*Research Article* *Open Access*

**Journal of Progressive Medicine and Health Care**

**ISSN: 3070-2305**

***DOI:*** *doi.org/10.63721/26JPMHC115*

# *Integrative Metabolomics and Machine Learning Reveal Predictive Biomarkers in Breast Cancer Patients*

**Fatemeh Mehdikhani[1,2,3], Maryam Refaei[4] and Sajad Alavimanesh[5*]**

[1]Westmead Clinical School, Faculty of Medicine and Health, The University of Sydney, Westmead, New South Wales, [2145], Australia

[2]Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney, Westmead, New South Wales, [2145], Australia

[3]Department of Dermatology, Westmead Hospital, Westmead, New South Wales, [2145], Australia

[4]College of Human Sciences, School of Medical and Health Sciences, Bangor University, Rathbone Building, College Road, Bangor, Gwynedd LL[57] [2]DF, UK

[5]Student Research Committee, Shahrekord University of Medical Sciences, Shahrekord, Iran

***Citation:*** *Fatemeh Mehdikhani, Maryam Refaei, Sajad Alavimanesh (2026) Integrative Metabolomics and Machine Learning Reveal Predictive Biomarkers in Breast Cancer Patients. J. of Pro Med and Hea Care 2(2), 1-12. WMJ/JPMHC-115*

## *Abstract*

***Background:*** *Breast cancer remains one of the most common causes of cancer mortality in women globally. Although therapeutic options have improved, resistance to treatment continues to limit clinical success, highlighting the need for innovative approaches. One defining feature of malignancy is metabolic reprogramming. In breast cancer, tumor cells exhibit metabolic patterns that differ markedly from those of normal tissue, and characterizing these differences may help uncover new therapeutic targets aimed at disrupting tumor growth and improving outcomes.*

***Objectives:*** *This study evaluated whether multiple machine learning methods could differentiate the metabolic signatures of breast cancer patients from those of healthy individuals, with the goal of identifying metabolite-based targets relevant to precision oncology.*

***Methods:*** *Plasma samples from 102 women with breast cancer and 99 control participants were profiled using targeted liquid chromatography–tandem mass spectrometry (LC-MS/MS). Six classification algorithms were trained and compared, and metabolite contributions to model performance were assessed using mean squared error–based importance measures.*

***Results:*** *Breast cancer samples showed reduced levels of several amino acids, including alanine, histidine, tryptophan, tyrosine, methionine, and proline. Among the evaluated models, Random Forest demonstrated the strongest performance (accuracy = 0.90, specificity = 0.85, sensitivity = 0.95). K- Nearest Neighbors produced similar sensitivity but lower specificity, while Logistic Regression provided balanced performance (specificity = 0.90, sensitivity = 0.86; accuracy = 0.88). Naïve Bayes and Support Vector Machine yielded in-*

*termediate accuracy (0.83). The Decision Tree model had the lowest sensitivity (0.76) but the highest positive predictive value (0.89). Feature-importance analysis consistently identified glutamic acid, ketocholesterol, cystine, ornithine, succinate, acetylcarnitine, asparagine, tryptophan, and palmitic acid as influential metabolites.*

***Conclusion:*** *Machine learning–based metabolic profiling revealed several metabolites that may represent actionable metabolic constraints in breast cancer. These findings support the potential of metabolomics-driven modeling to inform targeted interventions and individualized therapeutic strategies.*

***Corresponding author**: Sajad Alavimanesh, Student Research Committee, Shahrekord University of Medical Sciences, Shahrekord, Iran..

**Submitted:** 19.02.2026 **Accepted**: 13.02.2026 **Published**: 10.04.2026

**Keywords:** Breast Cancer, Metabolomics, Machine Learning, Metabolic Targeting, Biomarker Discovery

## 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 among the most common cancers affecting women globally and contributes substantially to cancer-related illness and death. [1] Data indicate that in 2022 approximately 2.3 million women were newly diagnosed with breast cancer and about 670,000 deaths were attributed to the disease. Projections suggest that by 2050, incidence may rise by roughly 38% and mortality by 68%, corresponding to an estimated 3.2 million new cases and 1.1 million deaths worldwide. [2] Although treatment approaches have advanced, treatment resistance continues to hinder progress, and improving patient outcomes remains a major clinical priority. [3, 4] In recent years, tumor metabolism has attracted growing interest, especially in relation to the Warburg effect, whereby cancer cells favor glycolysis for energy generation even when oxygen is available. [5] This metabolic alteration promotes rapid cellular growth and enhances survival in stressful environments, a phenomenon commonly referred to as metabolic reprogramming.[6] Because of its central involvement in tumor development and progression, altered metabolism is now recognized as a defining feature of cancer cells. [7] Multiple metabolic pathways are disrupted in breast cancer, with glycolysis markedly upregulated alongside increased expression of glucose transporters and key glycolytic enzymes. [8,9] In addition to glucose metabolism, certain amino acids, including serine and glutamine, play critical roles in supporting breast cancer growth. [10,11] Moreover, enhanced activity of the pentose phosphate pathway has been observed in breast tumors, where it supports nucleotide production and maintenance of cellular redox balance. [12] Furthermore, lipid metabolism is altered in breast cancer, with tumor cells exploiting fatty acids and upregulating lipid synthesis, including cholesterol uptake, to support membrane formation and energy generation. [13] These metabolic alterations allow cancer cells to survive and proliferate in adverse microenvironments, thereby promoting tumor growth. [14,15] Predictive medicine is an evolving discipline that utilizes technologies such as genomics, bioinformatics, and artificial intelligence to evaluate disease risk and tailor healthcare to individual patients. [16-20] By integrating genetic variants, lifestyle information, and clinical parameters, predictive models can pinpoint individuals at higher risk for conditions such as diabetes, cancer, and cardiovascular diseases, facilitating early detection and tailored preventive interventions. [21-23]

