



Stability-Guided Identification of Compact Molecular Biomarker Signatures for Breast Cancer Subtype Classification

Dr. Rajeshwari Ullagaddi

Assistant Professor, Food Science and Nutrition, Department of Life Sciences, Sri Sathya Sai University for Human Excellence, Kalaburagi, Karnataka, Pin Code: 585101, India, Email ID: rajeshwari.u@sssuhe.ac.in, ORCID ID : 0009-0004-3533-3910

Received Date: 11-06-2026 | Revised Date: 18-08-2026 | Accepted Date: 26-09-2026 | Published Date: 19-10-2026

ABSTRACT

Breast cancer is molecularly heterogeneous and accurate classification of breast cancer subtypes is important to understand the characteristics of the tumor and to facilitate precision medicine. The objective of this study was to determine stable molecular markers, create compact gene-expression signatures and test them for breast cancer subtype classification. A secondary-data analysis was performed using the METABRIC molecular data set (1,898 BC cases and 489 gene-expression features) in a retrospective fashion. Molecular differences were evaluated by the Kruskal–Wallis test and stability guided biomarker identification was performed by 50 iterations of stratified re-sample and re-ranking of features. Multinomial logistic regression with stratified five-fold cross validation was used to compare compact signatures with 5, 10, 15 and 20 genes. There were significant differences in gene-expression between molecular subtypes. Multiple biomarkers, such as GATA3, EGFR, AURKA, CCNE1 and CHEK1, showed selection to be consistent over resampling iterations. The 20-gene signature got a classification accuracy of 73.18%, a balanced accuracy of 66.13% and a macro-F1 score of 0.6760. The full panel of 489 genes obtained a slightly higher macro-F1 score of 0.7055. Recall across molecular subtypes was variable, and lowest for Normal-like tumors. These results show that there are compact molecular signatures that can be useful in subtype classification and can significantly decrease the feature complexity. Independent validation, further stability assessment, and biological verification, however, is needed to determine the reproducibility of the biomarkers and their potential applicability to breast cancer molecular classification.

Keywords: Breast Cancer; Molecular Biomarkers; Gene Expression; Stability-Guided Feature Selection; Molecular Subtype Classification.

1. Introduction

Breast cancer is a highly diverse disease with significant intra-tumoral variations in molecular properties, behavior, clinical outcome and response to therapy. Molecular classification has shed light on this heterogeneity, with the identification of molecular subtypes: Luminal A, Luminal B, HER2-enriched, Basal-like and other molecular groups. These subtypes have distinct patterns of gene expression, cellular signaling pathways and clinical course of the disease, thus highlighting the critical role of precise molecular characterization in the clinic. The identification of molecular subtypes is an integral part of breast cancer studies as it is used in disease stratification and a better understanding of tumour biology (Orrantia-Borunda et al., 2022). Pathological classification has also evolved, with further advances giving greater recognition to the biologically meaningful characteristics of tumors; however, these differences between conventional pathological categories and molecular classifications need to be considered (Rakha et al., 2023). Molecular classification can also help in precision medicine by aiding in the identification of biologically relevant therapeutic targets (Carvalho et al., 2025).

The knowledge of molecular biomarkers in breast cancer is essential to differentiate cancer types and to understand their biological nature. Gene-expression profiles can give information about molecular differences that are not completely captured by traditional clinical and histopathological measurements. Established and emerging biomarkers are increasingly used to improve breast cancer classification and diagnostic interpretation (Beňačka et al., 2022). Despite the substantial variation within the hormone receptor-positive and HER2-negative breast cancer, there is a need for refined classification approaches as demonstrated by molecular studies (Jin et al., 2023). In a similar fashion, molecular analysis of the HER2-positive tumours has defined different biological characteristics, which can be exploited to provide more tailored treatments (Li et al., 2024). Triple-negative breast cancer is also highly molecularly heterogeneous, including distinct genetic biomarkers and biological mechanisms linking to its classification (Lu et al., 2023). The prognostic significance of molecular classification has been reinforced by recent evidence in triple-negative breast cancer (TNBC), a type of breast cancer. Recent evidence also highlighted the importance of molecular classification in triple-negative breast cancer (TNBC).

Nevertheless, accurate classification of breast cancer subtypes remains difficult, due to the high number of potentially informative features of the molecular measurements. Molecular subtyping methods need to reflect differences in tumor biology and should be consistent and easy to interpret between patient cohorts. Magnetic resonance imaging (MRI) phenotype studies have shown links between MRI phenotypes and molecular subtype categories, which suggests that there is biological heterogeneity that can be assessed by different diagnostic measures (Ab Mumin et al., 2022). Likewise, multimodal ultrasound features have been observed to be correlated with breast cancer molecular subtyping (Zhu et al., 2023). Thus, molecular classification systems have grown beyond the traditional diagnostic markers to more holistic systems incorporating biological and computational data (Roy et al., 2023). However, there are still issues with classification reliability, biological interpretation and practical implementation of molecular subtype identification (Zhang, 2023).

