

Machine Learning Based Techniques for the Prediction of Axillary Lymph Node Metastases in Early Breast Cancer

Maisam Ali*, Muhammad Yaseen *, Sikandar Ali, Hee-Cheol Kim *

**Dept. of Digital Anti-Aging Healthcare, Inje University Gimhae, Republic of Korea*

maisamali053@gmail.com, shigriyaseen@gmail.com, sikandershigri77@gmail.com heeki@inje.ac.kr

Abstract— One of the most significant considerations in determining the prognosis of breast cancer is the involvement of lymph nodes. Although non-invasive imaging techniques like ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and F-18 fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET)/CT have been recommended for the assessment of the ALN status, their diagnostic performance lacks sufficient sensitivity in the detection of ALN metastasis. Accurate machine learning based techniques have been applied to the investigation of lymph node in the early stage of breast cancer. This occurrence is influenced by several tumor-specific elements. The primary objective of this research was to determine the clinical and pathological variables that validate the radiomics-based machine learning model can be used to predict whether individuals with early-stage breast cancer will have metastases in their axillary lymph nodes (ALNM). Various aspects were considered and analyzed while utilizing various machine learning algorithms to determine the involvement of the lymph node metastasis. For our experiments, we have examined three machine learning models such as Gradient Boosting Machine (GBM), KNN (K-Nearest Neighbor) and Decision Tree (DT). Accuracy, specificity, sensitivity, and F1-score were used to evaluate the performance of all three models. The accuracy of GBM was 97%, followed by that of KNN and Decision Tree (DT), which were 91% and 91%, respectively.

Keywords— Lymph node metastasis, Breast cancer, Machine learning, Deep learning, Gradient Boosting Machine (GBM).

I. INTRODUCTION

Despite acknowledging that breast cancer is a heterogeneous disease with diverse prognoses, developments in research and public health initiatives regarding cancer patient identification and therapy have increased survival rates globally [1]. One of the most significant prognostic factors influencing the long-term prognosis of breast cancer is an accurate preoperative assessment of axillary lymph-node (ALN) status. Although non-invasive imaging methods such as ultrasonography, computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET)/CT are available to evaluate the status of axillary lymph nodes (ALNs), these techniques are insufficiently sensitive to identify ALN metastasis. In the 19th century, the biological concept of breast cancer predominated. It described the

disease as a local occurrence inside the breast that propagated centrifugally along lymphatics to first and then second echelon nodes, with increasing systemic metastasis. A beneficial outcome of this strategy was radical local surgery, or the Halstead radical mastectomy [2]. To correctly stage the patient and achieve adequate local control, axillary lymph node dissection (ALND), which determines the ALN status, has been the standard of care for patients with invasive breast cancer. Except for T4 tumours, sentinel lymph node biopsy (SLNB), an alternative to ALND, has recently come to be the norm for axillary staging in all clinically node-negative patients (as determined by clinical examination, ultrasonography, and/or fine needle aspiration cytology). The staging of the axilla using the SLN approach has been shown to be possible, accurate, and adequate while avoiding the morbidity of an ALND. [3,4]. ALN dissection has been commonly used for the comprehensive examination of ALN status. However, it is unnecessary in patients without ALN metastases and can have serious adverse effects like lymphedema. Although sentinel lymph node biopsy, the standard procedure for determining the status of the axillary lymph nodes in patients with early breast cancer, has been associated to a false-negative rate of 5–10% [5]. The purpose of this study is to assess the probability of breast cancer will spread to the axillary lymph node. The method carried out by implementing the most well-known machine learning algorithms can produce an individual risk assessment for every patient, which helps doctors in making shrewd decisions about minimizing consequences from the spread of cancer to the lymph nodes and improving treatment according to each patient's specific needs.

II. RELATED WORKS

Several research attempts have been made by researchers to investigate axillary lymph node metastasis. All patients had managed to keep their blood glucose levels below 8.3 mmol/L (150 mg/dL) and had fasted for at least six hours before receiving F-18 FDG. The intravenous dose of 5.5 MBq/kg of F-18 FDG was followed by the PET/CT scan 60 minutes later. A