Machine learning algorithms improve the precision of predictions by uncovering complex patterns within

large datasets, thereby facilitating personalized treatment strategies and better patient outcomes. [24,25] Furthermore, pharmacogenomics, a central element of predictive medicine, enables the customization of drug therapies according to an individual's genetic profile, reducing adverse effects while maximizing therapeutic efficacy. As predictive medicine advances, it is transforming healthcare by emphasizing proactive prevention and precision care rather than solely reactive treatment. [26-28] Given the pronounced metabolic differences between breast tumors and normal tissue, profiling tumor metabolism offers significant promise for improving therapeutic strategies. Pinpointing these metabolic vulnerabilities may reveal novel intervention points—so-called metabolic bottlenecks—that can be targeted to impede tumor progression and enhance treatment effectiveness.

This study seeks to characterize the metabolic signatures of breast cancer patients and evaluate their clinical relevance through machine learning techniques. We hypothesize that distinct metabolic patterns can act as predictive metabolic bottlenecks, informing the development of more precise and effective personalized therapies. The novelty of this work lies in combining targeted plasma metabolomics using LC-MS/MS with multiple machine learning algorithms to uncover predictive metabolic markers in breast cancer. Unlike prior studies that examined amino acid or lipid metabolism in isolation, our approach simultaneously evaluates both pathways. Additionally, by introducing the concept of metabolic bottlenecks, we emphasize potential vulnerabilities that may serve as innovative therapeutic targets.

## Methods

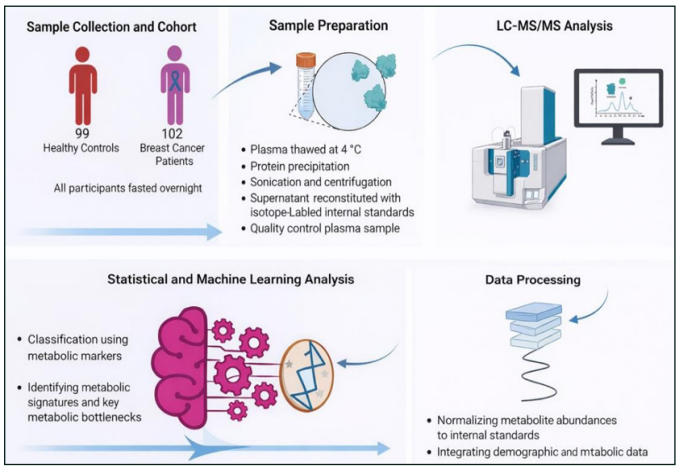


**Figure 1:** Workflow of the study design and analysis

### Study Design and Participants

This study analyzed a dataset containing both demographic and metabolic information from 201 participants, including 102 women diagnosed with breast cancer and 99 age-matched healthy controls. A schematic overview of the study workflow is provided in Figure 1. Biological samples were sourced from Bloodworks Northwest (Seattle, WA) and the Breast Specimen Repository at the Fred Hutchinson Cancer Research Center (FHCRC; Seattle, WA). All breast cancer samples were obtained from FHCRC, whereas control specimens comprised 31 samples from FHCRC and 68 from Bloodworks Northwest.

Because the specimens were commercially acquired, the analysis qualified for IRB exemption. However, informed consent had been secured from all donors by the respective institutions at the time of sample collection. The research adhered to the principles of the Declaration of Helsinki and received approval from the Inonu University Health Sciences Non-Interventional Clinical Research Ethics Committee (protocol no. 2024/5750).

Eligibility criteria for the breast cancer group included women aged 30–75 years with histopathologically confirmed primary breast cancer who had not undergone any cancer treatment, such as chemotherapy, radiotherapy, or hormonal therapy, prior to blood collection. Control participants were healthy women within the same age range who had no history of cancer or chronic metabolic conditions. To minimize metabolic variability, all participants fasted for at least eight hours before blood sampling.