Computational identification of biomarkers can help to solve these problems, and to select informative molecular features from complex gene-expression profiles. But biomarkers chosen from one analytical sample are not necessarily stable across varying study population or when the model training conditions change. To compensate for this, stability-guided feature selection has been developed to assess how often features are selected in the same training populations resampled many times. Computational methods have also been proven to be beneficial to explainable artificial intelligence approaches in identifying molecular biomarkers associated with breast cancer subtype classification (Rajpal et al., 2023). Moreover, molecular subtyping studies have underscored the need for reliable molecular signatures to delineate the biologically distinct nature of TNBC (Weng et al., 2024).

Compact biomarker signatures may lead to better model interpretability and decrease the number of molecular measurements needed for classification. A methodological challenge, however, is to strike a balance between the stability of the biomarkers, size of the signature and predictive performance. Hence, in the present study, breast cancer molecular data from secondary METABRIC source has been used to find stable gene-expression biomarkers, construct compact molecular signatures, and assess their efficacy in identifying breast cancer subtyping.

Study Objectives

- To identify stable molecular biomarkers associated with breast cancer subtypes using repeated feature selection and stability assessment of gene-expression profiles.
- To develop compact molecular biomarker signatures and evaluate their performance in classifying

different breast cancer molecular subtypes.

- To assess the relationship between biomarker selection stability, signature size, and classification performance to identify a reproducible and informative molecular biomarker panel.

2. Materials and Methods

2.1. Study Design and Data Source

The research type in this study was retrospective and quantitative, which involved secondary molecular data obtained from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) via Zenodo (Alharbi, 2021). The data contained gene-expression data, mutation profiles, clinical data, and molecular subtype classifications. A computational framework has been introduced to select stable molecular biomarkers, design gene-expression signatures of small size and validate their performance for breast cancer subtype classification.

2.2. Study Population and Molecular Variables

The study comprised 1,904 breast cancer records with 692 variables, 489 of which were gene-expression variables, 173 were mutation variables, and 30 were clinical and classification variables. Molecular subtype (pam50+_claudin-low_subtype) was the main outcome. Luminal A, Luminal B, HER2-enriched, Basal-like, claudin-low, Normal-like and NC were the categories. Of the 1954 records, 6 NC records were excluded because they were not classified into a subtype. Gene-expression measurements were used as the main predictors and clinical and mutation variables were used for descriptive analysis.

2.3. Data Preprocessing and Quality Assessment

The data preprocessing included the evaluation of missing values, duplicate records, the completeness of variables and the distribution of molecular features. Meaningless expression measurements were excluded, and median imputation was used for handling missing expression measurements. Z-score transformation was used to standardize the values of gene-expression. To avoid leakage of information, the estimated pre-processing parameters were used on training data and applied to validation data. The class imbalance was considered by examining the distribution of sub types before classification.

2.4. Stability-Guided Molecular Biomarker Selection

Repeated stratified resampling of L1-regularized multinomial logistic regression was employed in designing stability-guided feature selection. This process of feature selection was repeated 100 times with the re-sampled data set and the frequency of selection of each gene was determined. Biomarkers were sorted based on the consistency of their selection and a threshold of 70% was proposed to identify the highly reproducible ones. To avoid bias in the estimates of classification results, only training partitions were considered for feature selection.

2.5. Development of Compact Molecular Biomarker Signatures

Compact molecular signatures of 5, 10, 15 and 20 highly ranked genes were proposed for comparative evaluation. The signature for each was built based on the identified biomarkers ranked by their stability and obtained from training samples. The classification performance was analysed with the different signature sizes to assess to what extent smaller gene panels would be able to discriminate the subtypes. To maintain interpretability and decrease the molecular complexity, the smallest signature with competitive predictive performance was prioritized.

2.6. Breast Cancer Subtype Classification and Model Validation

The Multinomial logistic regression was chosen as the main classification model to be compared with Random Forest. Stratified nested fivefold cross validation was suggested to assess classification performance without compromising for overfitting. Model tuning and feature selection were done in inner folds, and outer folds were used for independent performance estimates. To measure the overall and subtype-specific classification performance, accuracy, balanced accuracy, macro-averaged precision, recall, F1-score and confusion matrices were chosen.

2.7. Statistical Analysis and Ethical Considerations

Data for categorical variables were reported as frequencies and percentages, and continuous variables were reported as means $\pm$ standard deviation or median with interquartile range. Differences in gene-expression across molecular subtypes were evaluated by Kruskal–Wallis test (with Benjamini–Hochberg correction for

multiple testing). Where applicable, a p value of <0.05 was considered statistically significant. The biomarker selection frequencies and classification performances were compiled based on the repeated validation. Secondary data and publicly available data were used and the data was de-identified and not directly involving the patients. Ethical concerns were taken care of as per institutional needs in case of secondary data research.

3. Results

3.1. Characteristics of the Study Population and Molecular Data

The original data comprised 1904 patient records of breast cancer patients and 692 variables, including 489 gene-expression variables. After discarding six cases that were reported as NC, 1898 cases were available for molecular subtyping. Most subtypes were luminal A (35.77%), luminal B (24.29%) and HER2-enriched (11.59%). Basal-like and claudin-low types were each 10.48% while Normal-like was 7.38% of the population analysed. No data were missing from the gene-expression matrix from the number of retained data. The distribution of the molecular subtypes is shown in Table 1 and Figure 1.