Discovery STE PET/CT scanner from GE Healthcare, Milwaukee, Wisconsin, was used to do the scans within a month of the surgical procedure. The cranial vertex to the proximal thighs were first imaged with a low-dose CT scan (peak voltage of 120 kVp, automatic tube current of 60–150 mA, and slice thickness of 3.75 mm) without contrast enhancement for attenuation correction. Following the acquisition of the CT scan, a 3-dimensional PET scan was conducted with an acquisition period of 3 minutes per bed position [6]. A retrospective review of all F-18 FDG PET/CT images was conducted using the AW server 3.2 (GE Healthcare, Milwaukee, WI, USA). To accurately identify genuine IDC lesions, the F-18 FDG PET/CT images were visually analyzed and compared with breast MRI images. The maximal SUV (SUVmax) was calculated using the following formula: $SUV_{max} = \text{maximum activity in region of interest (MBq/g)} / [\text{injected dosage (MBq)} / \text{body weight (g)}]$. In cases of ALN, focally enhanced F-18 FDG avid ALNs with a high F-18 FDG uptake ($SUV_{max} \geq 2$) were considered favorable findings, as a cut-off of 2.0 has been shown to exhibit the best accuracy in diagnosing ALN metastases in breast cancer [7]. The results showed that F-18 FDG PET/CT had a sensitivity of 55.8%, specificity of 93.0%, accuracy of 77.0%, positive predictive value (PPV) of 85.7%, and negative predictive value (NPV) of 73.6% for detecting ALN metastases.

III. MATERIALS AND METHODS

This section describes the materials and methods we used in our experiment.

A. Data Source

We used a data set from DAICON for this research study. The data set contains 549 patients' information with 28 different features.

B. Data Preparation

Data preparation is one of the key methods used in deep learning and machine learning. The degree to which the data has been pre-processed typically determines how accurate your model will be. In other words, the quality of your model is strongly correlated with the preprocessing of the data. Our dataset contained categorical and non-categorical values as well as null values and outliers. We eliminated outliers and null values. The categorical values and non-categorical values were converted using one-hot encoding and label encoding. We obtained balanced and clean data for further processing after the dataset underwent preprocessing. So that models could be used, we converted the raw data into an appropriate form. To choose the most pertinent features, we used an additional technique for tree feature selection. In our experiment, we used 21 attributes to facilitate the implementation of different machine learning models and to further develop explainable AI for the support and better

representation of this research study. Table 1 below displays the features information and descriptions.

TABLE 1. Feature descriptions of the dataset

	Features	Feature type
1	Age	Continuous
2	Diagnostic Name	Categorical
3	Cancer Location	Continuous
4	Cancer Size	Continuous
5	NG	Continuous
6	HG	Continuous
7	HG_score_1	Continuous
8	HG_score_2	Continuous
9	HG_score_3	Continuous
10	DCIS_or_LCIS	Categorical
11	DCIS_or_LCIS_type	Categorical
12	N_category	Categorical
13	ER	Categorical
14	ER_Allred_score	Continuous
15	PR	Categorical
16	PR_Allred_score	Continuous
17	KI-67 LI percent	Continuous
18	HER2	Categorical
19	HER2_IHC	Categorical
20	HER2_SISH	Categorical
21	HER2_SISH_ratio	Categorical

C. Machine Learning Algorithms

For the development of our prediction model, we used three different machine learning models. Based on input features, a lymph node metastatic decision tree classifier determines if a patient has lymph node metastasis. The decision tree algorithm develops a set of rules to categorize patients into two groups based on factors including tumour size, tumour location, and patient age. These groups are either patients with lymph node metastases or not. With a combination of decision trees that determine the mode or mean/average of each tree, RF is an ensemble learning approach for classification, regression, and other use cases [8]. The Gradient Boosting Machine (GBM) [9], by iteratively training decision trees to compensate for prediction errors, GBM for lymph node metastasis forecasts the likelihood of metastasis. Starting with a single prediction, it then aggregates forecasts, fits a tree to residuals, and repeats. The model is improved through this iterative process, which captures complicated patterns and yields a reliable assessment of the chance of lymph node metastasis. In the evaluation of lymph node metastasis, K-Nearest Neighbours (KNN) [10], predicts if a patient has metastasis by seeking the 'k' nearest patients in the dataset with comparable variables (such as tumour size, grade, etc.). The assumption is based on the dominant class among these neighbours. KNN is non-parametric, logical, and makes few assumptions about the data, enabling simple interpretation. In the case of lymph node metastasis

investigations, however, choosing an adequate "k" value and distance metric is essential for reliable predictions. We compared these models' performance metrics, including accuracy, specificity, sensitivity, and F1-score.

D. Performance Measures

The performance of machine learning models can be evaluated using certain metrics and instruments. The performance of models is typically measured using metrics like accuracy, specificity, sensitivity, and F1-score. These criteria were also used to assess how well our models performed, and the findings are fairly positive. Accuracy is the degree to which measurement results are close to the actual value. In other words, it is only the proportion of observations that were correctly predicted to all the observations. The percentage of relevant occurrences among the examples that were retrieved is known as specificity, which is exactness. The ratio of accurately anticipated positive observations to all the actual class observations is known as sensitivity. The sensitivity and specificity weighted average is known as the F1-score.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-Score} = \frac{2(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (4)$$

Where TP, TN, FP, FN denoted as true positive, true negative, false positive, false negative, respectively.

