Original Article

JHSMR Journal of Health Science and Medical Research

Prediction of Regional Lymph Node Metastasis from the Clinicopathological Features of Breast Carcinoma: Application of Deep Learning

Santosh Kumar Sharma, M.Sc., Ph.D. scholar[1,2], Ratna Chopra, M.D.[3], Sanjay Kumar, Ph.D.[1], Sompal Singh, M.D.[2]

[1]Department of Life Science, School of Basic Sciences and Research, Sharda University, Greater Noida, Uttar Pradesh 201310, India.
[2]Department of Pathology, North Delhi Municipal Corporation Medical College and Hindu Rao Hospital, Delhi 110007, India.
[3]Department of Surgery, North Delhi Municipal Corporation Medical College and Hindu Rao Hospital, Delhi 110007, India.

Received 28 November 2024 ● Revised 12 February 2025 ● Accepted 17 February 2025 ● Published online 11 June 2025

Abstract:

Objective: Breast cancer is the leading cause of cancer-related mortality in women globally, with invasive ductal carcinoma (IDC) as the predominant subtype. Regional lymph node (LN) metastasis significantly impacts prognosis, staging, and treatment strategies. However, the role of deep learning in predicting LN metastasis is underexplored. To develop and evaluate a deep learning model leveraging clinicopathological features for predicting LN metastasis in IDC patients, with an aim to enhance diagnostic accuracy and reduce reliance on invasive methods.

Material and Methods: A cross-sectional study was conducted on 351 IDC cases from a tertiary-care hospital. Input variables included clinicopathological features: age, tumor size, modified Bloom-Richardson grade, ER, PR, HER2 receptor status, Ki-67 index, and microvessel density (MVD). LN status was dichotomized using a cut-off ratio of 0.3. A neural network model with an input layer of 8 neurons, 3 hidden layers (50 neurons each), and ReLU activation was developed. Data were split into training (70%) and test (30%) sets. Predictive accuracy was evaluated using standard performance metrics.

Results: The mean age was 46.4±11.29 years and tumor volume averaged 44.9 cm^3. Low ER (35.6%) and PR (26.8%) positivity rates were observed, with HER2 positivity at 21.7%. The model achieved 78.3% accuracy in predicting LN metastasis. The F1 score of the model was 0.83.

Contact: Sompal Singh, M.D.
Department of Pathology, North Delhi Municipal Corporation Medical College and Hindu Rao Hospital, Malka Ganj, Delhi 110007, India.
E-mail: sompal151074@gmail.com

J Health Sci Med Res
doi: 10.31584/jhsmr.20251224
www.jhsmr.org

Conclusion: The study demonstrates the utility of deep learning models in predicting LN metastasis using clinicopathological data. With 78.3% accuracy, the model highlights AI's potential in oncology diagnostics, supporting personalized treatment approaches. Further integration of imaging and molecular data could enhance model performance and clinical applicability.

Keywords: artificial intelligence, breast cancer, deep learning, regional lymph node metastasis

Introduction

Breast carcinoma is the most frequently diagnosed malignancy and a primary cause of cancer-related mortality among women worldwide. It accounts for about 2.3 million new cases annually, highlighting its immense global burden[1]. Among various histological subtypes, infiltrating ductal carcinoma (IDC) is the most common, comprising 70–80% of invasive breast cancers. The disease often presents with regional lymph node (LN) metastasis, which is a key determinant of prognosis, staging, and therapeutic strategies[2].

IDC outcomes are influenced by a range of prognostic factors, including tumor size, histological grade, receptor status [estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2)], microvessel density, and proliferation indices such as Ki-67. Lymph node involvement, particularly the extent of nodal metastasis, is one of the most significant predictors of survival[3]. Furthermore, molecular subtyping has refined our understanding of tumor biology, providing insights into the varying behaviours of luminal-A, luminal-B, HER2-enriched, and triple-negative breast cancer (TNBC) subtypes[4].

Lymphatic drainage plays a pivotal role in breast cancer metastasis, facilitating the spread of malignant cells to distant organs. The axillary lymph nodes, with a drainage probability of 98.2%, represent the most common and clinically significant nodal site for staging and prognosis[5]. The positive lymph node ratio (LNR), defined as the ratio of positive lymph nodes to the total lymph nodes examined, has emerged as a significant prognostic tool. Unlike traditional nodal staging, which relies solely on the number of positive nodes, LNR provides a more nuanced risk stratification, especially in cases with limited lymph node dissection. A higher LNR is associated with worse overall survival and disease-free survival across various breast cancer subtypes[6]. It has proven particularly valuable in tailoring adjuvant therapies, such as chemotherapy and radiotherapy, to mitigate recurrence risks[7].