Table 1. Distribution of Breast Cancer Molecular Subtypes in the Study Population

Molecular subtype	Frequency (n)	Percentage (%)
Luminal A	679	35.77
Luminal B	461	24.29
HER2-enriched	220	11.59
Basal-like	199	10.48
Claudin-low	199	10.48
Normal-like	140	7.38
Total	1,898	100.00

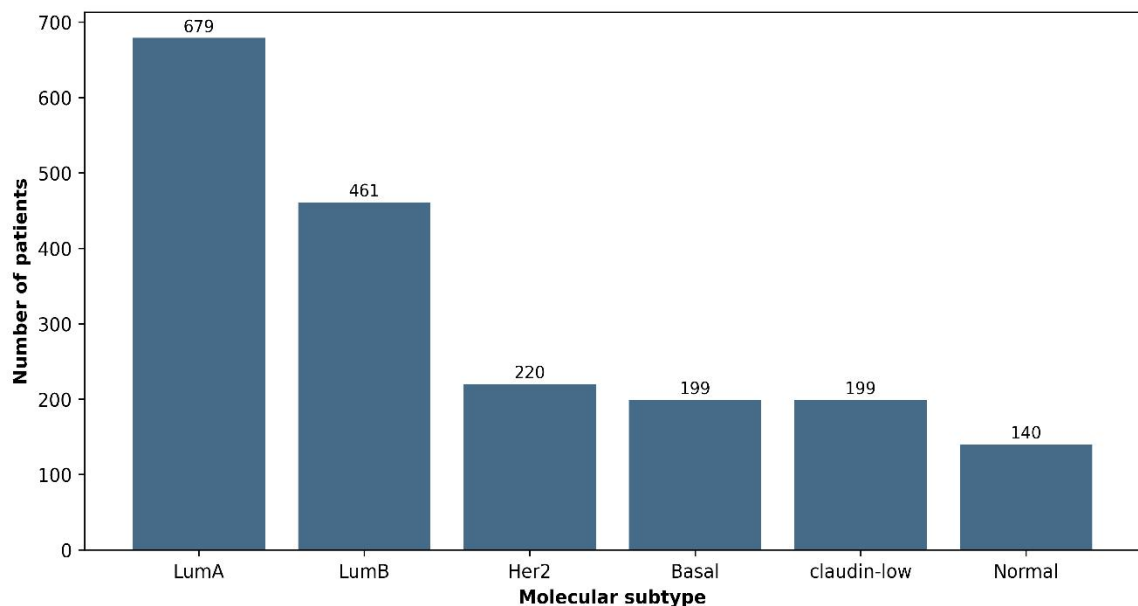


Figure 1. Distribution of Breast Cancer Molecular Subtypes

3.2. Molecular Expression Patterns Across Breast Cancer Subtypes

Comparisons of gene expression revealed significant differences between the breast cancer subtypes.

Several candidate biomarkers (GATA3, EGFR, MAPT, AURKA and CCNE1) showed highly significant differences ($P < 0.001$) by Kruskal–Wallis analysis. GATA3 had the most significant evidence for expression differences associated with subtype, followed by EGFR and MAPT. After Benjamini–Hochberg correction, all the genes mentioned in Table 2 were still statistically significant. The molecular heterogeneity across the six classification groups is shown as subtype-specific standardized expression patterns of selected candidate biomarkers in Figure 2.

Table 2. Statistical Comparison of Candidate Molecular Biomarkers Across Breast Cancer Subtypes

Gene	Kruskal–Wallis H	Adjusted p-value
GATA3	979.77	7.04×10^{-207}
EGFR	895.97	4.84×10^{-189}
MAPT	839.58	5.15×10^{-177}
AURKA	803.83	2.10×10^{-169}
CCNE1	783.89	3.46×10^{-165}
CHEK1	737.35	3.35×10^{-155}
IGF1R	726.23	7.30×10^{-153}
CDC25A	725.41	9.61×10^{-153}
CDK1	711.15	1.04×10^{-149}
E2F2	688.01	9.40×10^{-145}

