

Unraveling metabolic signatures in breast cancer: Machine learning for improved therapeutic targeting

Fatemeh Mehdikhani¹, Parnian Habibi^{2,*}, Sajad Alavimanesh³

¹Department of Clinical Biochemistry, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran

³Student Research Committee, Shahrekord University of Medical Sciences, Shahrekord, Iran

*Author for correspondence:
Email: willsa940@gmail.com

Received date: March 20, 2025

Accepted date: July 22, 2025

Copyright: © 2025 Mehdikhani F, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background: Breast cancer is one of the leading causes of cancer-related mortality among women worldwide. Despite advancements in treatment, therapeutic resistance remains a major challenge, necessitating novel approaches for more effective interventions. One of the hallmarks of cancer, particularly in breast tumors, is metabolic reprogramming, where altered metabolic pathways create distinct profiles compared to normal cells. Identifying these metabolic alterations can provide critical insights for developing targeted therapies aimed at disrupting tumor metabolism and improving patient outcomes.

Objectives: This study applies six machine learning algorithms to predict metabolic profiles in breast cancer patients compared to healthy individuals, providing a promising approach for identifying metabolic targets in precision therapy.

Method: Plasma samples from 102 breast cancer patients and 99 healthy individuals were analyzed using targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) to assess metabolic profiles. Six machine learning algorithms were applied to evaluate classification performance, and feature importance was determined using the Mean Squared Error (MSE) value.

Result: Our findings revealed a significant decrease in alanine, histidine, tryptophan, tyrosine, methionine, and proline levels in breast cancer patients. Among the machine learning models, Random Forest (RF) achieved the highest classification performance (accuracy: 0.90, specificity: 0.85, sensitivity: 0.95), followed by K-Nearest Neighbors (KNN) with similar sensitivity but lower specificity. Logistic Regression (LR) balanced specificity (0.90) and sensitivity (0.86) with an accuracy of 0.88. Naïve Bayes (NB) and Support Vector Machine (SVM) showed moderate accuracy (0.83), while Decision Tree (DT) had the lowest sensitivity (0.76) but the highest PPV (0.89). Feature importance analysis identified glutamic acid, ketocholesterol, cystine, ornithine, succinate, acetylcarnitine, asparagine, tryptophan, and palmitic acid as key metabolic markers.

Conclusion: This study draws attention to key predictive metabolic bottlenecks identified through machine learning models, which could aid in targeted therapy and personalized treatment based on patients' metabolic profiles.

Keywords: Breast cancer, Metabolic alterations, Classification algorithms, Metabolic vulnerability, Feature importance analysis

Abbreviation: mTOR: Mechanistic Target of Rapamycin; TNFα: Tumor Necrosis Factor Alpha; MEKK4: Mitogen-Activated Protein Kinase Kinase Kinase 4; TNBC: Triple-Negative Breast Cancer; KNN: K-Nearest Neighbors; NB: Naïve Bayes; SVM: Support Vector Machine; DT: Decision Tree; RF: Random Forest; LR: Logistic Regression

Introduction

Breast cancer is one of the most prevalent malignancies among the female population worldwide, accounting for a significant portion of cancer-related morbidity and mortality [1]. Despite advancements in treatment strategies, overcoming therapeutic resistance and improving patient

outcomes remain critical challenges [2]. In recent years, cancer metabolism has gained considerable attention, particularly due to the Warburg effect, a phenomenon in which cancer cells preferentially utilize glycolysis for energy production, even in the presence of oxygen [3]. This metabolic shift supports rapid cell proliferation and survival under various stress conditions, a process known as metabolic reprogramming [4]. Given its fundamental role in tumor progression, metabolism has emerged as a defining characteristic of cancer cells [5].

Several metabolic pathways are altered in breast cancer. Glycolysis is highly active, accompanied by elevated expression of glucose transporters and glycolytic enzymes [6,7]. Beyond glucose metabolism, specific amino acids, such as serine and glutamine, are essential for breast cancer growth [8,9]. Additionally, increased activation of the pentose phosphate pathway (PPP) has been reported in breast tumors, contributing to nucleotide biosynthesis and redox homeostasis [10]. Moreover, lipid metabolism is dysregulated, with breast cancer cells utilizing fatty acids and lipid synthesis, including cholesterol uptake, to sustain membrane biosynthesis and energy production [11]. These metabolic adaptations enable cancer cells to thrive in hostile microenvironments and contribute to tumor progression [12].

