YASJ Young Australians Science Journal

Assessing recurrent neural network accuracy and interpretability in breast cancer classification

Michaela Loukas[1]

[1]Marist Catholic College Penshurst, Sydney, New South Wales, Australia

ABSTRACT

Ribonucleic acid sequencing-based gene expression profiling enables molecular-level cancer diagnosis. However, most machine learning approaches lack the interpretability required for clinical adoption. This investigation examined how accurately a recurrent neural network could classify gene expression data to distinguish malignant from healthy breast tissue, determine molecular subtypes, and reveal key gene expression patterns. It was hypothesised that a multi-task recurrent neural network would achieve high diagnostic accuracy while providing gene-level interpretability. A bidirectional long short-term memory model was trained on The Cancer Genome Atlas Breast Cancer dataset (1,109 malignant and 109 solid tissue normal samples) following extensive preprocessing, including reducing 20,530 genes to 5,000 features. The model achieved 96.5% accuracy for malignancy classification, with 100% sensitivity and 70.6% specificity, and 75.4% accuracy for molecular subtype classification. Explainability analysis identified transferrin, adenosine diphosphate ribosylation factor interacting protein 2, and moesin as key genes, highlighting roles of iron metabolism and cytoskeletal regulation. These findings demonstrate that recurrent neural networks can achieve clinically applicable accuracy for transcriptomic breast cancer diagnosis whilst providing interpretable, gene-level insights, thereby addressing critical barriers to artificial intelligence adoption in oncology.

Young Australians Science Journal
Volume 1, Issue 1, 2026, pp. 9–18
https://doi.org/10.67642/ye0rwq74

Keywords: breast cancer; RNA sequencing; recurrent neural networks; bidirectional long short-term memory; explainable artificial intelligence; SHAP

Student authors: Michaela Loukas, Year 12

Address correspondence to authors@yasj.org.au

Type: Research Article
Open access: Yes
Licence: This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).
DOI: 10.67642/ye0rwq74

YASJ Editors' Note: Reviewers commended the integration of transcriptomic data, deep learning, and explainable artificial intelligence, particularly the connection between model performance and biologically interpretable gene-level findings. They noted that this computational proof of concept uses a single, imbalanced cohort, and that external validation and comparison with simpler baseline models would strengthen the findings.

INTRODUCTION

Breast cancer is the most frequently diagnosed malignancy among women worldwide and remains a leading cause of cancer-related mortality, accounting for approximately 15% of female cancer deaths globally (Zhang et al., 2025). In Australia, annual diagnoses have increased from approximately 13,500 cases in 2000 to over 20,000 in 2024, with mortality also rising despite advances in imaging, pathology, and treatment (Australian Institute of Health and Welfare, 2025). "While the overall 5-year survival rate in stage 1 breast cancer is 99%, it drops to 27% when distant metastases are detected" (Burcu et al., 2023). This highlights the need for earlier, molecular-level diagnostic strategies which detect malignancy before overt morphological change.

Breast cancer is biologically heterogeneous and

comprises distinct molecular subtypes: luminal A, luminal B, human epidermal growth factor receptor 2-enriched (HER2-enriched), and basal-like. These principal subtypes are currently classified using immunohistochemistry (IHC) markers for oestrogen, progesterone, and HER2 receptors (Orrantia-Borunda et al., 2022). However, IHC oversimplifies tumour biology and fails to capture transcriptional heterogeneity. Gene expression profiling reveals earlier and more nuanced distinctions, such as elevated estrogen-receptor 1 (ESR1) and guanine-adenine-thymine-adenine binding protein 3 (GATA3) in luminal A tumours, increased marker of proliferation Ki-67 (MKI67) in luminal B tumours, erb-b2 receptor tyrosine kinase 2 (ERBB2) amplification in HER2-enriched disease, and tumour protein p53 (TP53) mutations in basal-like cancers (Orrantia-Borunda et al., 2022). These transcriptional signatures often precede phenotypic changes, positioning gene-level analysis as a powerful tool for early detection.

Ribonucleic acid sequencing (RNA-seq) enables high-resolution transcriptomic profiling and surpasses microarrays in sensitivity, range, and the ability to detect low-abundance and novel transcripts (Mantione et al., 2014). Advances in high-throughput sequencing system (HiSeq) technologies have substantially reduced costs (Levy & Boone, 2019). Despite its biological value, clinical adoption remains limited due to the computational challenges posed by high-dimensional datasets and the need for robust analytical frameworks to extract clinically meaningful insights. (Acedo-Terrades et al., 2025).

