used in the study. For AI models, the fact that microarray data contain thousands of gene information belonging to few patients both lead to computational inefficiency and reduces the performance of prediction models. Additionally, it may be useless to use information about thousands of genes in clinical practice, and there may be many genes unrelated to the disease of interest in these datasets containing many genes. From this perspective, the identification of a small subset of genes with AI methods not only facilitates transfer to clinics but also limits the identification of false-positive predictive genes. For this reason, in this study, a methodology combining Boruta + LASSO methods was applied to identify candidate biomarker genes that

Table 2. Statistical analysis results of selected genes as a result of Boruta + LASSO

Genes*	Breast cancer		p-value	ES	OR (95% CI)
	No-metastasis	Metastasis			
PIB5PA	-0.005 (0.367)	-0.337 (0.495)	<0.001	0.828 (large)	0.093 (0.022-0.326)
SSX2	0.092 (0.414)	-0.07 (0.298)	0.001	0.717 (medium)	1.032 (0.993-NA)
OR1F1	0.112 (0.193)	0.029 (0.074)	<0.001	0.81 (large)	0.008 (0-0.167)
ALDH4A1	0.126 (0.194)	-0.071 (0.262)	<0.001	0.974 (large)	0.006 (0-0.072)
FGF18	0.06 (0.416)	-0.242 (0.284)	<0.001	0.995 (large)	0.023 (0.003-0.127)
WISP1	0.064 (0.278)	-0.079 (0.208)	<0.001	0.816 (large)	0.01 (0.001-0.124)
PRAME	-0.77 (0.228)	0.054 (1.17)	<0.001	0.786 (medium)	4.454 (2.145-10.27)
CEGP1	0.065 (0.484)	-0.755 (0.734)	<0.001	1.032 (large)	0.181 (0.076-0.394)
AL080059	-0.391 (0.344)	0.076 (0.436)	<0.001	1.318 (large)	57.248 (12.013-364.325)
NMU	-0.302 (0.256)	-0.06 (0.392)	<0.001	1.091 (large)	35.396 (6.389-280.071)
ATP5E	-0.054 (0.128)	0.054 (0.166)	<0.001	0.981 (large)	728.461 (121.328-8541.318)
SMARCE1	-0.005 (0.288)	-0.18 (0.238)	<0.001	0.985 (large)	0.012 (0.001-0.106)
FGD6	0.029 (0.235)	-0.121 (0.162)	<0.001	0.86 (large)	0.143 (0.012-0.979)
SLC37A1	-0.064 (0.253)	0.069 (0.315)	0.002	0.655 (medium)	23.439 (2.749-245.161)

*: Gene expression levels are summarized as "median (IQR)", OR: Odds ratio, CI: Confidence interval, ES: Effect size

Table 3. Results of performance measures for models created to predict breast cancer

Models	Accuracy	F1-score	Positive predictive value	Sensitivity	AUC
AdaBoost	0.95	0.963	1.000	0.926	0.988
XGBoost	0.90	0.929	0.929	0.929	0.976

AUC: Area under the ROC curve

may be associated with distant metastasis. In this way, 14 genes that may be associated with BC metastasis were identified. These predictive genes were *PIB5PA, SSX2, OR1F1, ALDH4A1, FGF18, WISP1, PRAME, CEGP1, AL080059, NMU, ATP5E, SMARCE1, FGD6,* and *SLC37A1*. Then, two different models, AdaBoost and XGBoost, were created using 14 genes determined for distant metastasis prediction. The accuracy, F1-score, positive predictive value, sensitivity, and AUC values obtained with the AdaBoost algorithm for BC metastasis prediction were 95%, 96.3%, 100%, 92.6%, and 98.8%, respectively. The model created with the XGBoost algorithm, on the other hand, obtained the accuracy, F1-score, positive predictive value, sensitivity, and AUC values of 90%, 92.9%, 92.9%, 92.9%, and 97.6%, respectively. The results showed that AdaBoost outperformed XGBoost in BC distant metastasis prediction.

Our gene selection results were generally compatible with the literature. In a study in the literature, it was reported that high PIB5PA levels are associated with limited tumor progression and better prognosis in patients with BC (16). Greve et al. (20) investigated the phenotypic and molecular changes associated with SSX2 expression in human melanoma and BC cells and showed that the *SSX2* gene has oncogenic potential. Additionally, the study highlighted the potential of this gene as a therapeutic target (20).