Predictive medicine is an emerging field that leverages advanced technologies such as genomics, bioinformatics, and artificial intelligence to assess disease risk and personalize healthcare [13–15]. By analyzing genetic variations, lifestyle factors, and clinical data, predictive models can identify individuals susceptible to diseases like diabetes, cancer, and cardiovascular disorders [16], allowing for early diagnosis and targeted prevention strategies [17–19]. Machine learning algorithms enhance predictive accuracy by detecting complex patterns in large datasets, enabling personalized treatment plans and improving patient outcomes [20–22]. Additionally, pharmacogenomics, a key component of predictive medicine, tailors drug therapies based on an individual's genetic makeup, minimizing adverse reactions and optimizing efficacy [23–26]. As predictive medicine continues to evolve, it is reshaping healthcare by shifting the focus from reactive treatment to proactive prevention and precision care [27–29].

Given the distinct metabolic alterations between breast tumors and normal tissues, metabolic profiling holds great potential for improving breast cancer treatment strategies. Identifying these metabolic vulnerabilities may provide novel therapeutic targets, serving as metabolic bottlenecks that can be exploited to hinder tumor progression and enhance treatment efficacy. This study aims to investigate metabolic signatures in breast cancer patients and assess their clinical relevance using machine learning approaches. By analyzing metabolic profiles, we hypothesize that distinct metabolic patterns can serve as predictive metabolic bottlenecks for breast cancer, ultimately contributing to the development of more effective and personalized treatment strategies.

Methods

Study design and participants

This study utilized a dataset comprising demographic and metabolic data from 201 individuals, including 102 breast cancer patients and 99 healthy controls. The study adhered to the principles of the Declaration of Helsinki and was approved by the Inonu University Health Sciences Non-Interventional Clinical Research

Ethics Committee (protocol code = 2024/5750) for data analysis. Prior to sample collection, informed consent was obtained from all participants in accordance with ethical guidelines. Blood samples were collected following an overnight fasting period to minimize metabolic variability. Plasma metabolic profiles were then analyzed using targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) to ensure precise quantification of metabolites. After data acquisition, the samples were categorized based on demographic characteristics and key metabolic markers for further analysis [30,31].

Sample preparation and LC-MS/MS analysis

Plasma samples were thawed overnight at 4°C before metabolite extraction. For each sample, 50 µL of plasma was transferred to a 2 mL vial, followed by protein precipitation using 300 µL of methanol. The mixture was vortexed, incubated at -20°C, and subjected to sonication and centrifugation. The supernatant was collected, dried using a vacuum concentrator, and reconstituted in a solvent containing stable isotope-labeled internal standards for system performance monitoring. A pooled quality control (QC) sample, composed of plasma from both breast cancer patients and healthy controls, was processed identically and analyzed at regular intervals. LC-MS/MS experiments were conducted using a Waters Acquity I-Class UPLC TQS-micro MS system. Chromatographic separation was performed on a Waters Xbridge BEH Amide column at 40°C with a flow rate of 0.3 mL/min. Samples were analyzed in both positive and negative ionization modes, using distinct mobile phase compositions for each mode. A gradient elution program was applied to optimize metabolite separation, and metabolite identities were confirmed by spiking standard compounds into plasma samples. Data acquisition and integration of extracted multiple reaction monitoring (MRM) peaks were performed using Target Lynx software [30].

Study measures

Metabolic markers: Plasma metabolites were analyzed using LC-MS/MS, targeting key metabolic pathways associated with breast cancer progression.

Other clinical variables: Age was included as a covariate in statistical models.

Statistical approach

Machine learning models: A machine learning-based classification approach was applied to metabolic data to categorize individuals into breast cancer patients and healthy controls. Six different machine learning models were evaluated:

K-Nearest Neighbors (KNN): K-Nearest Neighbors (KNN) is a machine learning algorithm that classifies data points based on their proximity to neighboring samples. In this distance-based model, a new sample is categorized according to the most frequent class among its k nearest neighbors [32].

Support Vector Machine (SVM): Support Vector Machine (SVM) is a machine learning algorithm used for data classification. It constructs a decision boundary, known as a hyperplane, to separate different classes and assigns data points to their respective categories. By maximizing the margin between classes, SVM minimizes overfitting and enhances generalization, making it useful for handling complex data structures [33].

Naïve Bayes (NB): Naïve Bayes is a statistical classification algorithm that predicts class membership probabilities. It estimates the probability of a sample being assigned to a particular class, assuming that each attribute contributes to the classification independently of the others. Naïve Bayes is based on Bayes' Theorem [34].

Decision Tree (DT): A decision tree is a supervised learning method that classifies data using a tree-like structure consisting of a root node, branch nodes, and leaf nodes. The dataset is recursively divided into smaller subsets, and the decision tree is built incrementally to improve classification accuracy.

Random Forest (RF): The Random Forest algorithm constructs multiple decision trees and classifies data based on a majority vote. Each decision tree is trained on a different subset of the data, reducing the risk of overfitting. This method enhances model stability and improves accuracy, making Random Forest a reliable and robust classification algorithm.