Exclusion criteria for both groups comprised pregnancy or breastfeeding, chronic systemic illnesses (including diabetes, cardiovascular, renal, or hepatic disorders), use of medications known to influence metabolism, and any signs of acute infection or inflammation at the time of sampling. Plasma metabolites were quantified using targeted liquid chromatography–tandem mass spectrometry (LC-MS/MS), ensuring high analytical precision. Following acquisition, samples were stratified according to demographic variables and metabolic features for downstream analysis. [29,30]

Sample Preparation and LC-MS/MS Analysis

Plasma samples were thawed at 4 °C overnight prior to metabolite extraction. For each specimen, 50 µL of plasma was placed into a 2 mL vial and proteins

were precipitated by adding 300 µL of methanol. The mixture was vortex-mixed, incubated at −20 °C, and then subjected to sonication followed by centrifugation. The resulting supernatant was transferred, evaporated to dryness using a vacuum concentrator, and reconstituted in a solution containing stable isotope–labeled internal standards to monitor analytical performance.

To evaluate instrument stability and reduce potential batch effects, a pooled quality control (QC) sample—prepared by combining plasma from both breast cancer patients and controls—was processed in the same manner as study samples and injected periodically throughout the LC-MS/MS sequence.

Metabolomic profiling was performed using a Waters Acquity I-Class UPLC system coupled with a TQS- micro triple quadrupole mass spectrometer. Chromatographic separation was achieved on a Waters Xbridge BEH Amide column maintained at 40 °C with a flow rate of 0.3 mL/min. Analyses were conducted in both positive and negative electrospray ionization modes, each employing mode-specific mobile phase compositions. A gradient elution protocol was implemented to enhance metabolite resolution, and compound identities were verified through spiking experiments using authentic standards. [31]

Data acquisition and processing of multiple reaction monitoring (MRM) signals were carried out with TargetLynx software. All plasma specimens were stored at −80 °C prior to analysis to preserve metabolite integrity. Stable isotope–labeled standards were added to each sample, and metabolite intensities were normalized against these standards to improve reproducibility. Absolute concentrations were not determined; instead, the study relied on relative quantification for all analyses. [32]

## Study Measures Metabolic Markers

Plasma metabolite profiling was performed using LC-MS/MS, focusing on metabolites involved in pathways relevant to breast cancer progression. In total, 99 metabolites were measured and included in subsequent analyses and model construction.

## 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. Several supervised machine learning algorithms were employed for classification analysis. The K-Nearest Neighbors (KNN) method assigns class labels based on similarity to nearby observations in the feature space, with the final class determined by the majority label among the closest neighbors. [33] Support Vector Machine (SVM) separates classes by constructing an optimal decision boundary, or hyperplane, that maximizes the margin between groups, thereby improving generalization and reducing overfitting in complex datasets. [34] The Naïve Bayes algorithm is a probabilistic approach that estimates class membership using Bayes' theorem while assuming that predictors contribute independently to the outcome. [35] Decision Tree models classify observations through a hierarchical structure of sequential decision rules, where the data are repeatedly partitioned into smaller subsets based on feature values. [36]

Random Forest extends this concept by generating an ensemble of decision trees built on random subsets of samples and predictors, with final predictions determined by majority voting across trees, enhancing stability and predictive accuracy. [37] Logistic Regression, a widely used statistical learning method, models the probability of a binary outcome as a function of predictor variables by estimating their contribution to the log-odds of the response, offering both interpretability and reliable performance in classification tasks. [38, 39]

## Model Robustness, Loss Functions, and Data Quality Control

To reduce the risk of overfitting and enhance generalizability, model development incorporated a five- fold cross-validation framework implemented with the caret package in R. The dataset was first divided into training (80%) and testing (20%) subsets, with cross-validation performed only within the training portion. Hyperparameter selection and model optimization were conducted across the cross-validation folds, and final performance was evaluated on the independent test set. Each algorithm optimized its inherent objective function during training: logistic regression minimized binary cross-entropy, support vector machines relied on hinge loss, and tree-based models (decision tree and random forest)

used Gini impurity to guide node splitting. Naïve Bayes estimated parameters through maximization of log- likelihood, whereas K-Nearest Neighbors, being non-parametric, assigned classes according to Euclidean distance without optimizing an explicit loss function [40, 41].

Before model fitting, data quality checks were performed to ensure reliability. The dataset contained no missing values. Potential outliers were screened using boxplots and z-score thresholds ($|z| > 3$). Extreme observations were reviewed for biological plausibility, and when appropriate, logarithmic transformation was applied to mitigate skewness and approximate normal distributions. These preprocessing steps were undertaken to improve data integrity and model robustness. Predictive performance was quantified using accuracy, sensitivity, specificity, positive predictive value, and negative predictive value [42].

### Hyperparameter Tuning

Hyperparameter selection was performed through grid search where applicable. For the K-Nearest Neighbors algorithm, the optimal number of neighbors (k) was chosen based on cross-validation results. For the support vector machine model, a radial basis function kernel was adopted in accordance with prior studies, while other parameters remained at default values. In the random forest model, the number of trees was set to 100, which provided stable and reliable predictions.