Conventional methods for LN assessment, including sentinel LN biopsy and axillary LN dissection, are invasive, costly, and associated with potential complications such as lymphedema[8]. Advances in imaging and biomarker analyses have improved non-invasive approaches, yet their diagnostic accuracy remains limited by inter-observer variability and technical constraints. This underscores the need for robust predictive tools capable of integrating diverse clinicopathological parameters to enhance diagnostic precision.

Deep learning, a subset of artificial intelligence, offers transformative potential in oncology. It excels in pattern recognition and predictive modeling, particularly with large and complex datasets[9]. In breast cancer, deep learning algorithms have shown promise in tasks like tumor classification, molecular subtyping, and recurrence prediction. However, its application in predicting regional LN metastasis remains underexplored. Developing models that integrate clinicopathological parameters with deep learning approaches could significantly enhance diagnostic accuracy and treatment planning, particularly in resource-constrained settings[10].

This study focuses on applying deep learning to predict regional LN metastasis in IDC cases based on clinicopathological features. By leveraging a neural network model, the research aims to evaluate its predictive performance and explore its potential as an adjunct to traditional diagnostic methods, thus improving patient outcomes and reducing the reliance on invasive procedures.

Material and Methods

The present study was conducted in the departments of pathology and general surgery in a tertiary-care medical college and hospital. A total of 351 consecutive cases of infiltrating ductal carcinoma of the breast from June 2022 to April 2024 were included in the study. Patients who did not provide their consent were excluded from the study. The clinicopathological features taken into consideration as input variables for the neural network were age, volume of tumor, modified Bloom Richardson (BR) score of tumor, ER positivity, PR positivity, HER2 positivity, Ki-67 score and microvessel density (MVD). Ratio of the number of positive LNs out of total LNs isolated was taken as the outcome variable. The outcome variable was dichotomized with a cut-off of 0.3. The dichotomization of the outcome variable (LN metastasis) at a cut-off of 0.3 refers to the threshold used to classify cases as either having significant lymph node involvement (≥0.3) or not (<0.3). This cut-off value is based on the lymph node ratio (LNR), calculated as the ratio of the number of positive lymph nodes to the total number of lymph nodes examined. A value of 0.3 was chosen based on clinical relevance. Additionally, this threshold was statistically tested in the study using the Receiver Operating Characteristic (ROC) curve analysis to evaluate its predictive significance.

The data analysis was done using Python language version 3.0 with the help of Jupyter notebook version 5.0. Python package scikit-learn was used to build a neural network. The input layer of the neural network comprised 8 neurons. There were 3 hidden layers with 50 neurons each. The activation function used was Rectified Linear Unit (ReLU) with an adaptive learning rate. The data were divided into training set (70%) and test set (30%). The predictive capability of the model developed was assessed by plotting the ROC curve. Evaluation metrics included accuracy, sensitivity, specificity, and area under the curve (AUC) of the ROC analysis in order to predict LN metastasis.

Modal accuracy was calculated using the test set where accuracy was defined as:

TP+TN/TP+TN+FP+FN), which takes into account true positive (TP), true negative (TN), false positive (FP), false negative (FN).

Results

The study included a total of 351 cases of IDC of the breast, all of which were histopathologically classified as the NOS (Not Otherwise Specified) subtype. The clinicopathological characteristics of these cases are summarized in Table 1.

The mean age of the patients was 46.4 years (S.D.±11.29), with a median age of 45 years, ranging from 20 to 78 years.

The mean tumor volume was 44.9 cm^3 (S.D.±51.7 cm^3), and the median volume was 31.5 cm^3, with a range of 0.6 to 440 cm^3.

The mean BR score was 7.2, with a median score of 7, ranging from 3 to 9.

Out of the 351 cases, 125 (35.6%) were ER positive, while 226 (64.4%) were ER negative. PR positivity was observed in 94 cases (26.8%), while 257 cases (73.2%) were PR negative.