Logistic Regression (LR): Logistic regression (LR) is a supervised classification algorithm that predicts the probability of an outcome based on continuous or categorical independent variables. It is useful for determining the presence or absence of a characteristic and assumes independent sampling and inclusion of all relevant predictors. As a linear model, LR establishes a weighted relationship between independent variables and the target class.

Model training and evaluation

The classification models were trained to distinguish between breast cancer patients and healthy individuals. The dataset was randomly split into training (80%) and testing (20%) subsets. Model performance was assessed using the following metrics: Accuracy, Sensitivity (True Positive Rate), Specificity (True Negative Rate), Positive Predictive Value (PPV), and Negative Predictive Value (NPV). To ensure robustness and prevent overfitting, a 5-fold cross-validation strategy was applied.

Statistical analysis

All statistical analyses were conducted using R software (version 4.4). Continuous variables with a normal distribution were analyzed using an independent t-test for comparison, with a significance threshold set at $p < 0.001$.

Results

Sample characteristics

The demographic and certain significant metabolic characteristics of the participants are summarized in **Table 1**. A total of 201 plasma samples were analyzed, including 102 samples from breast cancer patients and 99 from healthy controls. The average age of breast cancer patients was 55 years, while that of healthy individuals was 52 years. Additionally, the mean plasma metabolite

Table 1. Sample Characteristics.

	Healthy Samples Mean (SD)	Breast cancer patients Mean (SD)	p
Age	52 (12)	55 (10)	0.056
Alanine	519,367 (141,095)	377,036 (121,443)	<0.001
Histidine	1,210,992 (239,513)	1,080,024 (241,916)	<0.001
Tryptophan	1,196,444 (393,159)	978,986 (291,736)	<0.001
Acetylcarnitine	293,355 (167,744)	492,791 (225,758)	<0.001
Acetylglucosamine	3,440 (1,187)	2,668 (1,054)	<0.001
Adenosine	1,316 (1,224)	2,646 (3,235)	<0.001
Tyrosine	113,269 (43105)	76,600 (24,208)	<0.001
Anthranilic.acid	907 (652)	584 (546)	<0.001
Caffeine	390,620 (423,610)	104,374 (182,222)	<0.001
Carnitine	1,466,179 (347,252)	1,266,450 (344,290)	<0.001
Choline	832,622 (250,867)	684,913 (258,019)	<0.001
Creatinine	8,469,468 (1,899,658)	6,979,915 (1,742,566)	<0.001
Cystine	60,124 (48,938)	138,303 (46,830)	<0.001
Glutamic Acid	1,253,120 (695,248)	412,072 (216,153)	<0.001
Methionine	151,370 (67,815)	193,134 (57,982)	<0.001
Homoserine	514,689 (147,898)	409,621 (122,108.97)	<0.001
Hypoxanthine	193,253 (261,932)	402,543 (428,699)	<0.001
Isoleucine	6,439,637 (2,162,432)	4,789,288 (1,186,628)	<0.001
Kynurenic.acid	3,356 (1,581)	2,307 (1,282)	<0.001
Kynurenine	11,791 (4,565)	9,463 (3,686)	<0.001
L.Alloisoleucine	7,108,170 (2,376,444)	5,287,973 (1,312,879)	<0.001
Leucine	6,414,989 (2,157,010)	4,757,621 (1,186,610)	<0.001
Lysine	1,222,998 (340,572)	975,897 (311,635)	<0.001
Proline	621,213 (202,214)	438,417 (170,220)	<0.001
Norleucine	5,959,529 (1,997,544)	4,423,950 (1,098,147)	<0.001
Ornithine	595,850 (220,847)	341,889 (133,200)	<0.001
Phenylalanine	3,274,991 (993,388)	2,463,481 (551,597)	<0.001

levels were measured across participants. For instance, Alanine levels were 377,036 in patients versus 519,367 in healthy individuals, and Histidine levels were 1,080,024 in patients compared to 1,210,992 in healthy individuals, Tryptophan levels were 978,986 in patients compared to 1,196,444 in healthy individuals. A comprehensive analysis of all metabolic data can be found in **Table 1**.

Classification performance

The performance of six machine learning models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and Logistic Regression (LR) was evaluated using sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). The detailed results are summarized in **Figure 1**. Among the classifiers, RF achieved the highest accuracy (0.90) and specificity (0.85), with sensitivity equal to KNN (0.95). KNN also demonstrated high sensitivity (0.95) and a strong NPV (0.94), though its specificity was lower (0.75). LR showed a balance between specificity (0.90) and sensitivity (0.86) with an overall accuracy of 0.88. NB and SVM achieved similar accuracy scores (0.83), with NB showing slightly higher sensitivity (0.90) but lower specificity (0.75). The DT model had the lowest sensitivity (0.76) but the highest PPV (0.89), indicating strong performance in positive class prediction. Overall, RF demonstrated the highest classification performance, followed by LR and KNN, based on accuracy and sensitivity.