Deep learning (DL) approaches have shown promise in cancer classification, particularly in medical imaging (Kim et al., 2021). Convolutional neural networks (CNN) have achieved high diagnostic accuracy using imaging modalities (Yu et al., 2023). However, such methods often rely on high-radiation or late-stage morphological data, limiting their suitability for early detection (Yu et al., 2023). Recurrent neural networks (RNN), particularly those incorporating bidirectional long short-term memory units (Bi-LSTM), are well-suited to modelling complex dependencies within sequential data (Schuster & Paliwal, 1997) and have demonstrated strong predictive performance in oncology (Dasgupta et al., 2025). Nonetheless, most existing RNN applications exclude transcriptomic inputs (Kim et al., 2021).

Another barrier to clinical translation is the limited interpretability of DL models. In oncology, where decisions are high-stakes, opaque "black-box" predictions undermine trust and ethical deployment (Islam et al., 2024). Explainable artificial intelligence (AI) techniques such as SHapley Additive exPlanations (SHAP) enable gene-level attribution, allowing model predictions to be interpreted and contextualised within established biological frameworks (Lundberg & Lee, 2017). While some studies have applied DL to gene expression data, few integrate explainability, and even fewer employ large, well-annotated datasets such as The Cancer Genome Atlas (TCGA) (Alharbi & Vakanski, 2023).

Recent transcriptomic studies have also identified emerging breast cancer biomarkers, including transferrin (TF), adenosine diphosphate ribosylation factor interacting protein 2 (ARFIP2), and moesin (MSN), which are implicated in iron metabolism, cytoskeletal regulation, and treatment resistance but remain absent from standard clinical panels (Alhaddad et al., 2025; Shi et al., 2024). The detection and validation of such subtle, non-canonical signals highlight the need for models that combine predictive accuracy with biological interpretability.

This study therefore asks whether a RNN trained on RNA-seq gene expression data can accurately and interpretably classify breast cancer and its molecular subtypes. The aim of this research was to develop a Bi-LSTM RNN using RNA-seq data from TCGA dataset to distinguish malignant from healthy tissue and identify molecular subtypes while providing gene-level explanations. It was hypothesised that integrating biologically informed preprocessing, recurrent DL architecture, and explainable AI would yield clinically relevant diagnostic accuracy while revealing biologically meaningful gene expression drivers. Collectively, this work demonstrates the potential of interpretable DL to address key limitations in current breast cancer diagnostics and advance the clinical adoption of transcriptomic AI models.

MATERIALS AND METHODS

Dataset selection and computational infrastructure

Gene expression data were obtained from The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) cohort, comprising 1,109 primary breast tumours and 109 solid tissue normal sam-

ples (total n = 1,218), sequenced using the Illumina HiSeq platform (The Cancer Genome Atlas Program, 2018). Sample identities were assigned using TCGA barcode conventions, where "-01" denoted primary tumours, "-06" revealed metastatic samples and "-11" indicated normal breast tissue. All analyses were conducted in Python 3.10 using TensorFlow for DL implementation. Model training and experimentation were performed in Google Colab to enable graphics processing unit acceleration and reproducibility.

Batch effect assessment and data cleansing

To assess latent technical variation, dimensionality reduction was performed using t-distributed stochastic neighbour embedding (t-SNE). Samples were annotated by technical covariates including sequencing centre, tissue source site, vial number, and collection indicators. No batch effects were detected, hence, batch correction was not applied. Principal component analysis (PCA) was additionally used as a confirmatory linear assessment of variance structure. Outlier detection was conducted using Mahalanobis distance to identify samples with multivariate expression profiles deviating from the cohort centroid. Statistical outliers were subsequently contextualised through cohort-level validation against canonical oncogenes, tumour suppressors, and housekeeping genes using fold-change analysis. Three filtering strategies were evaluated, and the final dataset was selected to maximise molecular subtype separation while preserving sufficient sample size for model training.

Gene filtering and feature selection

Low-information genes were removed using a dual threshold approach, excluding genes with variance < 0.1 and those within the lowest 10% of mean absolute expression. This retained 18,477 genes while preserving biologically relevant tumour suppressors and regulatory genes. Differential expression analysis was conducted using two-sample Student's t-tests to compare malignant versus normal tissue and between molecular subtypes. Genes with $p < 0.05$ and $|\log_2 \text{fold change}| > 0.585$ were considered significant, yielding 15,925 differentially expressed genes. To reduce redundancy, Pearson correlation filtering was applied with a threshold of $|r| > 0.85$. Highly correlated gene pairs were assessed using F-tests to confirm comparable variance. Genes were then ranked using an analysis-of-variance F-statistic to quantify discriminatory power across molecular subtypes. The top 5,000 genes were retained for model input, resulting in an effective sample-to-feature ratio suitable for DL.