### Statistical Analysis

All statistical analyses were conducted in R (version 4.4.1). Two-sided tests were used throughout, and statistical significance was defined as $p < 0.05$. Variable distributions were evaluated using the Shapiro– Wilk test. Group comparisons for categorical variables were performed using the chi-square test, whereas continuous variables were analyzed using either the independent t-test or the Mann–Whitney U test depending on distributional assumptions.

## Results

### Sample Characteristics

The demographic profile and selected metabolically relevant variables of the study population are presented in Table 1. In total, 201 plasma samples were included in the analysis, comprising 102 from patients with breast cancer and 99 from healthy controls. The mean age of participants in the breast cancer group was 55 years, compared with 52 years in the control group. Average metabolite levels were also evaluated across groups. For example, alanine concentrations were lower in patients than in controls (377,036 vs. 519,367), as were histidine (1,080,024 vs. 1,210,992) and tryptophan (978,986 vs. 1,196,444). Detailed results for all measured metabolites are provided in Supplementary Table 1.

Metabolite measurements are reported as normalized relative abundances derived from quality-control–based scaling procedures, in which QC injections were averaged to maintain analytical consistency across runs. Although these values do not represent absolute molar concentrations, they are appropriate for comparative statistical evaluation and machine learning modeling.

### Classification Performance

The predictive performance of six machine learning

**Table 1:** Sample Characteristics

| | Healthy Samples Mean (SD) | Breast cancer patients Mean (SD) | p |
|---|---|---|---|
| Age | 52 (12) | 55 (10) | 0.056 |
| Alanine | 519367 (141095) | 377036 (121443) | <0.001 |
| Histidine | 1210992 (239513) | 1080024 (241916) | <0.001 |
| Tryptophan | 1196444 (393159) | 978986 (291736) | <0.001 |
| Acetyl carnitine | 293355 (167744) | 492791 (225758) | <0.001 |
| Acetylglucosamine | 3440 (1187) | 2668 (1054) | <0.001 |
| Adenosine | 1316 (1224) | 2646 (3235) | <0.001 |
| Tyrosine | 113269 (43105) | 76600 (24208) | <0.001 |
| Anthranilic. Acid | 907 (652) | 584 (546) | <0.001 |

| | | | |
|---|---|---|---|
| Caffeine | 390620 (423610) | 104374 (182222) | <0.001 |
| Carnitine | 1466179 (347252) | 1266450 (344290) | <0.001 |
| Choline | 832622 (250867) | 684913 (258019) | <0.001 |
| Creatinine | 8469468 (1899658) | 6979915 (1742566) | <0.001 |
| Cystine | 60124 (48938) | 138303 (46830) | <0.001 |
| Glutamic Acid | 1253120 (695248) | 412072 (216153) | <0.001 |
| Methionine | 151370 (67815) | 193134 (57982) | <0.001 |
| Homoserine | 514689 (147898) | 409621 (122108.97) | <0.001 |
| Hypoxanthine | 193253 (261932) | 402543 (428699) | <0.001 |
| Isoleucine | 6439637 (2162432) | 4789288 (1186628) | <0.001 |
| Kynurenic.acid | 3356 (1581) | 2307 (1282) | <0.001 |
| Kynurenine | 11791 (4565) | 9463 (3686) | <0.001 |
| L.Alloisoleucine | 7108170 (2376444) | 5287973 (1312879) | <0.001 |
| Leucine | 6414989 (2157010) | 4757621 (1186610) | <0.001 |
| Lysine | 1222998 (340572) | 975897 (311635) | <0.001 |
| Proline | 621213 (202214) | 438417 (170220) | <0.001 |
| Norleucine | 5959529 (1997544) | 4423950 (1098147) | <0.001 |
| Ornithine | 595850 (220847) | 341889 (133200) | <0.001 |
| Phenylalanine | 3274991 (993388) | 2463481 (551597) | <0.001 |

classifiers—K-Nearest Neighbors, Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest, and Logistic Regression—was assessed using sensitivity, specificity, accuracy, positive predictive value, and negative predictive value. A detailed comparison of these metrics is presented in Table 2, while receiver operating characteristic curves for all models are shown in Figure 2.

Among the evaluated approaches, the Random Forest model demonstrated the strongest overall sensitivity (0.95) and the highest negative predictive value (0.94), indicating superior ability to correctly identify breast cancer cases and rule out healthy controls. The Decision Tree model achieved the best specificity (0.90), overall accuracy (0.90), and positive predictive value (0.90), reflecting strong performance in correctly classifying non-cancer cases and confirming positive predictions. Logistic Regression also performed well, showing high sensitivity (0.90), negative predictive value (0.89), and good overall accuracy (0.85). Naïve Bayes displayed balanced performance, with sensitivity of 0.81, specificity of 0.80, and an accuracy of 0.80. Both Naïve Bayes and Support Vector Machine showed comparatively lower sensitivity (0.76) despite maintaining relatively high specificity (0.85). The K-Nearest Neighbors algorithm exhibited the weakest performance overall, with the lowest accuracy (0.78) and positive predictive value (0.77), suggesting reduced reliability in identifying positive cases.