HER2 status was determined based on staining intensity and divided into 4 categories as follows: absence of staining (category 0), weak intensity (category 1), equivocal (category 2), and strong positive (category 3). Among all the cases, 212 (60.4%) fell into category 0, 32

Table 1 Clinicopathological features of IDC-NOS patients

Clinicopathological features	Total number of cases	Mean (±S.D.)/ Median (range)
Age (years)	351	46.4 (±11.29)
Tumor volume (cm^3)	351	44.9 (±51.7)
MVD (microvessels/hpf)	351	17.1 (±9.9)
Ki-67 (%)	351	31.3 (±21.1)
BR score	351	7 (range 3-9)
	Number of cases	**Percentage (%)**
ER Positive	125	35.6
ER Negative	226	64.4
PR Positive	94	26.8
PR Negative	257	73.2
HER2 status		
Category 0	212	60.4
Category 1	32	9.1
Category 2	31	8.8
Category 3	76	21.7
LN ratio (≥0.3)	111	31.6
LN ratio (<0.3)	240	68.4

IDC-NOS=invasive ductal carcinoma-not otherwise specified, MVD=microvessel density, Ki-67=Kiel 67, BR=Bloom Richardson, ER=estrogen receptor, PR=progesterone receptor, HER2=human epidermal growth factor receptor-2, LN=lymph node

(9.1%) into category 1, 31 (8.8%) into category 2, and 76 (21.7%) into category 3. For analysis, categories 0, 1, and 2 were grouped as HER2-negative, while category 3 was classified as HER2 positive. Based on this grouping, 76 cases (21.7%) were HER2 positive, and 275 cases (78.3%) were HER2 negative.

The Ki-67 proliferation index, a marker of tumor cell proliferation, had a mean score of 31.3% (S.D.±21.1%) and a median score of 30%, ranging from 2% to 85%. Cases were further stratified into groups based on Ki-67 scores: 138 cases (39.3%) had a score below 20%, 119 cases (33.9%) had scores between 21-40%, 54 cases (15.4%) fell into the 41-60% group, and 40 cases (11.4%) had scores exceeding 60%.

The mean MVD index, a measure of tumor angiogenesis, was 17.1 (S.D.±9.9) microvessels per high-power field (hpf), with a median value of 15.0 microvessels per hpf, ranging from 3.0 to 100 microvessels per hpf.

In the present study, we found 111 cases of breast cancer with an LN ratio of ≥ 0.3, while 240 cases had an LN ratio of <0.3, comprising 31.6% and 68.4%, respectively.

The neural network model developed in this study for predicting regional LN metastasis had an input layer comprising 8 neurons, corresponding to the clinicopathological variables. The model included 3 hidden layers with 50 neurons each, utilizing the ReLU as the activation function. An adaptive learning rate optimizer was employed to enhance convergence. The dataset was divided into a training set (70%) and a test set (30%) in order to validate the model's performance. The accuracy of the neural network model developed in this study for predicting regional LN metastasis was 78.3% with AUC: 0.72; sensitivity: 40%; specificity: 70%; F1-score: 0.83. The results are depicted in Figure 1 and 2.

After a small unpublished pilot study by the authors assessing the utility of LNR to predict the survival of breast cancer patients, a cut-off value of 0.3 was obtained.

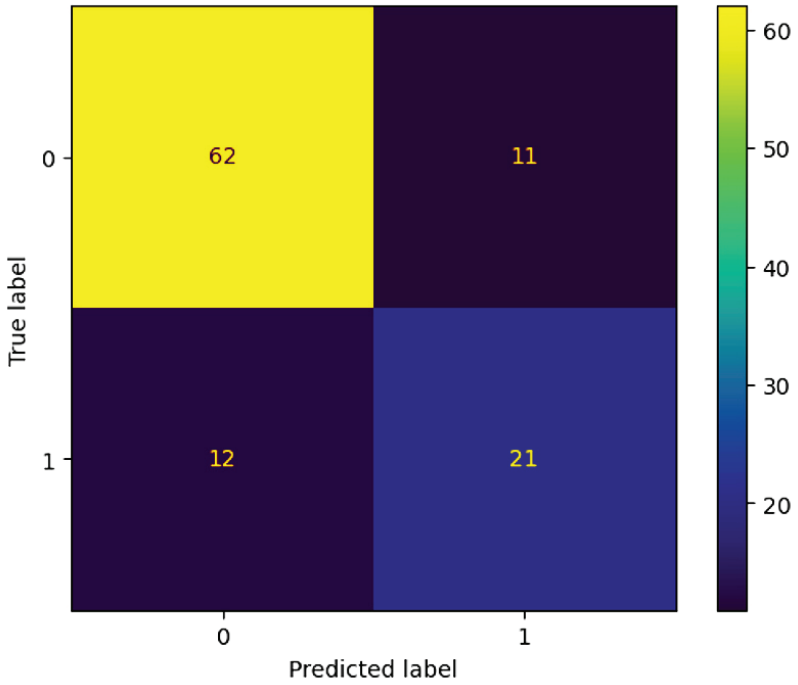

Figure 1 Prediction of the model as compared to the true label (0 means lymph node ratio (LNR) <0.3; 1 means ≥0.3)