Neural network architecture and training

Data were standardised and stratified into training (70%), validation (15%), and test (15%) sets. Input data were reshaped to a (samples, 1, features) format. A multi-task Bi-LSTM RNN was constructed, consisting of two Bi-LSTM layers (128 and 64 units), followed by dropout (0.30), a dense layer (32 units, rectified linear unit (ReLU) activation), and dual output layers for binary (sigmoid) and subtype (softmax) classification. L2 regularisation ($\lambda = 0.001$) was applied. The model was trained using the Adam optimiser (learning rate = 0.001) with early stopping (patience = 5 epochs). The final architecture comprised 905,495 trainable parameters.

Model evaluation and interpretability

Performance was evaluated using accuracy, sensitivity, specificity, precision, and F1 score, computed with sklearn.metrics. Confusion matrices were generated for both malignancy and subtype classification. Model interpretability was assessed using SHAP, with beeswarm plots used to identify gene-level contributions to predictions. SHAP-identified genes were cross-referenced with established and emerging breast cancer biomarkers to assess biological plausibility.

RESULTS

To ensure that downstream modelling was performed on biologically coherent and statistically reliable data, a systematic dataset curation and feature selection process was undertaken to improve sample-level quality while preserving discriminative molecular signal. Four sample filtering strategies were compared - unfiltered, conservative, moderate, and strict - each removing samples based on increasing confidence of biological implausibility or transcriptional dysregulation (Table 1).

Dataset quality was quantified using clustering separation score, compactness, and their ratio as an overall quality metric.

Progressive filtering resulted in monotonic im-

Table 1. Comparative analysis of dataset filtering strategies

Dataset	Samples	Separation Score	Compactness	Quality Ratio	Sample Retention
Original	956	78.80	331.21	0.24	100%
Conservative	955	78.89	332.85	0.24	99.9%
Moderate	952	80.30	289.27	0.28	99.6%
Strict	932	83.29	271.81	0.31	97.5%

provements in clustering quality, with the separation score increasing from 78.80 in the unfiltered dataset to 83.29 under strict filtering, and the quality ratio increasing from 0.24 to 0.31 (Table 1). However, increasing stringency was associated with reduced sample retention, decreasing from 100% in the unfiltered dataset to 97.5% under strict filtering. Conservative filtering yielded minimal improvement in clustering quality despite near-complete sample retention (99.9%), indicating that removal of only the most extreme outliers was insufficient to substantially improve dataset structure.

Visualisation using t-SNE demonstrated progressive refinement of molecular subtype clustering with increasing filtering stringency (Figure 1a).

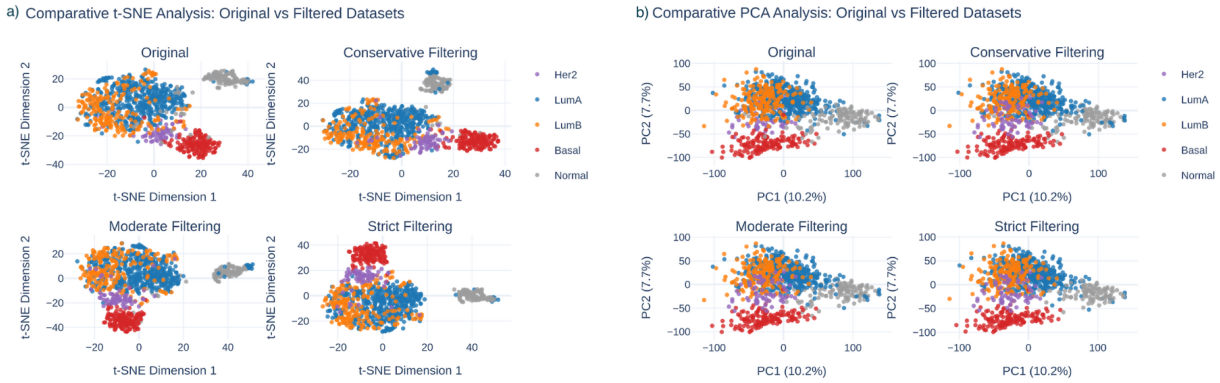


Figure 1. Effect of Progressive Filtering on TCGA-BRCA Sample Structure. (a) Comparative t-SNE plots of TCGA-BRCA sample clustering: original dataset vs. increasing filtering stringency (conservative, moderate, strict). More stringent filtering yields more defined clusters (colours denote subtype: grey = normal, blue = Luminal A, orange = Luminal B, purple = HER2-enriched, red = Basal-like). (b) Comparative PCA of Dataset Filtering Strategies. Each point represents a sample, projected onto the first two principal components (PC1 and PC2), which together explain 17.9% of total variance (PC1: 10.2%, PC2: 7.7%).

