Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Evaluation of conventional histopathological scoring in breast carcinoma using artificial intelligence technologies

Seda Cetinkaya Karabekir [a,*], Aylin Gokhan [b], Hilal Arslan [a], Burak Keskin [c], Adile Ferda Dagli [d]

[a] İzmir Bakırçay University, Faculty of Medicine, Department of Histology and Embryology, İzmir, Türkiye
[b] Ege University, Faculty of Medicine, Department of Histology and Embryology, İzmir, Türkiye
[c] İzmir Bakırçay University, Department of Management Information Systems, İzmir, Türkiye
[d] İzmir Bakırçay University, Faculty of Medicine, Department of Medical Pathology, İzmir, Türkiye
*Corresponding author: seda.karabekir@bakircay.edu.tr (Seda Cetinkaya Karabekir)

MAIN POINTS

- A deep learning–based model was developed to evaluate the Ki-67 proliferation index in breast carcinoma and compared with manual assessment.
- The model showed high accuracy and low error rates in low-proliferation cases (<20% Ki-67) (MAE: 5.31%; accuracy: 80%).
- AI-assisted evaluation has the potential to reduce interobserver variability, minimize human error, and ease the workload of pathologists during the preliminary assessment phase.
- Error analysis in high-proliferation samples revealed segmentation challenges due to dense nuclear clustering and faint immunonegative staining, indicating that advanced image-processing techniques could further improve model performance.

Cite this article as: Cetinkaya Karabekir S, Gokhan A, Arslan H, Keskin B, Dagli AF. Evaluation of conventional histopathological scoring in breast carcinoma using artificial intelligence technologies. *Ann Med Res*. 2026;33(1):33--42. doi: 10.5455/annalsmedres.2025.08.244.

ABSTRACT

Aim: Breast cancer is the most commonly diagnosed malignancy in women, and early detection plays a critical role in the success of treatment. The Ki-67 proliferation index is widely used to evaluate tumor cell proliferation; however, its manual scoring process is observer-dependent, time-consuming, and inherently subjective. This study aims to assess Ki-67 immunohistochemical staining using deep learning algorithms in an objective, rapid, and reproducible manner, and to compare the model's performance with conventional scoring methods.

Materials and Methods: In the first phase of the study, a dataset was created using digital images of Ki-67-stained histological sections obtained from patients diagnosed with breast cancer. These images were used to train a machine learning algorithm. In the second phase, 50 new Ki-67-stained tissue sections previously unseen by the model were digitized, and the model's predictions were compared with Ki-67 index values calculated by conventional manual assessment.

Results: The developed model achieved a mean absolute error (MAE) of 8.69%, a root mean square error (RMSE) of 13.00%, and a coefficient of determination (R^2) of 0.540 in overall prediction performance. For cases with low proliferation (Ki-67<20%), the model demonstrated high accuracy (MAE: 5.31%). Binary classification based on a 20% threshold yielded 80% accuracy, 80% sensitivity, 90% precision, and an F1 score of 0.84.

Conclusion: The use of artificial intelligence algorithms in Ki-67 assessment demonstrated successful performance, with an MAE of 8.69%, and has the potential to reduce pathologists' workload during the preliminary evaluation phase. The findings suggest that, with further refinement, the proposed model could contribute to more objective, consistent, and reproducible assessments in breast cancer diagnostics.

Keywords: Breast cancer, Ki-67, Artificial intelligence, Deep learning, Artificial neural network

Received: Aug 29, 2025 **Accepted:** Oct 31, 2025 **Available Online:** Jan 26, 2026

INTRODUCTION

Each year more than half a million women worldwide die from breast cancer. To reduce mortality associated with breast cancer, many countries have implemented screening programs over the past two decades. These programs have led to a nearly 30% reduction in breast cancer-related deaths by enabling earlier diagnosis, combined with advances in treatment. Nevertheless, breast cancer remains the leading cause of cancer-related mortality among women [1]. Breast cancer encompasses various subtypes [2]. According to the World Health Organization (WHO) Classification of Tumours of the Breast (5th edition, 2019), breast cancer is categorized into well-defined histological subtypes. These include ductal carcinoma in situ (DCIS), invasive carcinoma of no special type (NST), invasive lobular carcinoma, tubular carcinoma, mucinous carcinoma, medullary carcinoma, apocrine

https://doi.org/10.5455/annalsmedres.2025.08.244

carcinoma, metaplastic carcinoma, and other rare variants. In addition, molecular subtypes such as hormone receptor-positive/HER2-negative, HER2-positive, and triple-negative breast cancers are also recognized due to their prognostic and therapeutic significance. This classification provides a more standardized approach to diagnosis, research, and treatment planning [3].

Early and definitive diagnosis of breast cancer, including the identification of its specific subtype, is crucial in preventing disease progression and associated complications. Accurate diagnosis enables timely and effective treatment planning and ultimately reduces breast cancer-related mortality. Various imaging modalities are used for breast cancer detection, including mammography (MG) [4], breast thermography (BT) [5], magnetic resonance imaging (MRI) [6], positron emission tomography (PET), computed tomography (CT) [7], ultrasound (US) [8] and histopathology (HP) [9]. These methods are widely used for detecting breast cancer at early stages [10-12].