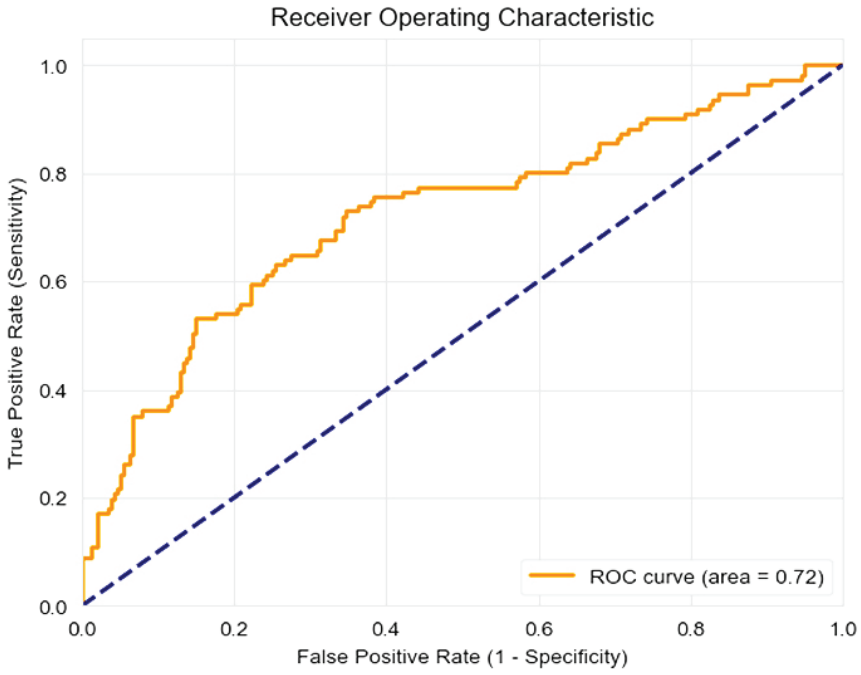

Figure 2 Receiver operating characteristic (ROC) curve of the results of the model

Discussion

The present study provides significant insights into the clinicopathological characteristics of IDC and highlights the potential of a deep learning model in predicting regional LN metastasis. The findings underscore the applicability of artificial intelligence (AI) in enhancing diagnostic accuracy, which could potentially reduce the need for invasive staging procedures.

Patient Characteristics

The mean age of the study population was 46.4±11.29 years, which is slightly younger compared to several previous studies, such as those by Shahriarirad et al.[11] and Zhou et al.[12], which reported mean ages of 50.9 and 48.6 years, respectively. This difference could reflect regional demographic variations in breast cancer incidence, emphasizing the need for context-specific modeling.

The mean tumor volume of 44.9 cm^3 and the predominance of grade II and III tumors align with findings from studies like Polat et al.[13] and Chen et al.[14]. The inclusion of detailed tumor volume measurements, beyond the traditional T-stage classification, adds depth to the predictive modeling and highlights the importance of nuanced pathological data in AI-based predictions.

ER and PR positivity were observed in 35.6% and 26.8% of cases, respectively, contrasting with higher positivity rates in studies by Shahriarirad et al.[11] (ER+ 74.3%; PR+ 71%) and Dihge et al.[15] (ER+ 86%; PR+ 74%). Similarly, HER2 positivity (21.7%) was comparable to the findings of Polat et al.[13] and Shiner et al.[16], reinforcing its moderate prevalence across cohorts. These differences in receptor status underline the importance of incorporating molecular subtypes into predictive models for better generalizability across diverse populations.

The Ki-67 index, a proliferation marker, showed a mean score of 31.3%, with 39.3% of cases falling below the 20% threshold. This finding aligns with prior research, such as Chen et al.[14], which also emphasized the role of high Ki-67 expression (>30%) in predicting LN metastasis. The stratification of Ki-67 into finer categories further refines its prognostic utility in AI algorithms.