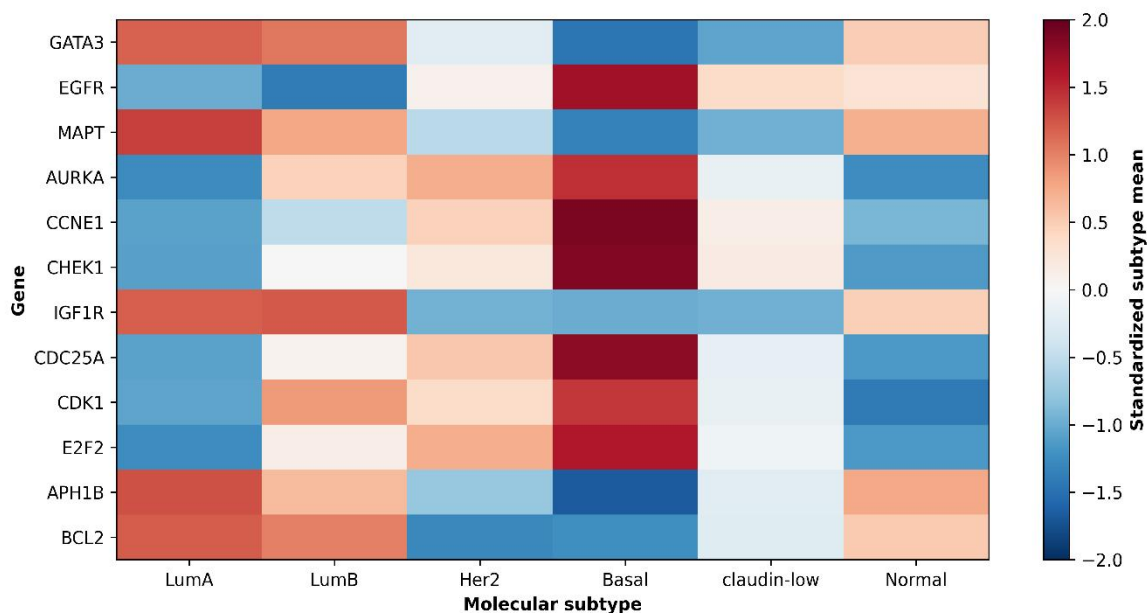


Figure 2. Molecular Expression Patterns Across Breast Cancer Subtypes

3.3. Stability-Guided Identification of Molecular Biomarkers

Candidate molecular biomarkers were evaluated with respect to their consistency by performing 50 stratified rounds of feature selection over 50 training samples. Genes were selected based on the univariate F-statistic and the top 20 genes were kept for each iteration. GATA3, EGFR, AURKA, CCNE1, CHEK1

and IGF1R were identified as completely selection consistent genes. These findings suggest that certain expression features were always well discriminatory with molecular subtype categories across the resampled observations. Table 3 gives the representative high stability biomarkers and Figure 3 shows the selection frequencies of the top 20 genes.

Table 3. Stability-Based Ranking of Selected Molecular Biomarkers

Gene	Selection frequency (%)	Stability category
APH1B	100.0	High
AR	100.0	High
AURKA	100.0	High
BCL2	100.0	High
CCNE1	100.0	High
CDC25A	100.0	High
CDK1	100.0	High
CDK6	100.0	High
CHEK1	100.0	High
E2F2	100.0	High
E2F3	100.0	High
EGFR	100.0	High
ERBB3	100.0	High
GATA3	100.0	High
IGF1R	100.0	High
MAPT	100.0	High
NOTCH1	100.0	High
TTYH1	100.0	High
CDKN2A	92.0	High
MMP7	68.0	Moderate

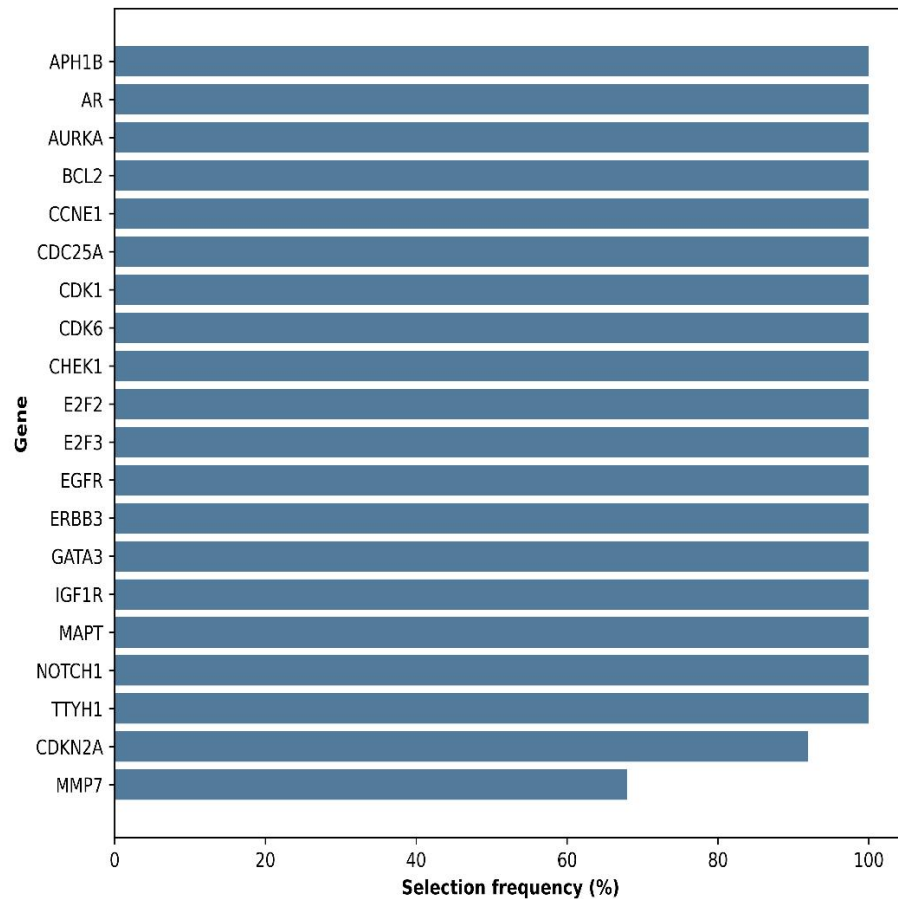


Figure 3. Selection Frequencies of Stable Molecular Biomarkers

3.4. Development and Comparative Evaluation of Compact Biomarker Signatures

The complete 489-gene expression panel was included as a benchmark, and models were tested with gene signatures comprised of 5, 10, 15, and 20 features. Multinomial logistic regression was applied with five-fold stratified cross validation, and the features have been ranked separately within each training folds. The classification performance tended to increase as the number of signatures increased. The 20-gene signature had an accuracy of 73.18%, a balanced accuracy of 66.13% and a macro F1 score of 0.6760. The total feature panel had a macro-F1 score of 0.7055. Table 4 and Figure 4 demonstrate that the 20-gene signature is the most efficient of the compact signatures tested but still achieves good performance.