Feature importance analysis

To determine the most influential features in classification performance, the Percentage Increase in Mean Squared Error (%IncMSE) was used as the importance metric. **Figure 2** presents the ranking of key features based on this measure, identifying the top 10%. Among them, glutamic acid (16.31) exhibited the highest value, emphasizing its strong contribution to model accuracy. Other significant features within the top 10% included ketocholesterol (12.45), cysteine (10.60), age (10.49), ornithine (6.65), succinate (5.96), acetylcarnitine (5.96), asparagine (5.38), tryptophan (5.25), and palmitic acid (5.24), all of which played a crucial role in model performance.

Discussion

In this study, we explored metabolic profiling as a potential tool for distinguishing breast cancer patients from healthy individuals. Given that metabolism is a hallmark of cancer, our aim was to identify metabolic signatures that could serve as biomarkers for breast cancer detection and potential therapeutic targets. We identified distinct metabolic alterations in the plasma samples of breast cancer patients compared to healthy individuals. Notably, the concentration of several amino acids, including alanine (Ala), histidine (His), tryptophan (Trp), tyrosine (Tyr), methionine (Met), and proline (Pro), was significantly reduced in breast cancer patients. Using machine learning models, we successfully classified individuals based on their metabolic profiles. Among the models tested, RF achieved the highest accuracy (0.90) and specificity (0.85), while KNN exhibited the highest sensitivity (0.95). The high sensitivity of RF and KNN suggests their potential for detecting cancer cases with minimal false negatives, which is crucial for early diagnosis. However, KNN had a higher false-positive rate (specificity: 0.75), which may lead to unnecessary follow-ups. LR showed a well-balanced performance (accuracy: 0.88, specificity: 0.90, sensitivity: 0.86), making it a suitable choice when minimizing both false positives and false negatives. In contrast, NB and SVM performed moderately (accuracy: 0.83), with NB favoring sensitivity (0.90) at the cost of specificity (0.75), making it more suitable for screening applications where missing positive cases is a greater concern than false positives. Feature importance analysis highlighted glutamic acid, 7-ketocholesterol (7-KC), cystine, age, ornithine, succinate, acetylcarnitine, asparagine, tryptophan, and palmitic acid as the most significant contributors to model accuracy [35,36].

Our findings align with previous research demonstrating altered amino acid metabolism in breast cancer. We observed a significant decrease in plasma levels of Ala, His, Trp, Tyr, Met, and Pro in breast cancer patients, consistent with a study by Shen *et al.*, which reported reduced levels of these amino acids in triple-negative breast cancer (TNBC) patients. Additionally, in ER+/PR+ breast cancer patients, levels of Ala and His were significantly lower than in healthy controls [37]. Another study analyzing plasma from patients with luminal A,

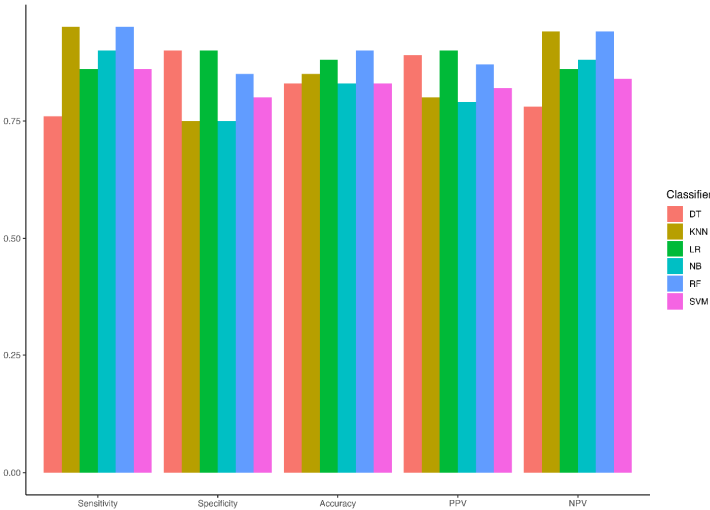


Figure 1. Performance Metrics of Machine Learning Models on the Test Dataset.