The unfiltered dataset exhibited diffuse and overlapping subtype clusters, whereas moderate filtering produced clearer separation between normal tissue, luminal, HER2-enriched, and basal-like samples while preserving expected areas of biological overlap. Although strict filtering produced the most compact clusters, boundaries appeared overly discrete, suggesting potential over-pruning of biologically valid intermediate samples. PCA was used as a complementary quality control measure and showed broadly consistent subtype organisation across filtering strategies, with the first two components explaining 17.9% of total variance (PC1: 10.2%, PC2: 7.7%; Figure 1b).

Based on the combined quantitative metrics and clustering visualisations, the moderate filtering strategy was selected for all downstream analyses as it achieved substantial improvement in dataset quality while retaining 99.6% of samples, thereby balancing data integrity with statistical power.

Following sample-level curation, feature selection reduced the dataset from 20,530 genes to 5,000 discriminative features (Table 2).

Table 2. Gene filtering pipeline results

Step	Method	Threshold / Criteria	Genes Retained
Initial Count	-	-	20 530
Low-Information Selective Filtering	Removing genes with mean absolute expression $\leq \pm$ 0.10	Var Threshold: 0.10	18 477
Differential Expression Analysis	T-test ($p < 0.05$)	$p < 0.05$	15 925
Quantile Normalisation	Standardised variance	$1.001 \rightarrow 0.088$	-
Signal-to-Noise (SNR) Filtering	SNR ratio	SNR > 1.0	492
Correlation filtering	Pearson correlation	$r > 0.85$	15 285
F-statistic ranking	Top discriminative features	Top 5 000	5 000

Differential expression analysis identified 15,925 significantly dysregulated genes between classes ($p < 0.05$, $\alpha = 0.05$), providing a statistically supported basis for dimensionality reduction. To assess biological validity of the final feature set, retention of established breast cancer biomarkers was evaluated (Table 3).

Table 3. Biomarker retention in final feature set

Biomarker Category	Total	Retained	Retention (%)	Key Examples Retained
PAM 50 Genes	10	3	30.0	ESR1, PGR, ERBB2
Tumour Suppressors	7	4	57.1	BRCA1, BRCA2, PTEN, CDKN2A
Oncogenes	6	3	50.0	MYC, CCND1, EGFR
Luminal Markers	5	5	100.0	ESR1, PGR, GATA3, FOXA1, TFF1
Basal Markers	6	2	33.3	EGFR, FOXC1
HER2-Enriched Pathway	5	3	60.0	ERBB2, PGAP3, LASP1
Overall	39	20	51.3	-

Binary classification of malignant versus normal breast tissue achieved an overall accuracy of 96.5% (Figure 2a).

All malignant samples were correctly classified (sensitivity = 100%), while 12 of 17 normal samples were correctly identified (specificity = 70.6%), resulting in five false positive predictions and no false negatives. Performance metrics reflected balanced classification despite class imbalance, with a weighted F1-score of 0.96 (Table 4).

Table 4. Malignancy classification performance metrics

	Precision	Recall	F1-Score
Healthy	1.00	0.71	0.83
Malignant	0.96	1.00	0.98
Accuracy	-	-	0.97
Macro Average	0.98	0.85	0.90
Weighted Average	0.97	0.97	0.96