Taken together, Random Forest provided the most robust classification performance, followed by Decision Tree and Logistic Regression, when considering both accuracy and sensitivity metrics.

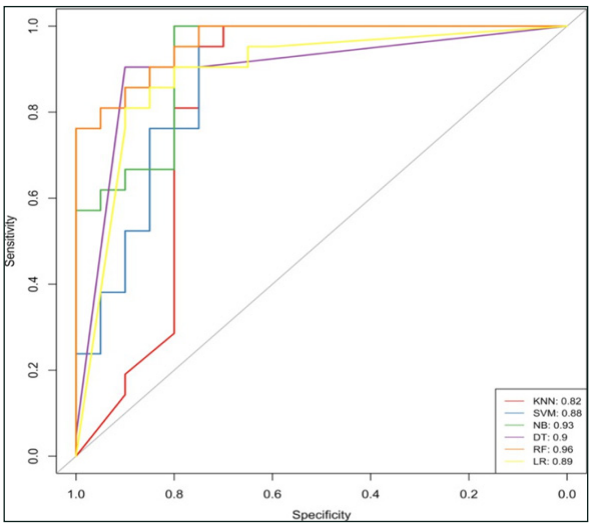


**Figure 2:** ROC plots for each model: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and Logistic Regression (LR).

**Table 2:** Performance Metrics of Machine Learning Models on the Test Dataset

| Classifier | Sensitivity | Specificity | Accuracy | PPV | NPV | F1 | AUC |
|---|---|---|---|---|---|---|---|
| KNN | 0.81 | 0.75 | 0.78 | 0.77 | 0.79 | 0.79 | 0.82 |
| SVM | 0.76 | 0.85 | 0.8 | 0.84 | 0.77 | 0.8 | 0.88 |
| NB | 0.81 | 0.8 | 0.8 | 0.81 | 0.8 | 0.81 | 0.93 |
| DT | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
| RF | 0.95 | 0.75 | 0.85 | 0.8 | 0.94 | 0.87 | 0.96 |
| LR | 0.9 | 0.8 | 0.85 | 0.83 | 0.89 | 0.86 | 0.89 |

## Feature Importance Analysis

To identify the most influential features contributing to classification performance, the Percentage Increase in Mean Squared Error (%IncMSE) was used as the feature importance metric. Figure 3 illustrates the ranking of variables based on this measure. Among all features, glutamic acid (16.31) showed the highest importance value, indicating its strong contribution to model accuracy. Other variables within the top 10% of importance 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). Collectively, these features played a substantial role in improving the predictive performance of the model.

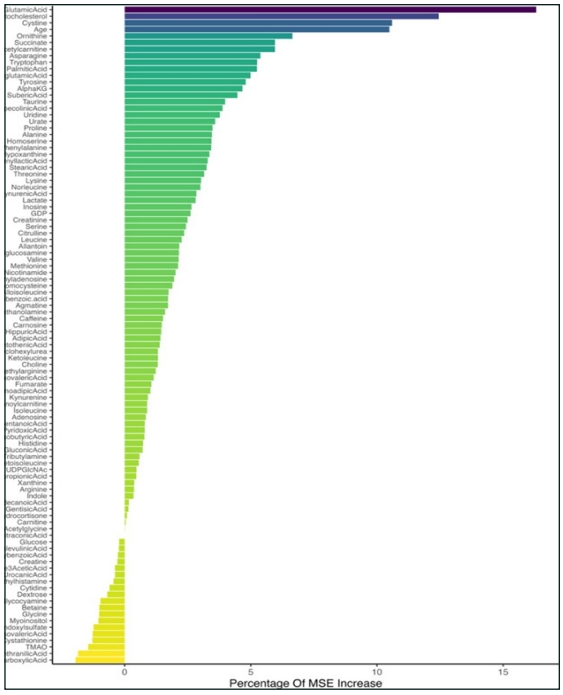


**Figure 3:** Features Ranked by Percentage Increase in Mean Squared Error (%IncMSE). Feature importance was assessed using the %IncMSE metric from the Random Forest model, where higher values indicate a greater contribution to classification accuracy.

## Discussion

Our findings indicate that metabolic profiling can effectively distinguish breast cancer patients from healthy individuals, supporting the notion that altered metabolism is a hallmark of cancer. Plasma analysis revealed significant reductions in several amino acids, including alanine (Ala), histidine (His), tryptophan (Trp), tyrosine (Tyr), methionine (Met), and proline (Pro), in breast cancer patients. These alterations may reflect cancer-related shifts in energy metabolism, protein turnover, and signaling pathways.