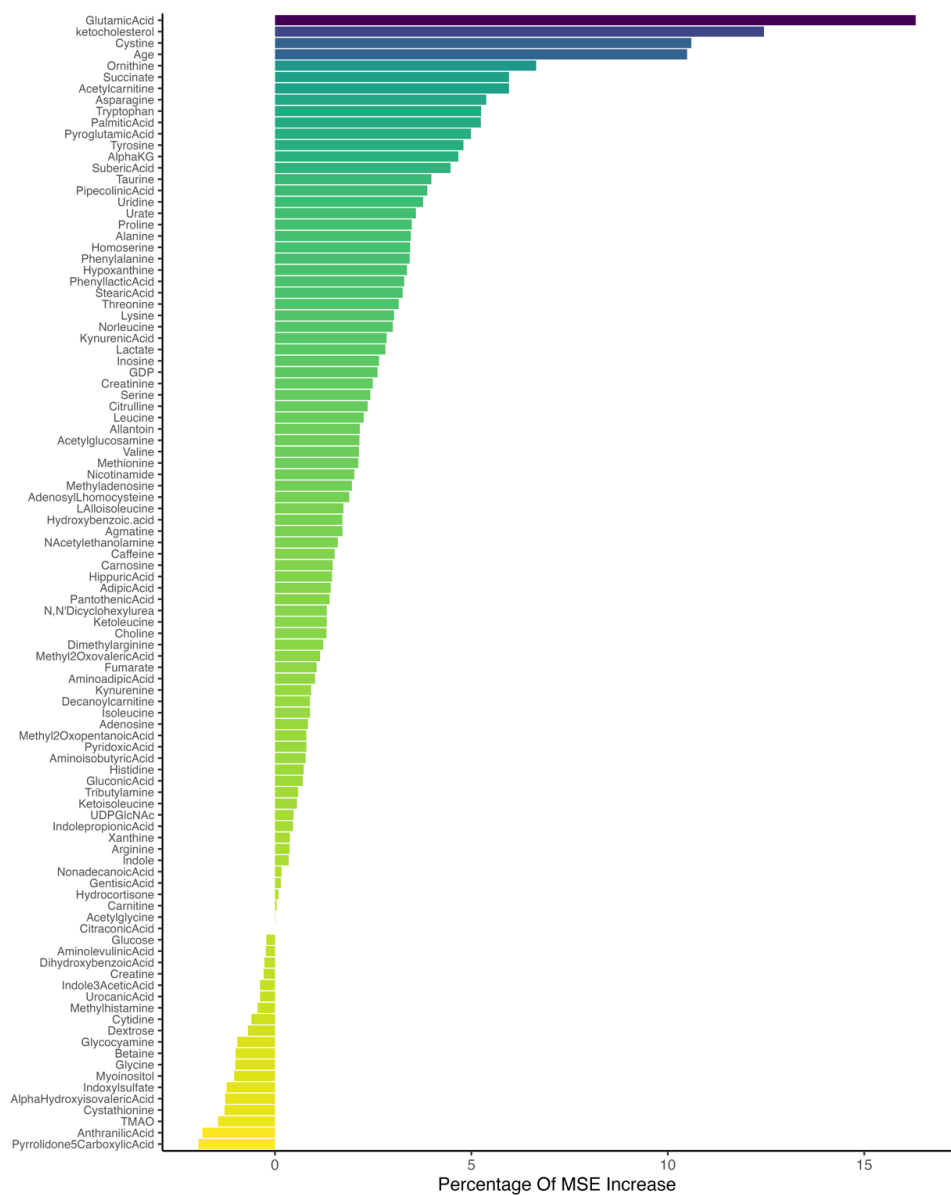


Figure 2. Ranking of the Most Important Features Based on %IncMSE.

TNBC, and HER2-positive breast cancer also observed a significant decrease in Trp levels compared to healthy individuals [38]. These findings suggest that amino acid metabolism is systematically altered in breast cancer, potentially due to increased tumor cell consumption and metabolic reprogramming.

Feature importance analysis identified glutamic acid as one of the most significant metabolic markers in breast cancer. Glutamic acid is a central metabolite in cancer metabolism, particularly in the glutaminolysis pathway, which fuels tumor growth by supplying carbon and nitrogen sources for biosynthesis and redox balance maintenance [39,40]. Studies indicate that glutaminase (GLS), the enzyme converting glutamine to glutamic acid, is upregulated in aggressive breast cancers such as TNBC, promoting proliferation and survival. Additionally, increased glutamic acid levels have been linked

to therapy resistance in endocrine-resistant breast cancer [41]. Given these roles, targeting glutaminase or glutamic acid metabolism could represent a promising therapeutic strategy, particularly in cancers that exhibit glutamine addiction.

Our study also identified 7-KC as a key metabolic feature in breast cancer. 7-KC is an oxidized cholesterol derivative implicated in cancer progression and drug resistance [42]. Research suggests that 7-KC reduces doxorubicin cytotoxicity in ER+ MCF-7 cells by upregulating P-glycoprotein via an ER α - and mTOR-dependent pathway, leading to reduced intracellular drug accumulation and decreased efficacy [42]. Furthermore, 7-KC has been shown to modulate tamoxifen response, slightly reducing its effect in ER+ cells while enhancing it in ER-negative BT-20 cells. It also promotes cancer cell migration and invasion, suggesting a role in

breast cancer metastasis [43]. However, recent studies indicate that 7-KC-loaded phosphatidylserine liposomes exhibit anticancer potential by inducing apoptosis and autophagy in melanoma and breast adenocarcinoma models, highlighting its dual role in cancer biology [44]. Overall, 7-KC influences both drug resistance and tumor progression, underscoring the need for further research into its therapeutic potential. Cystine, the oxidized dimer of cysteine, plays a crucial role in breast cancer progression, particularly in TNBC [45]. Our study identified cystine as a significant metabolic marker, aligning with studies showing that TNBC cells, especially mesenchymal subtypes, exhibit a strong dependency on cystine for survival. TNBC cells are highly sensitive to cystine deprivation, which induces programmed necrosis through TNF α and the MEKK4-p38-Noxa pathways. Interestingly, inhibiting these pathways reduces cell death, suggesting potential therapeutic strategies targeting cystine metabolism in aggressive breast cancer subtypes [45].