Application of machine learning and artificial intelligence

The neural network model developed in the present study achieved an accuracy of 78.3% with an AUC of 0.72 for predicting regional lymph node metastasis in breast carcinoma. Various studies in the literature assessed the role of machine learning and artificial intelligence in the prediction of one or more prognostic markers of breast carcinoma. Whereas, some of the studies have taken clinicopathological features as input variables in the machine learning model[11,14-18], other studies have taken either radiological images or histopathological slide images as the input variable[12,13,19,20]. Most of the studies tried to predict either regional lymph node metastasis or distant metastasis[11,12,14-17,19,20]. Occasional studies have directly predicted the overall survival and recurrence-free survival of breast cancer patients[15,18, 21-24] (Table 2).

The results of our study align with findings from Chen et al.[14], which demonstrated the significant role of clinicopathological features, including pathological type, HER-2 status, Ki-67 expression, hormone receptor status, and tumor size, in predicting axillary LN metastasis. Their multivariate logistic regression model identified these 5 predictors as central to predicting axillary LN metastasis, showing robust performance with an AUC of 0.725 in the training cohort. This consistency across studies highlights the utility of similar clinicopathological variables in estimating lymph node involvement in breast cancer patients.

Polat et al.[13] developed a neural network model based solely on clinicopathological features, achieving a moderate accuracy of 78.3% for predicting regional lymph node metastasis. Their model showed lower performance

Table 2 Summary of publications illustrating the utility of the machine learning algorithm for breast cancer

Authors name	Sample size	Machine learning model used	Input data set	Output variables	Highest accuracy
Chen et al.[14]	2,278 (train), 745 (validate)	Logistic Regression	Clinicopathological features	Axillary LN metastasis status	Training AUC=0.725; Validation AUC=0.786
Dihge et al.[15]	3,023	Gradient Boosting Machine (GBM)	Mixed (clinicopathological+ gene expression)	Nodal metastasis	AUC=0.72
Hu et al.[18]	421	CNN Models (e.g., ResNet-18)	WSIs, clinical data	Histological grade, Ki-67 expression, molecular features	AUC=0.68–0.90 for histological grade
Park et al.[17]	1,127 images	DenseNet121, Ensemble Models	Preprocessed CT images	Axillary LN metastasis	AUROC=0.968; Accuracy=93.8%
Polat et al.[13]	350	4D Hybrid CNN	Dynamic contrast-enhanced (DCE) MRI	cN and pN status	pN AUC=0.87
Shahriarirad et al.[11]	1,832	TabNet	Retrospective clinicopathological dataset	Sentinel LN involvement	Accuracy=75%
Shiner et al.[16]	175	Gradient Boosting Machine (GBM)	Clinical data from a retrospective cohort	Distant metastasis sites	AUC=0.75 (brain)
Xu Feng et al.[19]	1,058	DL-CNB+C (AMIL Framework, VGG16_ BN)	WSIs, clinicopathological data	Axillary LN metastasis status	AUC=0.831; Subgroup AUC=0.918 (age ≤50 years)
Ding et al.[20]	3,701	Multi-Modal Model Integration (MMMI)	Clinicopathological features +WSIs	LN metastasis status	AUC=0.809
Zhou et al.[12]	Not specified	Inception V3	Breast cancer ultrasound images	LN metastasis	AUC=0.90 (Test Set A)

LN=lymph node, AUC= area under curve, CT=computed tomography, CNN=convolutional neural network

in predicting clinical node (cN) status with an AUC of 0.55, consistent with findings from other studies like Shiner et al.[16], which found logistic regression models achieving AUCs of 0.74 and 0.70. The improvement in predictive performance with the addition of temporal features in the 4D model highlights the benefit of incorporating the dynamic aspects of tumor progression. Polat et al.[13] emphasized the value of integrating imaging data (3D and 4D models) for improved prediction outcomes, as shown by their 4D hybrid model yielding higher AUCs (0.87 for pathologic node (pN) status) compared to simpler models.

Xu et al.[19] focused on model selection for breast cancer metastasis prediction, with the VGG16_BN model outperforming others like ResNet50 and DenseNet121 across multiple metrics. This supports the idea that advanced convolutional neural networks (CNNs) can significantly boost predictive accuracy. Their findings are consistent with the model proposed by Polat et al.[13], which combined clinicopathological features with imaging data, achieving better performance in metastasis prediction. This synergy of clinical and imaging data aligns with the approach suggested by Zhou et al.[12], where multimodal deep learning models outperformed single-modality models.