Machine learning analyses further demonstrated the potential of metabolic signatures for accurate classification. Random Forest (RF) achieved the highest overall accuracy and specificity, while K-Nearest Neighbors (KNN) showed the highest sensitivity, emphasizing their utility in early detection where minimizing false negatives is critical. However, the lower specificity of KNN suggests a risk of increased false positives, which could lead to unnecessary follow-ups. Logistic Regression (LR) displayed a balanced performance, suggesting its suitability when both false positives and false negatives need to be minimized. Naive Bayes (NB) favored sensitivity at the expense of specificity, indicating its potential role in screening scenarios where missing true cases is more concerning than over-diagnosis.

Feature importance analyses highlighted metabolites such as glutamic acid, 7-ketocholesterol (7-KC), cystine, ornithine, succinate, acetylcarnitine,

asparagine, tryptophan, and palmitic acid, along with age, as key contributors to classification performance. These findings not only reinforce the relevance of metabolic dysregulation in breast cancer but also suggest specific biomarkers that could be further explored for diagnostic or therapeutic purposes. [43, 44]

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. [45] Another study analyzing plasma from patients with luminal A, TNBC, and HER2-positive breast cancer also observed a significant decrease in Trp levels compared to healthy individuals. [46] 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. [47, 48]

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. [49] 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. [50] Research suggests that 7-KC reduces doxorubicin cytotoxicity in ER+ MCF-7 cells by upregulating P-glycoprotein via an ERα- and mTOR- dependent pathway, leading to reduced intracellular drug accumulation and decreased efficacy (50). 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. [51]

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. [52] 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 (53). 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 TNFα 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. [53]

While our study identified several key metabolites with high discriminative potential for breast cancer diagnosis, we acknowledge that our findings remain at the correlation level. The absence of tissue-level validation and mechanistic exploration limits our ability to infer causal relationships between these plasma markers and breast cancer biology. Future studies should focus on confirming whether these differential metabolites mirror tumor-specific metabolic reprogramming through paired plasma–tissue metabolite profiling. Additionally, functional studies involving gene knockdown/overexpression in breast cancer cell lines, as well as pathway analyses targeting glutamine and cholesterol metabolism, will be crucial to elucidate underlying mechanisms and assess the translational relevance of these biomarkers as therapeutic targets.

## 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. Fourth, the breast cancer cohort included patients with different molecular subtypes (e.g., ER/PR/HER2 status) and clinical stages; this heterogeneity, although reflective of real-world populations, may have contributed to variability in the observed metabolic patterns. Finally, beyond age-matching, other potentially influential factors such as BMI, menopausal status, and lifestyle variables (diet, smoking, alcohol) were not available or considered in the dataset. The absence of these covariates may confound metabolic differences and should be addressed in future studies. 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, M.R., Writing–review and editing, Software, 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, et al. (2021) Global incidence and mortality of breast cancer: a trend analysis. Aging (Albany NY) 13: 5748-5803.
2. Kim J, Harper A, McCormack V, Sung H, Houssami N, et al. (2025) Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med 31: 1154-1162.
3. Kinnel B, Singh SK, Oprea-Ilies G, Singh R (2023) Targeted therapy and mechanisms of drug resistance in breast cancer. Cancers 15: 1320.
4. Shiralipour A, Khorsand B, Jafari L, Salehi M, Kazemi M, et al. (2022) Identifying key lysosome- related genes associated with drug-resistant breast cancer using computational and systems biology approach. Iranian Journal of Pharmaceutical Research IJPR 21: e130342.
5. Wu W, Zhao S (2013) Metabolic changes in cancer: beyond the Warburg effect. Acta Biochim Biophys Sin 45:18-26.
6. Payne KK (2022) editor Cellular stress responses and metabolic reprogramming in cancer progression and dormancy. Seminars in cancer biology Elsevier.
7. Pavlova NN, Thompson CB (2016) The emerging hallmarks of cancer metabolism. Cell metabolism