IV. OVERALL EXPERIMENTAL WORKFLOW

In this research, we initially retrieved the data from the Dacon dataset and performed preprocessing to remove outliers and null values from the data. After eliminating data abnormalities, we received 549 records. When preprocessing the data, we used label encoding and one hot encoding. We also chose 21 features for our experiment after selecting other features. The data was then divided into 20% for testing and 80% for training. We used Decision Tree, KNNNeighbor, and Gradient Boosting Machine. It was determined how to measure performance using metrics like accuracy, sensitivity, specificity, and F1-scores. In this study, the performance measures' graphical representation and ROC were also computed. Fig. 1 displays the overall process used in this research.

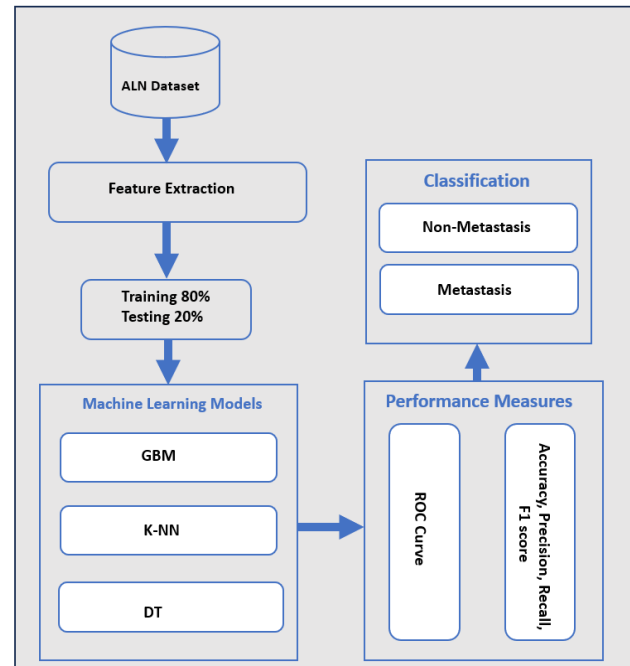


Figure 1. Overall workflow of the classification of the axillary lymph node metastasis.

V. RESULTS AND DISCUSSION

After removing the data that had missing values and did not satisfy the criteria set out by our models, we included 549 patient records in the analysis. After preprocessing the data, we chose 21 features for our investigation. A training set comprised 80% of the dataset, and a testing set comprised 20%. The gradient boosting machine, KNNneighbor, and decision tree were the three machine learning models we used. Interesting findings were obtained. Each model's accuracy, specificity, and recall were calculated. The performance metric for each of the three used machine learning algorithms is displayed in Table. 2

Table.2 The Overall Performance of all the Three Applied Machine Learning Models being Used for the Prediction and Classification of ALN metastasis.

Classifiers	Accuracy	Specificity	Sensitivity	F1 score
GBM	0.97	0.93	0.89	0.91
KNN	0.91	0.96	0.67	0.79
DT	0.91	0.90	0.87	0.88

In this study, we trained the machine learning ML models to predict the ALN metastasis of breast cancer patients using clinical informations. The GBM model outperformed the other three machine learning models, with an AUC of 0.97%, demonstrating strong discrimination performance. Fig 2 shows the ROC curve of all the three applied machine learning algorithms.

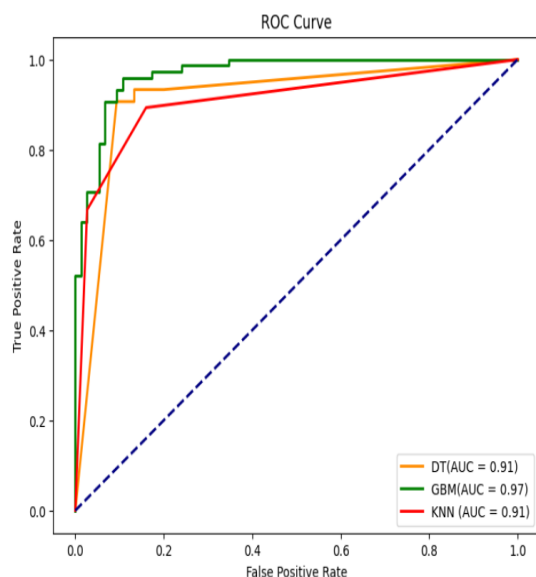


Figure 2. Receiver operating characteristics (ROC) of Classification Models

VI. CONCLUSION

In this study, the classification, and the prediction of the involvement of an axillary lymph node metastasis could be precisely assessed by three different machine learning models. In a conclusion, our models offer a precise and understandable prediction result that takes into consideration the distinctive characteristics of each patient and can assist doctors in customizing the treatment of patients with axillary lymph node metastasis in accordance with the patients' specific prognoses.