Ding et al.[20] introduced the concept of multi-modal learning by integrating clinicopathological parameters with whole-slide images (WSIs) for lymph node metastasis prediction. Their approach, which included a modal fusion module, showed a significant enhancement in predictive accuracy for different metastatic statuses (micrometastasis, macrometastasis, isolated tumor cells, no metastasis). This approach aligns with the findings from Zhou et al.[12], where combining multiple imaging modalities improved performance. The results from the study by Ding et al.[20] suggest that integrating both clinical data and digital

pathology can provide a more comprehensive prediction model, similar to the multimodal strategies used by Polat et al.[13].

Zhou et al.[12] evaluated various deep learning models on breast cancer images for predicting lymph node metastasis, achieving high AUCs across multiple models. Their results validate the findings of Polat et al.[13] by emphasizing the importance of model complexity and feature selection. The use of models like Inception V3 and ResNet101 underscores the impact of choosing the right architecture for accurate predictions. This aligns with the enhanced performance observed in Polat et al.'s[13] 4D hybrid model, which used temporal data and imaging features to boost prediction accuracy.

Dihge et al.[15] evaluated various machine-learning methods for predicting nodal metastasis, emphasizing gradient boosting machines (GBM) as the most effective. Their findings showed AUC values ranging from 0.65 to 0.73 across different molecular subtypes of breast cancer, indicating varied predictive performance. The reduced accuracy for HER2-enriched and TNBC subgroups suggests the models' sensitivity to the inherent heterogeneity within these groups, raising concerns about overfitting. This sensitivity is reflected in the performance discrepancies observed for ER+ with HER2-, HER2-enriched, and TNBC subgroups.

Hu et al.[18] focused on hormone receptor positive (HR+)/HER2- breast cancer, utilizing deep learning models to analyse multi-omics data. Their study demonstrated high predictive accuracy for histological grade and Ki-67, achieving AUCs up to 0.90 in the validation set. This approach showed superior performance compared to traditional methods, emphasizing the benefit of incorporating multi-omics data to improve prediction accuracy. However, their models for predicting T category and pathological N category exhibited less accuracy, revealing challenges in integrating morphological and genetic data for these features.

Park et al.[13] applied deep learning to classify lymph node images and predict nodal metastasis. Their study split into training and validation cohorts, provided significant insights into model robustness, with notable improvements in AUC from 0.72 to 0.76 across these groups. The comparison of different predictors highlighted that integrated models combining clinical features and deep learning features offered higher discriminative power for both overall survival and recurrence-free survival, underscoring the advantage of incorporating diverse data modalities in predictive models.

Shahriarirad et al.[11] demonstrated the advantage of TabNet models over logistic regression in predicting sentinel LN involvement, achieving superior AUCs and metrics such as sensitivity and specificity. This suggests that complex models like TabNet, which account for multiple factors including vascular invasion and tumor size, are better suited for capturing the nuanced features contributing to metastasis prediction. Their results align with those from Zhou et al.[12], which similarly demonstrated the strong performance of deep learning models (AUC up to 0.90) in predicting lymph node metastasis from breast cancer images. The importance of features like vascular invasion in Shahriarirad et al.'s[11] study aligns with the findings from Xu Feng et al.[19], indicating its significant role across various deep learning models.

Shiner et al.[16] explored machine learning classifiers for different distant metastasis sites and found that models performed similarly across training and independent test sets with AUCs around 0.74 to 0.75. This underscores the challenge of accurately predicting metastatic risk in breast cancer patients, especially when shifting focus from regional LN metastasis. Their findings highlight the potential limitations of single-site models and the need for more

robust feature selection strategies in order to enhance prediction accuracy. This aligns with the approach taken by Polat et al.[13] and Shahriarirad et al.[11], which combined multiple predictive factors for better performance. The comparison of the performance of our neural network model developed in the present study with similar previous studies is illustrated in Table 3.

Overall, these studies collectively highlight the significant advancements in the use of deep learning and multimodal approaches for predicting lymph node metastasis in breast cancer. The integration of traditional clinicopathological variables with advanced machine-learning methods enhances the predictive accuracy and utility of these models in clinical practice. While Chen et al.[14] focused on conventional clinicopathological factors, the other studies incorporated machine learning and multi-omics data, emphasizing the evolution towards more sophisticated predictive tools. Despite some model-specific limitations, such as overfitting in small subgroups, integrating these models into clinical practice holds promise for personalized breast cancer management.