Table 4. Comparative Classification Performance of Compact Molecular Biomarker Signatures

Number of genes	Accuracy (%)	Balanced accuracy (%)	Macro-F1 score
5	62.01	51.03	0.5125
10	69.55	60.35	0.6178
15	70.92	62.24	0.6362
20	73.18	66.13	0.6760
489 (full panel)	73.45	70.23	0.7055

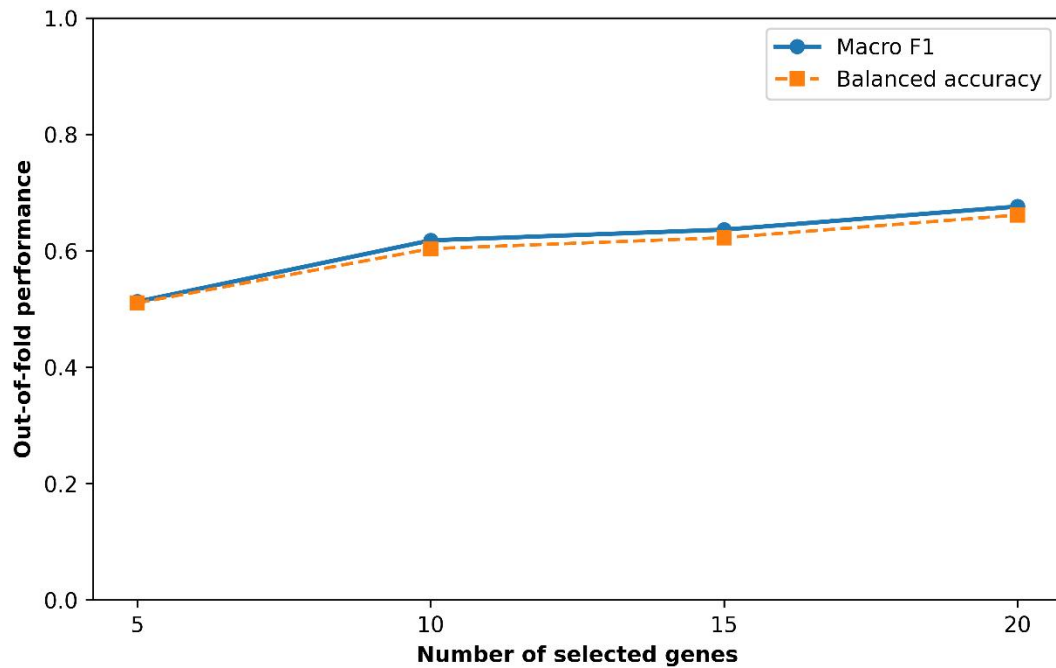


Figure 4. Classification Performance Across Different Biomarker Signature Sizes

3.5. Breast Cancer Subtype Classification Using the Optimal Compact Signature

Among the different signature sizes evaluated, the top-performing compact panel was the 20-gene signature. Subtype specific analysis showed variability in the classification performance. Luminal A had the best F1-score (0.8025), followed by Luminal B (0.7583), and Basal-like breast cancer (0.7268). The performance of tumors with HER2-enriched and claudin-low was intermediate, and the recall for Normal-like tumors was lowest (0.3786). The results of these studies suggest that some molecular biomarkers are better discriminators of breast cancer subtypes than others, and these are compact. The subtype-specific classification metrics are summarized in Table 5, and the confusion matrix based on out-of-fold predictions is shown in Figure 5.

Table 5. Subtype-Specific Classification Performance of the 20-Gene Molecular Signature

Molecular subtype	Precision	Recall	F1-score	Support (n)
Luminal A	0.7602	0.8498	0.8025	679
Luminal B	0.7511	0.7657	0.7583	461
HER2-enriched	0.6364	0.6364	0.6364	220
Basal-like	0.7460	0.7085	0.7268	199
Claudin-low	0.6906	0.6281	0.6579	199
Normal-like	0.6709	0.3786	0.4840	140