23: 27-47.
8. Willmann L, Schlimpert M, Halbach S, Erbes T, Stickeler E, et al. (2015) Metabolic profiling of breast cancer: Differences in central metabolism between subtypes of breast cancer cell lines. Journal of chromatography B 1000: 95-104.
9. Choi J, Jung W-H, Koo JS (2012) Metabolism-related proteins are differentially expressed according to the molecular subtype of invasive breast cancer defined by surrogate immunohistochemistry. Pathobiology 80: 41-52.
10. Kim S, Kim DH, Jung W-H, Koo JS (2013) Expression of glutamine metabolism-related proteins according to molecular subtype of breast cancer. Endocr Relat Cancer 20: 339-348.
11. Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, et al. (2011) Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nature genetics 43: 869-874.
12. Choi J, Kim E-S, Koo JS (2018) Expression of pentose phosphate pathway-related proteins in breast cancer. Disease markers: 9369358.
13. Hilvo M, Denkert C, Lehtinen L, Müller B, Brockmöller S, et al. (2011) Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. Cancer research 71: 3236-3245.
14. Nakahara R, Maeda K, Aki S, Osawa T (2023) Metabolic adaptations of cancer in extreme tumor microenvironments. Cancer science 114: 1200-1207.
15. Haghzad T, Khorsand B, Razavi SA, Hedayati M. A (2024) computational approach to assessing the prognostic implications of BRAF and RAS mutations in patients with papillary thyroid carcinoma. Endocrine 86: 707-722.
16. Hesami Z, Sabzehali F, Khorsand B, Alipour S, Sadeghi A, et al (2025) Microbiota as a State-of-the- art Approach in Precision Medicine for Pancreatic Cancer Management: A Comprehensive Systematic Review. iScience 112314.
17. Khorsand B, Savadi A, Naghibzadeh M (2020) SARS-CoV-2-human protein-protein interaction network. Informatics in medicine unlocked 20: 100413.
18. Khorsand B, Savadi A, Naghibzadeh M (2020) Comprehensive host-pathogen protein-protein interaction network analysis. BMC bioinformatics 21: 1-22.
19. Khorsand B, Savadi A, Zahiri J, Naghibzadeh M (2020) Alpha influenza virus infiltration prediction using virus-human protein-protein interaction network. Mathematical Biosciences and Engineering 17: 310929.
20. Irankhah L, Khorsand B, Naghibzadeh M, Savadi A (2020) Analyzing the performance of short-read classification tools on metagenomic samples toward proper diagnosis of diseases. Journal of bioinformatics and computational biology 22: 2450012.
21. Khorsand B, Hesami Z, Alipour S, Farmani M, Houri H et al (2025) Harnessing artificial intelligence for detection of pancreatic cancer: a machine learning approach. Clinical and Experimental Medicine 25: 228.
22. Razavi SA, Khorsand B, Salehipour P, Hedayati M (2024) Metabolite signature of human malignant thyroid tissue: A systematic review and meta-analysis. Cancer Medicine 202413: e7184.
23. Zareei S, Khorsand B, Dantism A, Zareei N, Asgharzadeh F, et al. (2024) PeptiHub: a curated repository of precisely annotated cancer-related peptides with advanced utilities for peptide exploration and discovery 20: baae092.
24. Khorsand B, Naderi N, Karimian SS, Mohaghegh M, Aghaahmadi A, et al (2025) Comprehensive transcriptomic analysis of hepatocellular Carcinoma: Uncovering shared and unique molecular signatures across diverse etiologies. Biochemistry and Biophysics Reports 43: 102123
25. Khorsand B, Khammari A, Shirvanizadeh N, Zahiri J, Arab SS, et al (2019) a mobile application for nucleotide sequence analysis. Biochemistry and Molecular Biology Education 47: 201-206
26. Samandari-Bahraseman MR, Hajibarati M, Khorsand B, Soltani N, Esmaeilzadeh-Salestani K, et al (2025) Deciphering the biosynthesis pathway of gamma terpinene cuminaldehyde and para cymene in the fruit of Bunium persicum. Scientific Reports 15: 22438.
27. Khorsand B, Savadi A, Naghibzadeh M (2020) Parallelizing assignment problem with DNA strands. Iranian Journal of Biotechnology 18: e2547.
28. Sadeghnezhad E, Sharifi M, Zare-maivan H, Khorsand B, Zahiri J, et al (2019) Cross talk