Study's limitations

The hospital in which this study was conducted is a tertiary-care hospital and medical college, catering to a large and varied population. However, in the future, large-scale multicenter studies should be undertaken in order to generate a generalized model. Incorporating imaging and multi-omics data alongside clinicopathological variables could improve predictive performance. Model performance can also be enhanced by employing more advanced neural network architectures, such as convolutional or transformer-based networks, which can capture complex relationships within the data.

Table 3 Comparison of model's performance with similar previous studies

Study	Data type	Model type	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	F1-Score	Remarks
Present Study	Clinicopathological	Neural Network	78.3	0.72	40	70	0.83	Single-center study
Chen et al.[14]	Clinicopathological	Logistic Regression	78.6	0.786	75	82	0.85	Multicenter study
Polat et al.[13]	DCE-MRI+Clinical	4D Hybrid CNN	78.3	0.87 (pN)	76	85	Not reported	Single-center study
Zhou et al.[12]	Ultrasound Images	Inception V3	Not reported	0.90	89	80	Not reported	Single-center
Xu et al.[19]	WSIs+Clinicopathological	VGG16_BN (AMIL Framework)	83.1	0.831	81	86	0.87	Single-center
Shahriarirad et al.[11]	Clinicopathological	TabNet	75	Not reported	72	78	0.76	Multicenter
Ding et al.[20]	WSIs+Clinicopathological	Multi-Modal Model	80.9	0.809	79	82	Not reported	Multicenter
Park et al.[17]	Preprocessed CT Images	DenseNet121, Ensemble Models	93.8	0.968	91	96	0.92	Single-center
Shiner et al.[16]	Clinical Data	Gradient Boosting Machine (GBM)	74	0.75	72	76	Not reported	Single-center

CT=computed tomography, VGG16_BN=visual geometry group 16 batch normalization, CNN=convolutional neural network

Conclusion

This study highlights the significant potential of deep learning models in predicting regional LN metastasis in IDC of the breast using clinicopathological features. The findings reinforce the importance of clinicopathological factors such as tumor volume, hormone receptor status, HER2 expression, Ki-67 index, and microvessel density as predictive variables for LN metastasis. The neural network model, with an accuracy of 78.3%, demonstrates the feasibility of integrating artificial intelligence into oncology diagnostics to enhance predictive accuracy. Incorporating additional data modalities, such as imaging and molecular profiling, may enhance the model's performance and generalizability across diverse patient populations.

This study advances the application of deep learning in breast cancer management by providing an innovative tool for predicting LN metastasis. Its integration into clinical workflows could improve staging precision, inform treatment strategies, and contribute to personalized care, ultimately improving patient outcomes. Future research should focus on expanding the dataset, addressing limitations, and exploring multimodal data integration in order to further enhance predictive accuracy and clinical utility.

Ethical approval

The present study was conducted after receiving approval from the Institutional Ethics Committee of NDMC Medical College and Hindu Rao Hospital, Delhi (IEC/NDMC/2022/109 dated 17.06.2022). Written informed consent was obtained from the patients.

Author contributions

S.K.S.: Data curation, Literature search, Formal analysis, Investigation, Methodology, Validation, Writing-original draft, Writing-review & editing; S.S., R.C.: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing-review & editing; S.K.: Literature search, Data curation, Methodology, Validation, Writing-review & editing. All authors read and approved the submitted version.

Data availability

Datasets are available on request.

Financial disclosure

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgement

The authors would like to thank everyone who participated in this study.

Funding sources

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 Cancers in 185 countries. CA Cancer J Clin. 2021;71:209-49. doi: 10.3322/caac.21660.
2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023;73:17-48. doi: 10.3322/caac.21763
3. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nat Rev Dis Primers 2019;5:66. doi: 10.1038/s41572-019-0111-2.
4. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52. doi: 10.1038/35021093.