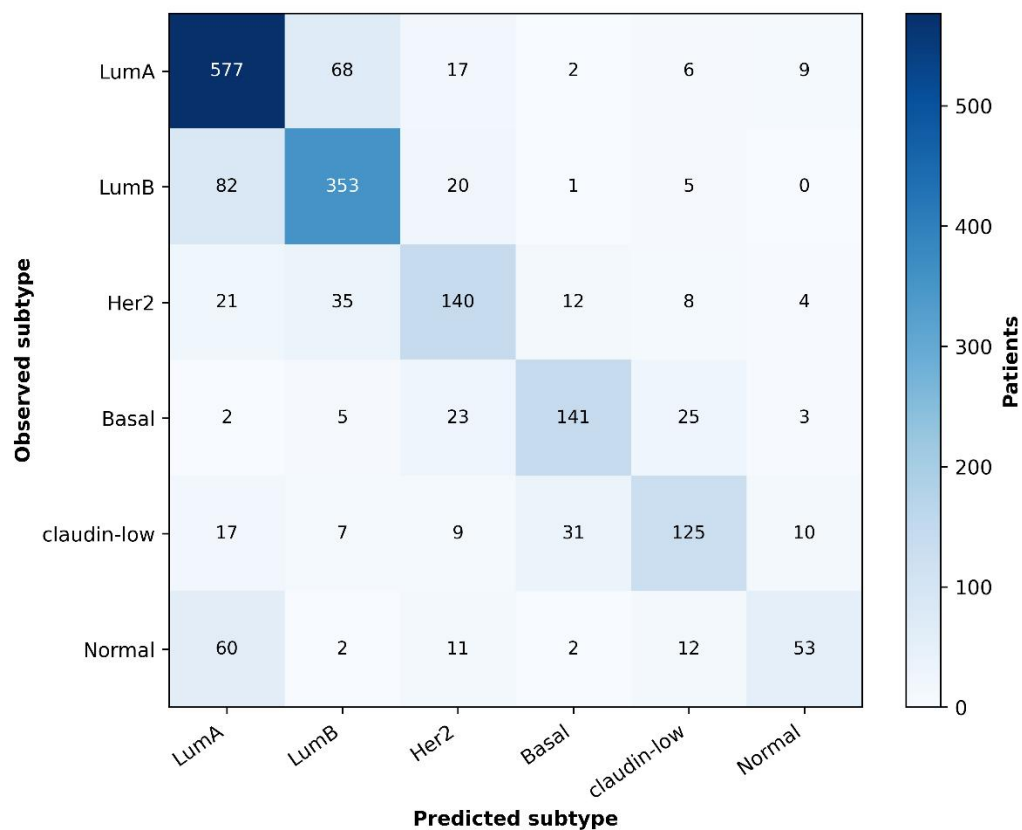


Figure 5. Confusion Matrix of Breast Cancer Molecular Subtype Classification

The analysis is based on 1,898 breast cancer records and 489 gene-expression features for stable molecular biomarkers to classify the subclasses. Important gene-expression differences between the molecular subtypes were found, and some of the biomarkers were highly stable. The 20-gene compact signature was the most efficient panel in terms of the classification performance, reaching 73.18% accuracy and 0.6760 macro-F1 score among all the compact signatures evaluated. Overall, the results indicate the potential for compact molecular biomarker signatures for breast cancer subtype classification, with varying levels of classification in different subtypes.

4. Discussion

In the present work, stable molecular markers were explored, and compact gene-expression signatures were developed for classification of breast cancer subtypes using molecular METABRIC data. 1,898 breast cancer records were analysed together with 489 features in gene-expression. Several genes showed selection in all repeated resampling, which suggests that they may be important to discriminate subtypes. The 20 gene panel showed the best classification performance of all the compact signatures tested with an accuracy of 73.18% and a macro-F1 result of 0.6760. These results indicate that the molecular signatures of reduced molecules can be useful to distinguish the subtype of breast cancer. Computational studies have shown that integration of multi-omics data and deep learning can be useful when predicting molecular subtypes (Choi & Chae, 2023). Multimodal deep learning has also shown promise in finding possible associations between biological aspects and breast cancer subtypes (Liu et al., 2022). More recent classification schemes have also underscored the need to further optimize the molecular subtype boundaries to better enable the biological interpretation (Turova et al., 2025).

There were several genes that had greatly different expressions between the different breast cancer types: GATA3, EGFR, MAPT, AURKA, and CCNE1. These biomarkers can be indicators of molecular mechanisms related to cellular differentiation, growth control and growth of tumor cells. The fact that they are consistently identified indicates that the selected molecular features play a role in discriminating the biologically heterogeneous breast cancer groups. Studies have shown that Luminal A and B tumors differ from each other and have biological features of their own (Lafci et al., 2023). Molecular heterogeneity has also been seen through imaging studies, with different characteristics of diffusion among breast cancer subtypes (Meyer et al., 2022). In addition, subtype-specific immune landscapes in triple negative breast cancer reveal the complexity of the molecular differences and their potential implications in therapy

(Syrnioti et al., 2024).

The analytical framework included stability guided biomarker identification. Several genes were found to have 100% selection frequency in 50 resampling iterations, showing that their discriminatory relevance does not change under the feature-ranking method selected. But, high selection frequency does not mean that there is necessarily biological cause or independent population reproducibility. Measurement variability and consistency under analytical conditions must be taken into consideration for reliable molecular classification. Artificial intelligence has been shown to help enhance the consistency of breast cancer receptor marker interpretation, highlighting the need for reliable diagnostic evaluations (Jung et al., 2024). Machine learning analyses have also pointed to the value of interpretable features in the process of subtype classification (Ma et al., 2022). Further, the discrepancies in the molecular subtype allocation between biopsy and surgical specimens highlight the need to evaluate the consistency of molecular typing.