Limitations

Despite the promising findings of our study, several limitations should be acknowledged. First, our sample size may not fully capture the heterogeneity of breast cancer, and larger, multi-center studies are needed to validate the metabolic signatures identified. Second, while we used machine learning models for classification, further optimization and external validation are required to confirm their clinical applicability. Incorporating additional datasets from independent cohorts could improve model generalizability. Third, our study relied on plasma metabolomics, which provides valuable insights but does not directly reflect tumor-specific metabolic changes. Integrating tumor tissue metabolomics or single-cell analysis could offer a more comprehensive understanding of metabolic alterations. Finally, although we identified key metabolic features linked to breast cancer, functional studies are needed to elucidate their precise mechanistic roles in tumor progression and therapy resistance.

Conclusion and Future Directions

This study highlights the potential of metabolic profiling and machine learning in breast cancer detection and classification. Our findings reveal significant alterations in amino acid metabolism, with decreased levels of alanine, histidine, tryptophan, tyrosine, methionine, and proline in breast cancer patients compared to healthy individuals. Feature importance analysis identified glutamic acid, 7-KC, and cystine as key metabolic markers, providing new insights into cancer metabolism and therapeutic targeting. Among the machine learning models tested, RF demonstrated the highest classification performance, followed by LR and KNN. These findings highlight the potential of metabolic markers and machine learning approaches in advancing non-invasive breast cancer diagnostics. Future studies should focus on validating these findings in larger cohorts, integrating multi-omics approaches, and exploring the mechanistic role of identified metabolites in breast cancer progression and treatment resistance.

Ethics Statement

This study was approved by the Inonu University Health Sciences Non-Interventional Clinical Research Ethics Committee and conducted in compliance with institutional guidelines, local regulations, and the ethical principles outlined in the Declaration of Helsinki. Prior to participation, all individuals provided written informed consent.

Author Contributions

F.M., Conceptualization, Formal Analysis, Methodology, Writing—original draft, Writing—review and editing, P.H., Writing—review and editing, Visualization, S.A. Software, Writing—review and editing.

Funding

Not applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Huang J, Chan PS, Lok V, Chen X, Ding H, Jin Y, et al. Global incidence and mortality of breast cancer: a trend analysis. *Aging (Albany NY)*. 2021 Feb 11;13(4):5748–803.
2. Kinnel B, Singh SK, Oprea-Ilie G, Singh R. Targeted Therapy and Mechanisms of Drug Resistance in Breast Cancer. *Cancers (Basel)*. 2023 Feb 19;15(4):1320.
3. Wu W, Zhao S. Metabolic changes in cancer: beyond the Warburg effect. *Acta Biochim Biophys Sin (Shanghai)*. 2013 Jan;45(1):18–26.
4. Payne KK. Cellular stress responses and metabolic reprogramming in cancer progression and dormancy. *Semin Cancer Biol*. 2022 Jan;78:45–8.
5. Pavlova NN, Thompson CB. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab*. 2016 Jan 12;23(1):27–47.
6. Willmann L, Schlimpert M, Halbach S, Erbes T, Stickeler E, Kammerer B. Metabolic profiling of breast cancer: Differences in central metabolism between subtypes of breast cancer cell lines. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2015 Sep 1;1000:95–104.
7. Choi J, Jung WH, Koo JS. Metabolism-related proteins are differentially expressed according to the molecular subtype of invasive breast cancer defined by surrogate immunohistochemistry. *Pathobiology*. 2013;80(1):41–52.
8. Kim S, Kim DH, Jung WH, Koo JS. Expression of glutamine metabolism-related proteins according to molecular subtype of breast cancer. *Endocr Relat Cancer*. 2013 May 21;20(3):339–48.
9. Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ, et al. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat Genet*. 2011 Jul 31;43(9):869–74.
10. Choi J, Kim ES, Koo JS. Expression of Pentose Phosphate Pathway-Related Proteins in Breast Cancer. *Dis Markers*. 2018 Feb 25;2018:9369358.
11. Hilvo M, Denkert C, Lehtinen L, Müller B, Brockmöller S, Seppänen-Laakso T, et al. Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res*. 2011 May 1;71(9):3236–45.
12. Nakahara R, Maeda K, Aki S, Osawa T. Metabolic adaptations of cancer in extreme tumor microenvironments. *Cancer Sci*. 2023 Apr;114(4):1200–7.
13. Haghazad T, Khorsand B, Razavi SA, Hedayati M. A computational approach to assessing the prognostic implications of BRAF and RAS mutations in patients with papillary thyroid carcinoma. *Endocrine*. 2024 Nov;86(2):707–22.
14. Khorsand B, Khammari A, Shirvanizadeh N, Zahiri J, Arab SS.