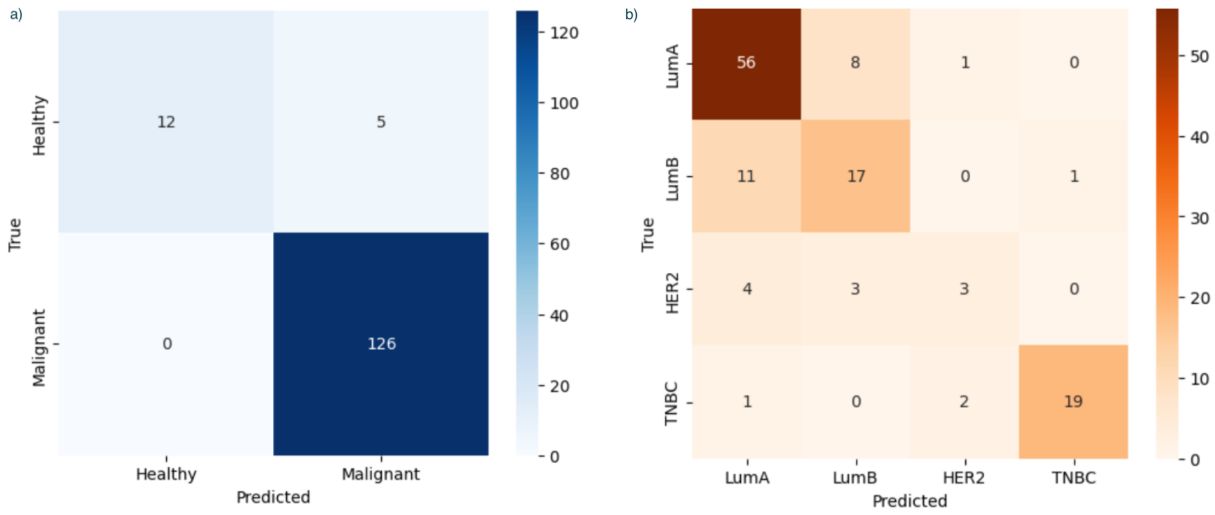


Figure 2. Confusion Matrices For Breast Cancer Classification. (a) Confusion Matrix for Malignancy Classification. The matrix demonstrates high sensitivity (1.00) for malignant tissue detection with minimal false positive predictions. (b) Confusion Matrix for Molecular Subtype Classification. Note: TNBC (Triple-Negative Breast Cancer) labels represent the Basal-like subtype.

Subtype classification achieved an overall accuracy of 75.4%, with performance varying across molecular subtypes (Figure 2b). Luminal A and basal-like tumours demonstrated the highest classification performance, with recall values of 86.2% and 86.4% respectively (Table 5).

Table 5. Molecular subtype classification performance metrics

	Precision	Recall	F1-Score
Luminal A	0.78	0.86	0.82
Luminal B	0.61	0.59	0.60
HER-2	0.50	0.30	0.38
Basal-Like	0.95	0.86	0.90

Luminal B exhibited intermediate performance (recall = 58.6%), while the HER2-enriched subtype showed the lowest recall (30.0%). Misclassification patterns were dominated by confusion between luminal A and luminal B subtypes, which together accounted for 61.3% of all classification errors (Figure 2b). Conversely, basal-like tumours showed comparatively low overlap with other subtypes.

Substantial gene overlap was observed across subtypes, particularly among luminal and HER2-enriched tumours, whereas basal-like predictions were driven by more distinct features (Figure 3).

Across all subtypes, TF emerged as a consistently influential feature, while ARFIP2 and Bardet-Biedl syndrome 12 (BBS12) appeared in multiple subtype signatures. In contrast, several genes demonstrated subtype exclusivity, including cadherin 19 (CDH19), hydroxysteroid 17-beta dehydrogenase 8 (HSD17B8), and kelch like family member 13 (KLHL13) for luminal A; serine peptidase inhibitor Kazal type 4 (SPINK4) and Solute Carrier Family 9 Member A3 regulator 1 (SLC9A3R1) for luminal B; neurofilament light chain (NEFL), ribosomal protein L26 like 1 (RPL26L1), and MSN for HER2-enriched tumours; and maternally expressed 3 (MEG3), delta/notch like epidermal growth factor repeat containing (DNER), and purinergic receptor P2X 6 (P2RX6) for basal-like tumours (Table 6).