5. Natale G, Stouthandel MEJ, Van Hoof T, Bocci G. The lymphatic system in breast cancer: anatomical and molecular approaches. Medicina (Kaunas) 2021;57:1272. doi: 10.3390/medicina57111272.
6. Liu J, Li Y, Zhang W, Yang C, Yang C, Chen L, et al. The prognostic role of lymph node ratio in breast cancer patients received neoadjuvant chemotherapy: a dose-response meta-analysis. Front Surg 26;9:971030. doi: 10.3389/fsurg.2022.971030.
7. Kim JY, Ryu MR, Choi BO, Park WC, Oh SJ, Won JM, et al. The prognostic significance of the lymph node ratio in axillary lymph node positive breast cancer. J Breast Cancer 2011;14:204-12. doi: 10.4048/jbc.2011.14.3.204.
8. Ozmen V, Cabioglu N. Sentinel lymph node biopsy for breast cancer: current controversies. Breast J 2006;12(Suppl 2):S134-42. doi: 10.1111/j.1075-122X.2006.00327.x.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44. doi: 10.1038/nature14539.
10. Ahn JS, Shin S, Yang SA, Park EK, Kim KH, Cho SI, et al. Artificial intelligence in breast cancer diagnosis and personalized medicine. J Breast Cancer 2023;26:405-35. doi: 10.4048/jbc.2023.26.e45.
11. Shahriarirad R, Meshkati Yazd SM, Fathian R, Fallahi M, Ghadiani Z, Nafissi N. Prediction of sentinel lymph node metastasis in breast cancer patients based on preoperative features: a deep machine learning approach. Sci Rep 2024;14:1351. doi: 10.1038/s41598-024-51244-y.
12. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph node metastasis prediction from primary breast cancer us images using deep learning. Radiology 2020;294:19-28. doi: 10.1148/radiol.2019190372.
13. Polat DS, Nguyen S, Karbasi P, Hulsey K, Cobanoglu MC, Wang L, et al. Machine learning prediction of lymph node metastasis in breast cancer: performance of a multi-institutional MRI-based 4D convolutional neural network. Radiol Imaging Cancer 2024;6:e230107. doi: 10.1148/rycan.230107.
14. Chen W, Wang C, Fu F, Yang B, Chen C, Sun Y. A model to predict the risk of lymph node metastasis in breast cancer based on clinicopathological characteristics. Cancer Manag Res 2020;12:10439-47. doi: 10.2147/CMAR.S272420.
15. Dihge L, Vallon-Christersson J, Hegardt C, Saal LH, Häkkinen J, Larsson C, et al. Prediction of lymph node metastasis in breast cancer by gene expression and clinicopathological models: development and validation within a population-based cohort. Clin Cancer Res 2019;25:6368-81. doi: 10.1158/1078-0432.16.
16. Shiner A, Kiss A, Saednia K, Jerzak KJ, Gandhi S, Lu FI, et al. Predicting patterns of distant metastasis in breast cancer patients following local regional therapy using machine learning. Genes (Basel) 2023;14:1768. doi: 10.3390/genes14091768.
17. Park TY, Kwon LM, Hyeon J, Cho BJ, Kim BJ. Deep learning prediction of axillary lymph node metastasis in breast cancer patients using clinical implication-applied preprocessed CT Images. Curr Oncol 2024;31:2278-88. doi: 10.3390/curroncol31040169.
18. Hu J, Lv H, Zhao S, Lin CJ, Su GH, Shao ZM. Prediction of clinicopathological features, multi-omics events and prognosis based on digital pathology and deep learning in HR^{+}/$HER2^{-}$ breast cancer. J Thorac Dis 2023;15:2528-43. doi: 10.21037/jtd-23-445.
19. Xu F, Zhu C, Tang W, Wang Y, Zhang Y, Li J, et al. Predicting axillary lymph node metastasis in early breast cancer using deep learning on primary tumor biopsy slides. Front Oncol 2021 14;11:759007. doi: 10.3389/fonc.2021.759007.
20. Ding Y, Yang F, Han M, Li C, Wang Y, Xu X, et al. Multi-center study on predicting breast cancer lymph node status from core needle biopsy specimens using multi-modal and multi-instance deep learning. npj Breast Cancer 2023;9:58. doi: 10.1038/s41523-023-00562-x.
21. Shouket T, Mahmood S, Hassan MT, Iftikhar A. Overall and disease-free survival prediction of postoperative breast cancer patients using machine learning techniques. In: 2019 22nd International Multitopic Conference (INMIC). Islamabad: IEEE; 2019;p.1–6. doi: 10.1109/INMIC48123.2019.9022756.
22. Kalafi EY, Nor NAM, Taib NA, Ganggayah MD, Town C, Dhillon SK. Machine learning and deep learning approaches in breast cancer survival prediction using clinical data. Folia Biol (Praha) 2019;65:212-20.
23. Fu B, Liu P, Lin J, Deng L, Hu K, Zheng H. Predicting invasive disease-free survival for early-stage breast cancer patients using follow-up clinical data. IEEE Trans Biomed Eng 2018. doi: 10.1109/TBME.2018.2882867.
24. Sun D, Li A, Tang B, Wang M. Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome. Comput Methods Programs Biomed 2018;161:45-53. doi: 10.1016/j.cmpb.2018.04.008.