The more genes selected (from 5 genes to 20 genes) the better performance the classification got. The macro-F1 scores of the 20-gene signature (0.6760) and the full 489-gene panel (0.7055) indicated that reducing the number of features significantly reduced the predictive performance but was still possible. This discovery bolsters the possible utility of smaller molecular signatures that could be used in lieu of larger feature panels. Previous studies have included multimodal deep learning for breast cancer sub-type classification (Ben Rabah et al., 2025), hybrid CNN and transformer architectures (Brahmareddy & Selvan, 2025), patient graph-based prediction (Furtney et al., 2023), and graph neural networks for molecular classification (Li & Nabavi, 2024). However, different study populations, molecular measures, definitions of subtypes and validation methods still limit the possibility of direct comparisons.

Smaller signatures of molecules may be helpful in creating more straightforward and understandable methods of classification in breast cancer. However, there were differences in the recall for each subtype: Normal-like tumors exhibited relatively low recall, suggesting that there are some issues with the reliability of the classification. Alternative computational methods for predicting molecular subtype have been explored in deep learning investigations of mammographic images, as well as histopathological slides (Nissar et al., 2024; Tafavvoghi et al., 2025). The study's drawbacks included its retrospective nature, the small gene-expression panel, the imbalance in the proportion of subtypes and no independent external validation. Moreover, the use of multiple univariate feature ranking does not necessarily define the reproducibility of a multivariable biomarker signature. Independent molecular cohorts nested cross validation and experimental validation are important considerations for future investigations to ensure that compact molecular biomarker signatures are reliable and may have clinical applications.

5. Conclusion

The present study helped to show the potential of stability guided identification of molecular biomarkers for breast cancer subtype classification from METABRIC gene-expression data. When they analysed 1,898 breast cancer records and 489 gene-expression features, the researchers found that there are significant molecular differences between breast cancer subtypes. Multiple genes (GATA3, EGFR, AURKA, CCNE1 and CHEK1) showed high discrimination power and were selected repeatedly in the same resampling methods. These results indicate that it is important to select molecular characteristics that are stable in various training sets. Compact molecular biomarker signatures were developed that showed significant feature reduction without losing effective classification performance. The 20-gene signature had the best accuracy (73.18%) and macro-F1 score (0.6760) among the compact signatures. The full 489-gene panel achieved slightly higher performance although the signature was smaller and simpler and might be easier to work with. There were differences of the classification performance between different molecular subtypes, with some still proving more difficult to classify accurately. Results justify the use of repeated feature selection and computational classification methods to find informative molecular signatures. The retrospective study design, limited gene-expression panel, and lack of external validation, however, confines the generalizability of the results. Additional analyses should be performed with independent patient cohorts, in addition to advanced stability analyses, and experimental validation should be incorporated to confirm the reproducibility and biological value of the selected biomarkers. In conclusion, compact molecular signatures are a promising way to enhance interpretable classification of breast cancer subtypes.

References

1. Ab Mumin, N., Hamid, M. T. R., Wong, J. H. D., Rahmat, K., & Ng, K. H. (2022). Magnetic resonance imaging phenotypes of breast cancer molecular subtypes: a systematic review. *Academic radiology*, 29, S89-S106.
2. Ben Rabah, C., Sattar, A., Ibrahim, A., & Serag, A. (2025). A multimodal deep learning model for the