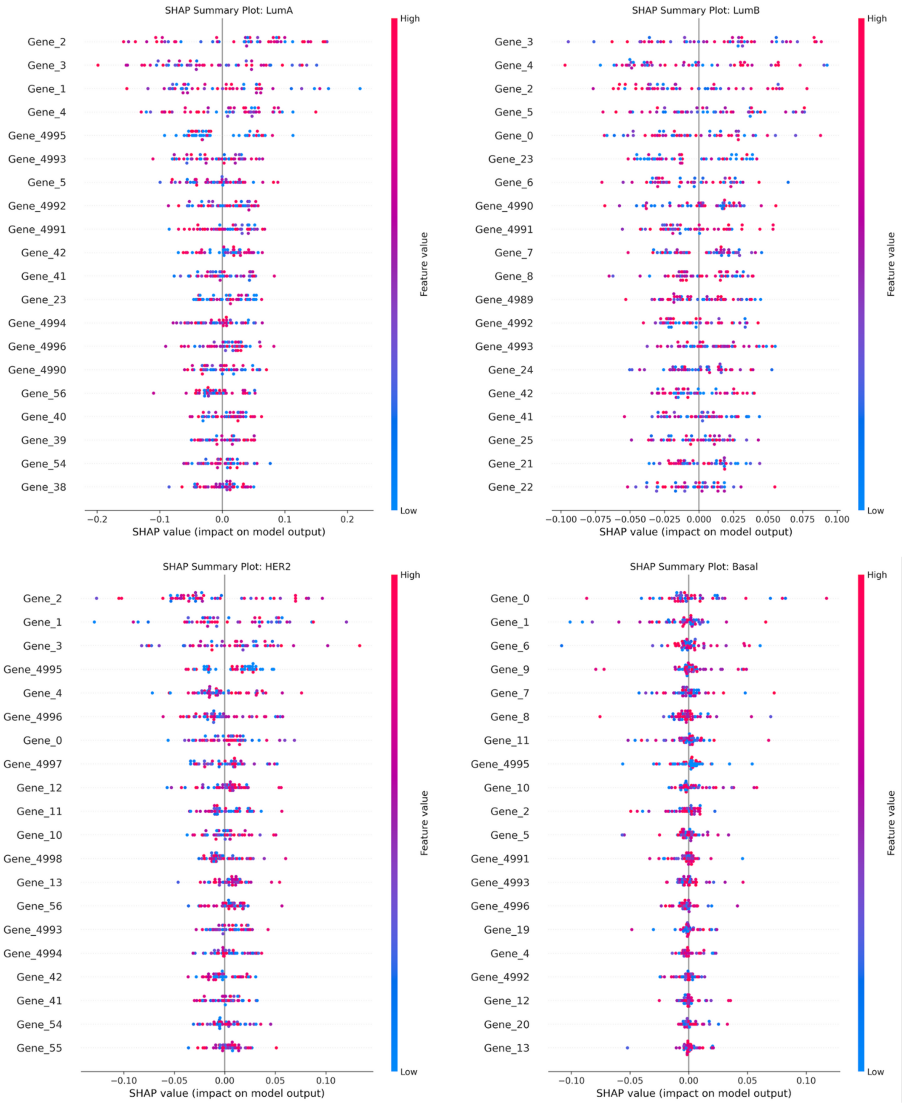


Figure 3. SHAP summary plots showing shared and subtype-specific gene contributions across breast cancer molecular subtypes. Luminal A (top left), Luminal B (top right), HER2-enriched (bottom left), and Basal-like (bottom right).

Table 6. Top 10 SHAP-identified genes by molecular subtype

Rank	Corresponding Genes			
	Luminal A	**Luminal B**	**HER2-Enriched**	**Basal-Like**
1	TF	ARFIP2	TF	ACAT2
2	ARFIP2	HYMAI	BBS12	BBS12
3	BBS12	TF	ARFIP2	SSPN
4	HYMAI	BTRC	LHCGR	MEG3
5	LHCGR	ACAT2	HYMAI	PSENEN
6	HSD17B8	SPINK4	NEFL	DNER
7	BTRC	SSPN	ACAT2	HNRPAB
8	KLHL13	SLC9A3R1	RPL26L1	LHCGR
9	PER2	PER2	MSN	P2RX6
10	CDH19	PSENEN	HNRNPAB	TF

DISCUSSION

This investigation evaluated whether a RNN trained on RNA-seq gene expression data could accurately distinguish malignant from healthy breast tissue, classify molecular subtypes, and identify key genes driving these predictions. Overall, the model achieved high diagnostic performance, demonstrating 96.5% accuracy and 100% sensitivity for malignancy detection (Figure 2a), alongside moderate specificity (70.6%). Subtype classification accuracy reached 75.4% (Figure 2b), with performance varying across molecular subtypes. Explainability analysis revealed biologically plausible gene signatures that aligned with both established and emerging breast cancer biomarkers (Figure 3). Collectively, these findings support the hypothesis that transcriptomic differences between healthy and malignant tissue can be leveraged by RNNs to enable accurate and interpretable breast cancer classification.

The absence of false negatives in malignancy classification is clinically significant, as it indicates that no malignant samples were incorrectly classified as healthy (Figure 2a). While specificity was lower, resulting in some false positives, this trade-off is acceptable in a screening or triage context where sensitivity is prioritised. Importantly, the model operates on molecular data and therefore has the potential to detect malignancy prior to observable anatomical changes, distinguishing it from imaging-based approaches such as CNN-driven breast imaging systems (Yu et al., 2023). These results suggest that RNA-seq-based models could complement existing diagnostic workflows as an early molecular screening tool rather than a definitive standalone test.