ALDH1A1 is an essential element in the retinoic acid signaling pathway that regulates self-renewal and differentiation of normal stem cells and may play an important role in cancer progression. Liu et al. (21) emphasized that high expression of ALDH1A1 mRNA in tumor tissues may be an independent predictor of a positive triple-negative BC outcome. Marcato et al. (22) In another study, they showed that ALDH1A3 expression could predict metastasis in BC patients. Song et al. (23) showed that the *FGF18* gene promotes epithelial-mesenchymal transition and migration in BC cells and emphasized that FGF18 expression may be a potential prognostic therapeutic marker for BC.

WISP1 genetic polymorphisms were highlighted in a study in the literature to be associated with platinum-based chemotherapy toxicity and sensitivity to platinum-based chemotherapy responses in patients with lung cancer (24). It has been reported that WISP1 can also predict a patient's susceptibility to cervical cancer and hepatocellular carcinoma (25,26). Wang et al. (27) emphasized that WISP1 polymorphisms play a critical role in BC. In another study, Sokol et al. (28) emphasized that the expression of the *SMARCE1* gene in patients diagnosed with early-stage BC would be a strong indicator of recurrence and metastasis. Additionally, they reported that SMARCE1 expression identifies early-stage breast, ovarian, and lung cancers that are likely to progress and metastasis.

Epping et al. (29) emphasized that PRAME expression is a prognostic marker for the clinical outcome of BC. The results of the study showed that PRAME was an independent predictor of shortened metastasis-free interval in patients not receiving adjuvant chemotherapy. PRAME expression was associated with tumor grade and negative estrogen receptor status. Lu et al. (30) reported that CEGP1 expression is associated with locoregional tumor recurrence or distant metastasis in patients with BC.

In a recent study, it was emphasized that the *AL080059* gene is one of the ten prognostic marker genes that differ between normal and tumor tissues of patients with BC (31). Galber et al. (32) reported that ATP synthase contributes to cancer development or metastasis Amino acid changes in ATP synthase encoded by the *ATP6* gene have been detected in pancreatic cancer cells (33), thyroid (34), cervical, bladder, and head/neck cancers, as well as in leukemia (35) and acute myeloid leukemia (36) patients. In a study, it was observed that the *A6L* gene, which is derived from ATP8, is mutated in ovarian, breast, cervical, and thyroid cancers (35). Additionally, Grzybowska-Szatkowska et al. (37) found homoplasmic mutations in the *ATP6* and *ATP8* genes in patients with BC. However, there was no information in the literature that ATP5E, one of the 14 genes we selected, is directly related to BC. Future studies should examine whether ATP5E is a predictive biomarker for patients with BC.

Garczyk et al. (38) have identified NMU as a drug response biomarker candidate for patients with BC. Additionally, they reported that NMU may be a putative therapeutic target to reduce the metastatic spread of BC cells (38). Another study conducted on the *FGD6* gene, which was selected as a biomarker candidate in our study, reported that it is an independent prognostic risk factor for the survival of patients with gastric cancer (39). However, no study was found that reported the association of this gene with BC. In future studies, investigating whether the *FGD6* gene is associated with BC and metastasis may be important for future therapeutic targets.

In the literature, several different studies have been found that predict metastasis with the dataset we used in this study. For example, in a study using the same data set, a variable selection was made and distant metastasis was predicted with 88.55% accuracy (40). In another study using the same data, the Elastic net method was used and an accuracy rate of 59% was obtained in the prediction of metastasis (41). It can be said that the AdaBoost model created in the current study has a more successful performance in estimating distant metastasis in BC patients compared to the literature.

Study Limitations

As with all retrospective case-control studies, this study has some limitations. Most genomic analysis are usually conducted with few samples, as they require high budgets for each sample. For this reason, a limitation of this study is the sample size. Secondly, the open-access data set was used in this study, which means that some variables can be ignored since all possible factors cannot be accessed in such studies. In future studies, it should be aimed to present a model that can be created for predicting BC metastasis to users with a web-based interface.

Conclusion

To sum up, the identification of genes that successfully predict BC distant metastases with AI methods in the study may be decisive for future therapeutic targets and may help clinicians better adapt adjuvant chemotherapy to their patients. Additionally, the predictive model AdaBoost created can distinguish patients at risk of distant metastasis.

Ethics Committee Approval: Ethical approval was obtained from the Inonu University Institutional Review Board in Non-Interventional Clinical Research (approval number: 2022/3645, date: 07.06.2022).