between energy cost and expression of Methyl Jasmonate-regulated genes: from DNA to protein. Journal of Plant Biochemistry and Biotechnology 28: 230-243.
29. Jasbi P, Wang D, Cheng SL, Fei Q, Cui JY, et al (2019) Breast cancer detection using targeted plasma metabolomics. Journal of chromatography B 1105: 26-37.
30. Yagin FH, Gormez Y, Al-Hashem F, Ahmad I, Ahmad F, et al (2024) 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. Frontiers in Molecular Biosciences 11: 1426964.
31. Halimi H, Hesami Z, Asri N, Khorsand B, Rostami-Nejad M, et al. (2025) Exploring the biliary microbiome in hepatopancreatobiliary disorders: a comprehensive systematic review of microbial signatures and diagnostic potential. BMC gastroenterology 26: 55.
32. Mousazadeh M, Khorsand B, Modarres Mousavi SM, Mahmoudi Aznaveh H, Fouani MH, et al. (2025) Liposomal Cancer Drug Database (LCDD): a comprehensive resource for liposome research in cancer therapy and diagnosis. Database: baaf042
33. Hourfar H, Taklifi P, Razavi M, Khorsand B (2025) Machine Learning-Driven Identification of Molecular Subgroups in Medulloblastoma via Gene Expression Profiling. Clinical Oncology 103789.
34. Khorsand B, Vaghf A, Salimi V Z M, Ghoreishi SA, et al. (2025) Enhancing ischemic stroke management: leveraging machine learning models for predicting patient recovery after Alteplase treatment. Brain Injury 39: 671-677.
35. Khorsand B, Rajabnia M, Jahanian A, Fathy M, Taghvaei S, et al. (2025) Enhancing the accuracy and effectiveness of diagnosis of spontaneous bacterial peritonitis in cirrhotic patients: a machine learning approach utilizing clinical and laboratory data. Advances in Medical Sciences. 70: 1-7.
36. Hematpour A, Habibi P, Alavimanesh S, Dadkhah K, Babaie K, et al. (2025) Machine learning approach to predict protein-protein interactions between human and hepatitis E virus: revealing links to hepatocellular carcinoma. bioRxiv 23.639757.
37. Jalali S, Dadkhah K, Ghazi MM. (2025) Peritoneal Metastasis Prediction in Gastric Cancer: A Machine Learning Approach. medRxiv 04. 11.25325702.
38. Khorsand B, Ghanbarian E, Rabin L, Sajjadi SA, Ezzati A, et al. (2025) Incremental Value of Plasma Biomarkers in Predicting Clinical Decline Among Cognitively Unimpaired Older Adults: Results from the A4 trial. medRxiv 22.25332015.
39. Khorsand B, Teichrow D, Ghanbarian E, Zheng L, Sajjadi SA, et al. (2025) Scalable Markers for Early Cognitive Decline: Plasma p-tau217, Subjective Cognitive Concerns, and Digital Testing: Results from the A4/LEARN studies. medRxiv10. 14.25338009.
40. Ranjbarzadeh R, Bagherian Kasgari A, Jafarzadeh Ghoushchi S, Anari S, Naseri M, et al. (2021). Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi- modalities brain images. Scientific reports11:10930.
41. Ranjbarzadeh R, Tataei Sarshar N, Jafarzadeh Ghoushchi S, Saleh Esfahani M, Parhizkar M, et al. (2023) MRFE-CNN: Multi-route feature extraction model for breast tumor segmentation in Mammograms using a convolutional neural network. Annals of Operations Research 328: 1021- 1042.
42. Hesami Z, Atashrooz M, Sardarzehi R, Looha MA, Khorsand B, et al. (2025) The oro- and nasopharyngeal microbiota as a revolutionary perspective on mental disorders and related psychopathology: a systematic review and meta-analysis. Journal of Translational Medicine 23: 726.
43. Hamdan D, Nguyen TT, Leboeuf C, Meles S, Janin A, et al. (2019). Genomics applied to the treatment of breast cancer. Oncotarget 10: 4786.
44. Sinha S, Sharma S, Vora J, Shrivastava N. Emerging. (2020) role of sirtuins in breast cancer metastasis and multidrug resistance: Implication for novel therapeutic strategies targeting sirtuins. Pharmacological Research 158: 104880.
45. Shen J, Yan L, Liu S, Ambrosone CB, Zhao H (2013) Plasma metabolomic profiles in breast cancer patients and healthy controls: by race and tumor receptor subtypes. Translational oncology 6: 757.
46. Díaz-Beltrán L, González-Olmedo C, Luque-

Caro N, Díaz C, Martín-Blázquez A, et al. (2021) Human plasma metabolomics for biomarker discovery: Targeting the molecular subtypes in breast cancer. Cancers 13: 147

47. Nan D, Yao W, Huang L, Liu R, Chen X, et al. (2025) Glutamine and cancer: metabolism, immune microenvironment, and therapeutic targets. Cell Communication and Signaling. 23: 45
48. Choi H, Gupta M, Hensley C, Lee H, Lu Y-T, et al. (2023) Disruption of redox balance in glutaminolytic triple negative breast cancer by inhibition of glutamate export and glutaminase. bioRxiv 19.567663.
49. Demas DM, Demo S, Fallah Y, Clarke R, Nephew KP, et al. (2019) Glutamine metabolism drives growth in advanced hormone receptor positive breast cancer. Frontiers in oncology 2: 686.
50. Wang C-W, Huang C-C, Chou P-H, Chang Y-P, Wei S, et al. (2017) 7-ketocholesterol and 27-hydroxycholesterol decreased doxorubicin sensitivity in breast cancer cells: estrogenic activity and mTOR pathway. Oncotarget 8: 66033-66050.
51. Spalenkova A, Ehrlichova M, Wei S, Guengerich FP, Soucek P (2023) Effects of 7-ketocholesterol on tamoxifen efficacy in breast carcinoma cell line models in vitro. The Journal of steroid biochemistry and molecular biology 232: 106354
52. Favero GM, Tortelli Jr TC, Fernandes D, Prestes AP, Kmetiuk LN, et al. (2018) Abstract A50: 7- Ketocholesterol loaded-phosphatidylserine liposome induces cell death, autophagy, and growth inhibition of melanoma and breast adenocarcinoma. Clinical Cancer Research 24: A50-A
53. Tang X, Ding C-K, Wu J, Sjol J, Wardell S, et al. (2017) Cystine addiction of triple-negative breast cancer associated with EMT augmented death signaling. Oncogene 36: 4235-4242.