ACKNOWLEDGMENT

This research was supported by a grant from Korea Health Technology R&D Project through the Korea Health Industry Development Institute validation (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant No: HI21C0977).

REFERENCES

- [1] Arnold M, Rutherford MJ, Bardot A, Ferlay J, Andersson TM-L, Myklebust TÅ, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol.* (2019) 20:1493–505. doi: 10.1016/S1470-2045(19)30456-5
- [2] Fisher, B. "Seminars of Bernard Fisher 1960-nature of cancer as systemic disease." *Bull Soc Int Chir* 31.6 (1972): 604-9.
- [3] Peeters, M., et al. "Guideline for the esophageal and gastric cancer: scientific support of the College of Oncology." *Good Clinical Practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE)* 21.03 (2008): 2008.
- [4] Rietman, J. S., Dijkstra, P. U., Geertzen, J. H., Baas, P., De Vries, J., Dolsma, W., ... & Hoekstra, H. J. (2003). Short-term morbidity of the upper limb after sentinel lymph node biopsy or axillary lymph node dissection for Stage I or II breast carcinoma. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 98(4), 690-696.
- [5] Krag DN, Anderson SJ, Julian TB, Brown AM, Harlow SP, Ashikaga T, et al. Technical outcomes of sentinel-lymph-node resection and conventional axillary-lymph-node dissection in patients with clinically node-negative breast cancer: results from the NSABP B-32 randomised phase III trial. *Lancet Oncol.* 2007;8:881–8. Song, Bong-Il. "A machine learning-based radiomics model for the prediction of axillary lymph-node metastasis in breast cancer." *Breast Cancer* 28 (2021): 664-671.
- [6] Taira N, Ohsumi S, Takabatake D, Hara F, Takashima S, Aogi K, et al. Determination of indication for sentinel lymph node biopsy in clinical node-negative breast cancer using preoperative 18F-fluorodeoxyglucose positron emission tomography/computed tomography fusion imaging. *Jpn J Clin Oncol.* 2009;39:16–21.
- [7] Carkaci S, Adrada BE, Rohren E, Wei W, Quraishi MA, Mawlawi O, et al. Semiquantitative analysis of maximum standardized uptake values of regional lymph nodes in inflammatory breast cancer: is there a reliable threshold for differentiating benign from malignant? *Acad Radiol.* 2012;19:535–41.
- [8] Breiman, L. Random Forests. *Mach. Learn.* 2001, 45, 5–32. [Google Scholar] [CrossRef]
- [9] Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; pp. 785–794. [Google Scholar]
- [10] Xu, Z., Xie, Y., Wu, L., Chen, M., Shi, Z., Cui, Y., ... & Liu, Z. (2023). Using Machine Learning Methods to Assess Lympho vascular Invasion and Survival in Breast Cancer: Performance of Combining Preoperative Clinical and MRI Characteristics. *Journal of Magnetic Resonance Imaging*.



Mr. Maisam Ali received his B.E degree in Electrical and communication Engineering from Hamdard University, Pakistan. He is currently pursuing his master's degree from Inje University, South Korea. His research interests are artificial intelligence, machine learning, deep learning, computer vision.

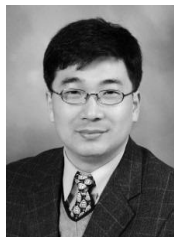


Mr. Muhammad Yaseen received his B.E degree in Electrical Engineering from Hamdard University, Pakistan. He is currently pursuing his master's degree from Inje University, South Korea. His research interests are artificial intelligence, machine learning, deep learning, computer vision.



Mr. Sikandar Ali received his B.E. degree in Computer Engineering from Mehran University of Engineering & Technology, Pakistan. He got his MS from the Department of Computer Science from Chung Buk National University, the Republic of Korea. Furthermore, he is now a Ph.D. candidate at Inje University South Korea in the department of Digital Anti-Aging Healthcare. His research interests include

artificial intelligence, data science, big data, machine learning, deep learning, reinforcement learning, Computer vision, and medical imaging.



Prof. Hee-Cheol Kim received his BSc at the Department of Mathematics, MSc at the Department of Computer Science in Sogang University in Korea, and Ph.D. at Numerical Analysis and Computing Science, Stockholm University in Sweden in 2001. He is a Professor at the Department of Computer Engineering and Head of the Institute of Digital Anti-aging Healthcare, Inje University in Korea. His research interests include machine learning, deep

learning, Computer vision, and medical informatics.