Informed Consent: Retrospective study.

Peer-review: Externally peer-reviewed.

Authorship Contributions: Concept - S.A.; Design - S.A., F.H.Y.; Data Collection or Processing - S.A., F.H.Y., C.C.; Analysis or Interpretation - F.H.Y., C.C.; Literature Search - S.A., F.H.Y.; Writing - S.A., F.H.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021; 71: 209-49.
2. Rojas K, Stuckey A. Breast Cancer Epidemiology and Risk Factors. Clin Obstet Gynecol 2016; 59: 651-72.
3. Momenimovahed Z, Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer (Dove Med Press) 2019; 11: 151-64.
4. Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, et al. Risk Factors and Preventions of Breast Cancer. Int J Biol Sci 2017; 13: 1387-97.
5. Kwong A, Chen JW, Shin VY. A new paradigm of genetic testing for hereditary breast/ovarian cancers. Hong Kong Med J 2016; 22: 171-7.
6. Tevaarwerk AJ, Gray RJ, Schneider BP, Smith ML, Wagner LI, Fetting JH, et al. Survival in patients with metastatic recurrent breast cancer after adjuvant chemotherapy: little evidence of improvement over the past 30 years. Cancer 2013; 119: 1140-8.
7. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell 2017; 32: 169-84.
8. Holleczek B, Stegmaier C, Radosa JC, Solomayer EF, Brenner H. Risk of loco-regional recurrence and distant metastases of patients with invasive breast cancer up to ten years after diagnosis - results from a registry-based study from Germany. BMC Cancer 2019; 19: 520.
9. Belkacemi Y, Hanna NE, Besnard C, Majdoul S, Gligorov J. Local and Regional Breast Cancer Recurrences: Salvage Therapy Options in the New Era of Molecular Subtypes. Front Oncol 2018; 8: 112.
10. Anwar SL, Avanti WS, Nugroho AC, Choridah L, Dwianingsih EK, Harahap WA, et al. Risk factors of distant metastasis after surgery among different breast cancer subtypes: a hospital-based study in Indonesia. World J Surg Oncol 2020; 18: 117.
11. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2016; 34: 1134-50.
12. De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng CKY, Nuciforo P, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. Ann Oncol 2014; 25: 1729-35.
13. Savas P, Teo ZL, Lefevre C, Flensburg C, Caramia F, Alsop K, et al. The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program "CASCADE". PLoS Med 2016; 13: e1002204.
14. Yue W, Wang Z, Chen H, Payne A, Liu X. Machine learning with applications in breast cancer diagnosis and prognosis. Designs 2018; 2: 13.
15. Vaka AR, Soni B, Reddy S. Breast cancer detection by leveraging Machine Learning. ICT Express 2020; 6: 320-4.
16. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415: 530-6.
17. Cevallos M, Egger M. STROBE (STrengthening the Reporting of OBservational studies in Epidemiology). Guidelines for reporting health research: a user's manual; 2014. pp. 169-79.
18. Ed. Cohen J. "The earth is round (p<.05)," What if there were no significance tests? 2016; 69-82.
19. Obi N, Werner S, Thelen F, Becher H, Pantel K. Metastatic Breast Cancer Recurrence after Bone Fractures. Cancers (Basel) 2022; 14: 601.
20. Greve KB, Lindgreen JN, Terp MG, Pedersen CB, Schmidt S, Mollenhauer J, et al. Ectopic expression of cancer/testis antigen SSX2 induces DNA damage and promotes genomic instability. Mol Oncol 2015; 9: 437-49.
21. Liu Y, Baglia M, Zheng Y, Blot W, Bao PP, Cai H, et al. ALDH1A1 mRNA expression in association with prognosis of triple-negative breast cancer. Oncotarget 2015; 6: 41360-9.
22. Marcato P, Dean CA, Pan D, Araslanova R, Gillis M, Joshi M, et al. Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. Stem Cells 2011; 29: 32-45.
23. Song N, Zhong J, Hu Q, Gu T, Yang B, Zhang J, et al. FGF18 enhances migration and the epithelial-mesenchymal transition in breast cancer by regulating Akt/GSK3β/β-catenin signaling. Cell Physiol Biochem 2018; 49: 1019-32.
24. Chen J, Yin J, Li X, Wang Y, Zheng Y, Qian C, et al. WISP1 polymorphisms contribute to platinum-based chemotherapy toxicity in lung cancer patients. Int J Mol Sci 2014; 15: 21011-27.
25. Lin YH, Hsiao YH, Yang SF, Liu YF, Hsu CF, Wang PH. Association between genetic polymorphisms of WNT1 inducible signaling pathway protein 1 and uterine cervical cancer. Reprod Sci 2018; 25: 1549-56.
26. Chen CT, Lee HL, Chiou HL, Chou CH, Wang PH, Yang SF, et al. Impacts of WNT1-inducible signaling pathway protein 1 polymorphism on hepatocellular carcinoma development. PLoS One 2018; 13: e0198967.
27. Wang Y, Yang SH, Hsu PW, Chien SY, Wang CQ, Su CM, et al. Impact of WNT1-inducible signaling pathway protein-1 (WISP-1) genetic polymorphisms and clinical aspects of breast cancer. Medicine (Baltimore) 2019; 98:e17854.
28. Sokol ES, Feng YX, Jin DX, Tizabi MD, Miller DH, Cohen MA, et al. SMARCE1 is required for the invasive progression of in situ cancers. Proc Natl Acad Sci U S A 2017; 114: 4153-8.
29. Epping MT, Hart AA, Glas AM, Krijgsman O, Bernards R. PRAME expression and clinical outcome of breast cancer. Br J Cancer 2008; 99: 398-403.
30. Lu Y, Tong Y, Huang J, Lin L, Wu J, Fei X, et al. Diverse Distribution and Gene Expression on the 21-Gene Recurrence Assay in Breast Cancer Patients with Locoregional Recurrence Versus Distant Metastasis. Cancer Manag Res. 2021; 13: 6279-89.