- classification of breast cancer subtypes. *Diagnostics*, 15(8), 995.
3. Beňačka, R., Szabóová, D., Guľašová, Z., Hertelyová, Z., & Radoňák, J. (2022). Classic and new markers in diagnostics and classification of breast cancer. *Cancers*, 14(21), 5444.
4. Brahmareddy, A., & Selvan, M. P. (2025). TransBreastNet a CNN transformer hybrid deep learning framework for breast cancer subtype classification and temporal lesion progression analysis. *Scientific Reports*, 15(1), 35106.
5. Carvalho, E., Canberk, S., Schmitt, F., & Vale, N. (2025). Molecular subtypes and mechanisms of breast cancer: precision medicine approaches for targeted therapies. *Cancers*, 17(7), 1102.
6. Choi, J. M., & Chae, H. (2023). moBRCA-net: a breast cancer subtype classification framework based on multi-omics attention neural networks. *BMC bioinformatics*, 24(1), 169.
7. Dogra, A. K., Prakash, A., Gupta, S., & Gupta, M. (2025). Prognostic significance and molecular classification of triple negative breast cancer: a systematic review. *European Journal of Breast Health*, 21(2), 101.
8. Furtney, I., Bradley, R., & Kabuka, M. R. (2023). Patient graph deep learning to predict breast cancer molecular subtype. *IEEE/ACM transactions on computational biology and bioinformatics*, 20(5), 3117-3127.
9. Jin, X., Zhou, Y. F., Ma, D., Zhao, S., Lin, C. J., Xiao, Y., ... & Shao, Z. M. (2023). Molecular classification of hormone receptor-positive HER2-negative breast cancer. *Nature Genetics*, 55(10), 1696-1708.
10. Jung, M., Song, S. G., Cho, S. I., Shin, S., Lee, T., Jung, W., ... & Kim, S. W. (2024). Augmented interpretation of HER2, ER, and PR in breast cancer by artificial intelligence analyzer: enhancing interobserver agreement through a reader study of 201 cases. *Breast Cancer Research*, 26(1), 31.
11. Lafci, O., Celepli, P., Öztekin, P. S., & Koşar, P. N. (2023). DCE-MRI radiomics analysis in differentiating luminal A and luminal B breast cancer molecular subtypes. *Academic radiology*, 30(1), 22-29.
12. Li, B., & Nabavi, S. (2024). A multimodal graph neural network framework for cancer molecular subtype classification. *BMC bioinformatics*, 25(1), 27.
13. Li, Y. W., Dai, L. J., Wu, X. R., Zhao, S., Xu, Y. Z., Jin, X., ... & Shao, Z. M. (2024). Molecular characterization and classification of HER2-positive breast cancer inform tailored therapeutic strategies. *Cancer Research*, 84(21), 3669-3683.
14. Liu, T., Huang, J., Liao, T., Pu, R., Liu, S., & Peng, Y. (2022). A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. *Irbm*, 43(1), 62-74.
15. Lu, B., Natarajan, E., Balaji Raghavendran, H. R., & Markandan, U. D. (2023). Molecular classification, treatment, and genetic biomarkers in triple-negative breast cancer: a review. *Technology in cancer research & treatment*, 22, 15330338221145246.
16. Ma, M., Liu, R., Wen, C., Xu, W., Xu, Z., Wang, S., ... & Chen, W. (2022). Predicting the molecular subtype of breast cancer and identifying interpretable imaging features using machine learning algorithms. *European Radiology*, 32(3), 1652-1662.
17. Meyer, H. J., Wienke, A., & Surov, A. (2022). Diffusion-weighted imaging of different breast cancer molecular subtypes: a systematic review and meta-analysis. *Breast Care*, 17(1), 47-54.
18. Nissar, I., Alam, S., Masood, S., & Kashif, M. (2024). MOB-CBAM: A dual-channel attention-based deep learning generalizable model for breast cancer molecular subtypes prediction using mammograms. *Computer Methods and Programs in Biomedicine*, 248, 108121.
19. Orrantia-Borunda, E., Anchondo-Núñez, P., Acuña-Aguilar, L. E., Gómez-Valles, F. O., & Ramírez-Valdespino, C. A. (2022). Subtypes of breast cancer. *Breast Cancer* [Internet].
20. Raghad Alharbi. (2021). Breast Cancer Gene Expression Profiles (METABRIC). <https://doi.org/10.5281/zenodo.4768795>.
21. Rajpal, S., Rajpal, A., Saggar, A., Vaid, A. K., Kumar, V., Agarwal, M., & Kumar, N. (2023). XAI-MethylMarker: Explainable AI approach for biomarker discovery for breast cancer subtype classification using methylation data. *Expert Systems with Applications*, 225, 120130.
22. Rakha, E. A., Tse, G. M., & Quinn, C. M. (2023). An update on the pathological classification of breast cancer. *Histopathology*, 82(1), 5-16.
23. Roy, M., Fowler, A. M., Ulaner, G. A., & Mahajan, A. (2023). Molecular classification of breast cancer. *PET clinics*, 18(4), 441-458.
24. Syrnioti, A., Petousis, S., Newman, L. A., Margioulas-Siarkou, C., Papamitsou, T., Dinas, K., & Koletsa, T. (2024). Triple negative breast cancer: molecular subtype-specific immune landscapes with therapeutic implications. *Cancers*, 16(11), 2094.

25. Tafavvoghi, M., Sildnes, A., Rakaee, M., Shvetsov, N., Bongo, L. A., Busund, L. T. R., & Møllersen, K. (2025). Deep learning-based classification of breast cancer molecular subtypes from H&E whole-slide images. *Journal of pathology informatics*, 16, 100410.
26. Turova, P., Kushnarev, V., Baranov, O., Butusova, A., Menshikova, S., Yong, S. T., ... & Kotlov, N. (2025). The Breast Cancer Classifier refines molecular breast cancer classification to delineate the HER2-low subtype. *NPJ Breast Cancer*, 11(1), 19.
27. Weng, L., Zhou, J., Guo, S., Xu, N., & Ma, R. (2024). The molecular subtyping and precision medicine in triple-negative breast cancer---based on Fudan TNBC classification. *Cancer cell international*, 24(1), 120.
28. Xu, M. M., Yan, M. Y., Chen, M. Y., Chen, M. X., & Gong, M. K. (2025). Expression and subtype discordance between core needle biopsy and surgical specimen in breast cancer. *Journal of Surgical Research*, 307, 42-52.
29. Zhang, X. (2023). Molecular classification of breast cancer: relevance and challenges. *Archives of pathology & laboratory medicine*, 147(1), 46-51.
30. Zhu, J. Y., He, H. L., Jiang, X. C., Bao, H. W., & Chen, F. (2023). Multimodal ultrasound features of breast cancers: correlation with molecular subtypes. *BMC Medical Imaging*, 23(1), 57.