Subtype classification performance reflected known biological relationships between breast cancer subtypes. Luminal A and basal-like tumours were classified with the highest accuracy, while luminal B and HER2-enriched subtypes exhibited greater misclassification. Notably, 61.3% of subtype classification errors occurred between luminal A and luminal B tumours (Figure 2b). This pattern is consistent with the biological continuum between these oestrogen receptor-positive subtypes, which share hormone-regulated pathways and differ primarily in proliferation-associated gene expression (Orrantia-Borunda et al., 2022). Consequently, the observed confusion is likely attributable to genuine molecular overlap rather than model inadequacy.

HER2-enriched tumours showed the lowest classification performance (Figure 2b and Table 5), which may reflect both the small sample size for this subtype and its inherent transcriptional heterogeneity. The dispersion of misclassified HER2-enriched samples across multiple predicted classes suggests variable gene expression patterns that were not fully captured by the selected feature set. These findings align with previous research highlighting the molecular diversity of HER2-positive disease and the limitations of single-modality transcriptomic classification (Orrantia-Borunda et al., 2022).

SHAP analysis provided gene-level insight into model decision-making, revealing both shared and subtype-specific molecular drivers. Several genes, including TF, ARFIP2, and BBS12, appeared across multiple subtype signatures, suggesting involvement in common regulatory pathways underlying breast cancer progression (Table 6). The identification of TF as a consistently influential feature aligns with emerging evidence linking iron metabolism dysregulation to tumour growth and therapeutic resistance (Alhaddad et al., 2025). Similarly, ARFIP2 and MSN have been independently associated with trastuzumab resistance and poor prognosis, supporting the biological plausibility of the model's explanations (Shi et al., 2024). On the contrary, a subset of genes demonstrated subtype exclusivity, including CDH19 and HSD17B8 for luminal A, SPINK4 for luminal B, NEFL and MSN for HER2-enriched tumours, and MEG3 and P2RX6 for basal-like disease (Table 6). These findings extend established Prediction Analysis of Microarray 50 (PAM50) panels by highlighting additional genes that may contribute to refined subtype discrimination.

The preprocessing pipeline played a critical role in balancing data quality with biological preservation. Moderate filtering improved clustering structure while retaining 99.6% of samples, avoiding over-pruning of biologically valid intermediate cases (Table 1). This hybrid approach addressed limitations of purely computational feature selection noted in previous studies (Acedo-Terrades et al., 2025.

Several limitations may have influenced the findings. Class imbalance between malignant and healthy samples likely contributed to reduced specificity. Subtype imbalance, particularly for HER2-enriched tumours, constrained classification performance and limits generalisability. Ad-

ditionally, reliance on transcriptomic data alone excluded clinical variables such as tumour stage, treatment history, and imaging features that are routinely integrated into diagnostic decision-making. The exclusion of a non-breast cancer control dataset further limits assessment of cross-cancer specificity. Future work should prioritise larger, more balanced multi-cohort datasets and integrate multimodal inputs, including clinical and imaging data, to enhance robustness. Longitudinal gene expression analysis could also enable modelling of disease progression and treatment response.

CONCLUSION

This study demonstrated that recurrent neural networks can accurately and interpretably classify breast cancer using RNA-seq gene expression data. The proposed multi-task bidirectional long short-term memory achieved high diagnostic performance for malignancy detection with perfect sensitivity, alongside reliable molecular subtype classification, with performance patterns reflecting known biological relationships between subtypes. Importantly, the preprocessing and feature selection strategy improved dataset quality while preserving clinically relevant biomarkers, ensuring alignment with established molecular pathways. Explainability analysis further validated the model by identifying biologically meaningful genes consistent with current breast cancer literature. Together, these findings support the research hypothesis that explainable deep learning approaches can leverage transcriptomic data to deliver clinically relevant breast cancer classification while maintaining transparency essential for medical adoption.

ACKNOWLEDGEMENTS

I would like to thank my Science Extension teacher and project supervisor, Ms Thompson, for running the Science Extension course solely for me. Without her initiative, this research opportunity would not have been possible. Her generosity with time and guidance, from identifying gaps in the literature to the final edits, was invaluable. I thank Safaa Al Haj Hussein and Teah Goodhand for their mentorship throughout this project. Their monthly Zoom meetings provided essential feedback on the scientific process and helped ensure a rigorous data preprocessing pipeline. I am grateful for their careful review of my drafts and ongoing support. I also thank Ms Cartwright and Ms English for reviewing my first complete draft and providing constructive feedback that significantly strengthened the quality of my report.