31. Song Q, Jing H, Wu H, Zou B, Zhou G, Kambara H. Comparative Gene Expression Analysis of Breast Cancer-Related Genes by Multiplex Pyrosequencing Coupled with Sequence Barcodes. Advances and Clinical Practice in Pyrosequencing. Springer; 2016. pp. 315-25.

32. Galber C, Acosta MJ, Minervini G, Giorgio V. The role of mitochondrial ATP synthase in cancer. Biol Chem 2020; 401: 1199-214.

33. Jones JB, Song JJ, Hempen PM, Parmigiani G, Hruban RH, Kern SE. Detection of mitochondrial DNA mutations in pancreatic cancer offers a "mass"-ive advantage over detection of nuclear DNA mutations. Cancer Res 2001; 61: 1299-304.

34. Máximo V, Soares P, Lima J, Cameselle-Teijeiro J, Sobrinho-Simoes M. Mitochondrial DNA somatic mutations (point mutations and large deletions) and mitochondrial DNA variants in human thyroid pathology: a study with emphasis on Hürthle cell tumors. Am J Pathol 2002; 160: 1857-65.

35. Jiménez-Morales S, Pérez-Amado CJ, Langley E, Hidalgo-Miranda A. Overview of mitochondrial germline variants and mutations in human disease: Focus on breast cancer (Review) Int J Oncol 2018; 53: 923-36.

36. Wu S, Akhtari M, Alachkar H. Characterization of mutations in the mitochondrial encoded electron transport chain complexes in acute myeloid leukemia. Sci Rep 2018; 8: 13301.

37. Grzybowska-Szatkowska L, Slaska B, Rzymowska J, Brzozowska A, Floriańczyk B. Novel mitochondrial mutations in the ATP6 and ATP8 genes in patients with breast cancer. Mol Med Rep 2014; 10: 1772-8.

38. Garczyk S, Klotz N, Szczepanski S, Denecke B, Antonopoulos W, von Stillfried S, et al. Oncogenic features of neuromedin U in breast cancer are associated with NMUR2 expression involving crosstalk with members of the WNT signaling pathway. Oncotarget 2017; 8: 36246-65.

39. Zeng J, Li M, Shi H, Guo J. Upregulation of FGD6 predicts poor prognosis in gastric cancer. Front Med (Lausanne) 2021; 8: 672595.

40. Hameed SS, Hassan R, Hassan WH, Muhammadsharif FF, Latiff LA. HDG-select: A novel GUI based application for gene selection and classification in high dimensional datasets. PloS One 2021; 16: e0246039.

41. Zemmour C, Bertucci F, Finetti P, Chetrit B, Birnbaum D, Filleron T, et al. Prediction of early breast cancer metastasis from DNA microarray data using high-dimensional cox regression models. Cancer Inform 2015; 14:(Suppl 2): 129-38.