1. OligoCOOL: A mobile application for nucleotide sequence analysis. *Biochem Mol Biol Educ.* 2019 Mar;47(2):201–6.
15. Khorsand B, Savadi A, Naghibzadeh M. Parallelizing Assignment Problem with DNA Strands. *Iran J Biotechnol.* 2020 Jan 1;18(1):e2547.
16. Hesami Z, Sabzehali F, Khorsand B, Alipour S, Sadeghi A, Asri N, et al. Microbiota as a state-of-the-art approach in precision medicine for pancreatic cancer management: A comprehensive systematic review. *iScience.* 2025 Mar 28;28(5):112314.
17. Houri H, Aghdaei HA, Firuzabadi S, Khorsand B, Soltanpoor F, Rafieepoor M, et al. High Prevalence Rate of Microbial Contamination in Patient-Ready Gastrointestinal Endoscopes in Tehran, Iran: an Alarming Sign for the Occurrence of Severe Outbreaks. *Microbiol Spectr.* 2022 Oct 26;10(5):e0189722.
18. Irankhah L, Khorsand B, Naghibzadeh M, Savadi A. Analyzing the performance of short-read classification tools on metagenomic samples toward proper diagnosis of diseases. *J Bioinform Comput Biol.* 2024 Oct;22(5):2450012.
19. Kharaghani AA, Harzandi N, Khorsand B, Rajabnia M, Kharaghani AA, Houri H. High prevalence of Mucosa-Associated extended-spectrum β -Lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* among Iranian patients with inflammatory bowel disease (IBD). *Annals of Clinical Microbiology and Antimicrobials.* 2023;22(1):86.
20. Shiralipour A, Khorsand B, Jafari L, Salehi M, Kazemi M, Zahiri J, et al. Identifying Key Lysosome-Related Genes Associated with Drug-Resistant Breast Cancer Using Computational and Systems Biology Approach. *Iran J Pharm Res.* 2022 Oct 15;21(1):e130342.
21. Soltanyzadeh M, Khorsand B, Baneh AA, Houri H. Clarifying differences in gene expression profile of umbilical cord vein and bone marrow-derived mesenchymal stem cells; a comparative in silico study. *Informatics in Medicine Unlocked.* 2022 Jan 1;33:101072.
22. Zareei S, Khorsand B, Dantism A, Zareei N, Asgharzadeh F, Zahraee SS, et al. PeptiHub: a curated repository of precisely annotated cancer-related peptides with advanced utilities for peptide exploration and discovery. *Database (Oxford).* 2024 Sep 20;2024:baae092.
23. Khorsand B, Asadzadeh Aghdaei H, Nazemalhosseini-Mojarad E, Nadalian B, Nadalian B, Houri H. Overrepresentation of Enterobacteriaceae and *Escherichia coli* is the major gut microbiome signature in Crohn's disease and ulcerative colitis; a comprehensive metagenomic analysis of IBDMDb datasets. *Front Cell Infect Microbiol.* 2022 Oct 4;12:1015890.
24. Razavi SA, Khorsand B, Salehipour P, Hedayati M. Metabolite signature of human malignant thyroid tissue: A systematic review and meta-analysis. *Cancer Med.* 2024 Apr;13(8):e7184.
25. Sadeghnezhad E, Sharifi M, Zare-maivan H, Khorsand B, Zahiri J. Cross talk between energy cost and expression of Methyl Jasmonate-regulated genes: from DNA to protein. *Journal of Plant Biochemistry and Biotechnology.* 2019 Jun 1;28(2):230–43.
26. Samandari Bahraseman MR, Khorsand B, Esmailzadeh-Salestani K, Sarhadi S, Hatami N, Khaleghdoust B, et al. The use of integrated text mining and protein-protein interaction approach to evaluate the effects of combined chemotherapeutic and chemopreventive agents in cancer therapy. *PLoS One.* 2022 Nov 11;17(11):e0276458.
27. Khorsand B, Savadi A, Zahiri J, Naghibzadeh M. Alpha influenza virus infiltration prediction using virus-human protein-protein interaction network. *Math Biosci Eng.* 2020 Apr 15;17(4):3109–29.
28. Khorsand B, Savadi A, Naghibzadeh M. Comprehensive host-pathogen protein-protein interaction network analysis. *BMC Bioinformatics.* 2020 Sep 10;21(1):400.
29. Khorsand B, Savadi A, Naghibzadeh M. SARS-CoV-2-human protein-protein interaction network. *Inform Med Unlocked.* 2020;20:100413.