REFERENCES

Acedo-Terrades, A., Perera-Bel, J., & Nonell, L. (2025). The importance of data transformation in RNA-seq preprocessing for bladder cancer subtyping. *BMC Research Notes*, *18*(1). https://doi.org/10.1186/s13104-025-07138-x

Alhaddad, L., Saifo, M., & Alsayed, R. (2025). Special status of iron transfer proteins in the serum of breast cancer patients. *Discover Oncology*, *16*(1). https://doi.org/10.1007/s12672-025-02799-3

Alharbi, F., & Vakanski, A. (2023). Machine learning methods for cancer classification using gene expression data: A review. *Bioengineering*, *10*(2), 173. https://doi.org/10.3390/bioengineering10020173

Australian Institute of Health and Welfare. (2025). *Cancer data in Australia*. https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/overview

Burcu, B., Ertas, I., Sener, A., Demircioglu, Z. G., Cerekci, E., & Kaya, C. (2023). Retrospective analysis of parameters affecting metastatic breast cancer. *The Medical Bulletin of Sisli Etfal Hospital*, *57*(4), 479–484. https://doi.org/10.14744/SEMB.2023.94803

Dasgupta, S., Dutta, M., Halder, A., Khan, A., Saha, G., & Pal, R. K. (2025). Computational analysis of gene expression data using bidirectional long short-term memory for disease diagnosis. *Innovations in Systems and Software Engineering*, *21*, 93–107. https://doi.org/10.1007/s11334-022-00492-0

Islam, T., Sheakh, Md. A., Tahosin, Mst. S., Hena, Most. H., Akash, S., Bin Jardan, Y. A., FentahunWondmie, G., Nafidi, H.-A., & Bourhia, M. (2024). Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-57740-5

Kim, J.-Y., Lee, Y. S., Yu, J., Park, Y., Lee, S. K., Lee, M., Lee, J. E., Kim, S. W., Nam, S. J., Park, Y. H., Ahn, J. S., Kang, M., & Im, Y.-H. (2021). Deep learning-based prediction model for breast cancer recurrence using adjuvant breast cancer cohort in tertiary cancer center registry. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/fonc.2021.596364

Levy, S. E., & Boone, B. E. (2019). Next-generation sequencing strategies. *Cold Spring Harbor Perspectives in Medicine*, *9*(7), a025791. https://doi.org/10.1101/cshperspect.a025791

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. *30*, 4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Mantione, K. J., Kream, R. M., Kuzelova, H., Ptacek, R., Raboch, J., Samuel, J. M., & Stefano, G. B. (2014). Comparing bioinformatic gene expression profiling methods: Microarray and RNA-seq. *Medical Science Monitor Basic Research*, *20*, 138–142. https://doi.org/10.12659/msmbr.892101

Orrantia-Borunda, E., Anchondo-Nuñez, P., Acuña-Aguilar, L. E., Gómez-Valles, F. O., & Ramírez-Valdespino, C. A. (2022). Subtypes of breast cancer. In *Breast cancer* (pp. 31–42). Exon Publications. https://doi.org/10.36255/exon-publications-breast-cancer-subtypes

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681. https://doi.org/10.1109/78.650093

Shi, X., Sheng, Y., Fei, H., Wei, B., Zhang, Z., Xia, X., Mao, C., & Si, X. (2024). Combined transcriptome and proteome analysis reveals MSN and ARFIP2 as biomarkers for trastuzumab resistance of breast cancer. *Breast Cancer Research and Treatment*, *207*(1), 187–201. https://doi.org/10.1007/s10549-024-07355-1

The Cancer Genome Atlas Program. (2018). *The Cancer Genome Atlas Program*. National Cancer Institute. https://www.cancer.gov/ccg/research/genome-sequencing/tcga

Yu, X., Dong, M., Yang, D., Wang, L., Wang, H., & Ma, L. (2023). Deep learning for differentiating benign from malignant tumors on breast-specific gamma image. *Technology and Health Care*, *31*, 61–67. https://doi.org/10.3233/thc-236007

Zhang, Y., Ji, Y., Liu, S., Li, J., Wu, J., Jin, Q., Liu, X., Duan, H., Feng, Z., Liu, Y., Zhang, Y., Lyu, Z., Song, F., Song, F., Yang, L., Liu, H., & Huang, Y. (2025). Global burden of female breast cancer: New estimates in 2022, temporal trend and future projections up to 2050 based on the latest release from GLOBOCAN. *Journal of the National Cancer Center*, *5*(3), 287–296. https://doi.org/10.1016/j.jncc.2025.02.002