30. Jasbi P, Wang D, Cheng SL, Fei Q, Cui JY, Liu L, et al. Breast cancer detection using targeted plasma metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2019 Jan 15;1105:26–37.
31. Yagin FH, Gormez Y, Al-Hashem F, Ahmad I, Ahmad F, Ardigò LP. Biomarker discovery and development of prognostic prediction model using metabolomic panel in breast cancer patients: a hybrid methodology integrating machine learning and explainable artificial intelligence. *Front Mol Biosci.* 2024 Dec 18;11:1426964.
32. Hourfar H, Taklifi P, Razavi M, Khorsand B. Machine Learning-Driven Identification of Molecular Subgroups in Medulloblastoma via Gene Expression Profiling. *Clin Oncol (R Coll Radiol).* 2025 Apr;40:103789.
33. Khorsand B, Vaghf A, Salimi V, Zand M, Ghoreishi SA. Enhancing ischemic stroke management: leveraging machine learning models for predicting patient recovery after Alteplase treatment. *Brain Inj.* 2025;39(8):671–7.
34. Khorsand B, Rajabnia M, Jahanian A, Fathy M, Taghvaei S, Houri H. Enhancing the accuracy and effectiveness of diagnosis of spontaneous bacterial peritonitis in cirrhotic patients: A machine learning approach utilizing clinical and laboratory data. *Adv Med Sci.* 2025 Mar;70(1):1–7.
35. Hamdan D, Nguyen TT, Leboeuf C, Meles S, Janin A, Bousquet G. Genomics applied to the treatment of breast cancer. *Oncotarget.* 2019 Jul 30;10(46):4786–801.
36. Sinha S, Sharma S, Vora J, Shrivastava N. Emerging role of sirtuins in breast cancer metastasis and multidrug resistance: Implication for novel therapeutic strategies targeting sirtuins. *Pharmacol Res.* 2020 Aug;158:104880.
37. Shen J, Yan L, Liu S, Ambrosone CB, Zhao H. Plasma metabolomic profiles in breast cancer patients and healthy controls: by race and tumor receptor subtypes. *Transl Oncol.* 2013 Dec 1;6(6):757–65.
38. Díaz-Beltrán L, González-Olmedo C, Luque-Caro N, Díaz C, Martín-Blázquez A, Fernández-Navarro M, et al. Human Plasma Metabolomics for Biomarker Discovery: Targeting the Molecular Subtypes in Breast Cancer. *Cancers (Basel).* 2021 Jan 5;13(1):147.
39. Nan D, Yao W, Huang L, Liu R, Chen X, Xia W, et al. Glutamine and cancer: metabolism, immune microenvironment, and therapeutic targets. *Cell Commun Signal.* 2025 Jan 24;23(1):45.
40. Choi H, Gupta M, Hensley C, Lee H, Lu YT, Pantel A, et al. Disruption of redox balance in glutaminolytic triple negative breast cancer by inhibition of glutamate export and glutaminase. *bioRxiv [Preprint].* 2023 Nov 19:2023.11.19.567663.
41. Demas DM, Demo S, Fallah Y, Clarke R, Nephew KP, Althouse S, et al. Glutamine Metabolism Drives Growth in Advanced Hormone Receptor Positive Breast Cancer. *Front Oncol.* 2019 Aug 2;9:686.
42. Wang CW, Huang CC, Chou PH, Chang YP, Wei S, Guengerich FP, et al. 7-ketocholesterol and 27-hydroxycholesterol decreased doxorubicin sensitivity in breast cancer cells: estrogenic activity and mTOR pathway. *Oncotarget.* 2017 Aug 2;8(39):66033–50.
43. Spalenkova A, Ehrlichova M, Wei S, Peter Guengerich F, Soucek P. Effects of 7-ketocholesterol on tamoxifen efficacy in breast carcinoma cell line models in vitro. *J Steroid Biochem Mol Biol.* 2023 Sep;232:106354.
44. Favero GM, Tortelli Jr TC, Fernandes D, Prestes AP, Kmetiuk

LN, Otake AH, et al. Abstract A50: 7-Ketocholesterol loaded-phosphatidylserine liposome induces cell death, autophagy, and growth inhibition of melanoma and breast adenocarcinoma. Clinical Cancer Research. 2018 Jan 1;24(1_Supplement):A50.

45. Tang X, Ding CK, Wu J, Sjol J, Wardell S, Spasojevic I, et al. Cystine addiction of triple-negative breast cancer associated with EMT augmented death signaling. Oncogene. 2017 Jul 27;36(30):4235–42.