Among these, histopathological evaluation plays a critical role in diagnosing, staging, and determination of breast cancer by analyzing tissue specimens. Accurate staging of the cancer is essential for selecting the most appropriate treatment strategy. During this process, certain prognostic markers such as the proliferation index (PI) and mitotic activity are used to evaluate tumor biology. These markers help predict disease progression and guide treatment decisions. Ki-67 is a nuclear antigen expressed in proliferating but not quiescent cells, making it a reliable marker for calculating the PI. The Ki-67 proliferation index is determined by evaluating immunohistochemically stained tumor sections under a microscope. The most densely stained region, referred to as the "hot spot," is selected, and the proportion of Ki-67-positive nuclei relative to the total number of nuclei is calculated and reported as the proliferation index [13, 14].

However, histopathological assessment of breast tissue is highly subjective and depends on the experience of the pathologist [15-17]. Manual Ki-67 scoring is associated with significant interobserver variability, particularly around key clinical thresholds (e.g., 14% or 20%), and lacks standardized criteria for the number of high-power fields to be evaluated [18]. These limitations introduce uncertainty into molecular subtype classification and therapeutic decision-making [17]. Therefore, there is growing interest in the application of artificial intelligence (AI)-based approaches to reduce subjectivity in pathological assessment, particularly in breast pathology [15-17].

Numerous AI-based tools utilizing machine learning (ML) and deep learning (DL) have been developed for tasks such as breast cancer classification, detection, and segmentation [19].

Recent comprehensive reviews and applied studies highlight that artificial intelligence can be integrated into all stages of breast pathology, including classification, grading, biomarker quantification, and risk prediction; in particular, CNN and Vision Transformer (ViT) based approaches have been shown to provide robust feature learning at the whole-slide image (WSI) level [20-22]. In parallel, methods have been reported for automatic scoring and risk stratification in fine distinctions such as HER2-low status and in key biomarkers such as Ki-67 [23]. Nevertheless, significant barriers to routine clinical adoption remain, including validation, generalizability, and standardization. Studies published in 2024–2025 emphasize that digital image analysis tools especially for Ki-67 evaluation have not yet provided sufficient evidence to be deemed fully suitable for clinical practice, underscoring the need for large-scale, multicenter validation [24, 25].

Artificial intelligence was first conceptualized by Alan Turing in 1950 and later defined by John McCarthy in 1956 as “the science and engineering of making intelligent machines [26, 27]. Today, AI represents a broad domain within computer science aimed at developing machines capable of performing tasks that typically require human intelligence [28]. These systems can store experiences, learn from them, reason, create, judge, and make decisions. A critical subset of AI, machine learning (ML)—particularly in conjunction with deep learning (DL)—has enabled significant progress in image processing [29].

DL employs multilayered architectures such as convolutional neural networks (CNNs), which mimic human visual perception and excel at tasks such as classification and segmentation [30]. Due to their ability to extract meaningful features from raw data, DL-based models have proven highly effective in analyzing pathology images. In this context, AI applications in pathology are primarily based on the recognition of visual patterns via CNNs and the extraction of diagnostic information from digitized slides [29].

The classification, detection, and staging of breast cancer are traditionally conducted by pathologists based on theoretical knowledge and clinical experience. However, this process is labor-intensive and time-consuming. Although accumulated observational expertise is invaluable, the potential for human error and time constraints associated with manual protocols cannot be overlooked. In this context, the present study aims to investigate the feasibility of evaluating Ki-67 a key biomarker in breast carcinoma diagnosis using artificial intelligence algorithms after initial manual assessment by pathologists. This approach aims to minimize human error and enhance the reliability of pathological evaluations.

MATERIALS AND METHODS

This study was approved by the Non-Interventional Clinical Research Ethics Committee of İzmir Bakırçay University (Decision No: 1357, Date: 13.12.2023). The research was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki, with full respect for human rights.

https://doi.org/10.5455/annalsmedres.2025.08.244

Between January 2021 and March 2025, a total of 150 patients with histopathologically confirmed breast carcinoma whose Ki-67 index had been assessed at the Department of Pathology, İzmir Bakırçay University Çiğli Training and Research Hospital, were included. All slides were digitized at 40× magnification, and only cases meeting predefined technical quality criteria (adequate focus, contrast, and background) were analyzed. Inclusion in the training set additionally required the availability of archival FFPE (formalin-fixed, paraffin-embedded) tissue blocks. Cases with pronounced staining artifacts, excessive background, or other technical defects rendering images unsuitable for analysis were excluded. An a priori power analysis (G*Power 3.1) for a two-tailed Pearson correlation with $\alpha = 0.05$, power = 0.90, and an anticipated effect size $r \approx 0.26$ indicated a required total sample size of 150.

Dataset construction

The study focused on the Ki-67 proliferation index of patients with breast cancer evaluated during routine diagnostic procedures at the Department of Pathology. A total of 100 digital images of breast cancer tissue sections immunohistochemically stained for Ki-67 were used. Imaging was performed using a ZEISS Axiolab 5 digital laboratory microscope integrated with a Zeiss AxioCam ERc 5s camera, and image data were recorded using ZEN 3.4 (Blue Edition) software. These images formed the digital dataset used to train the machine learning (ML) model.

Histopathological evaluation

Immunohistochemically stained breast cancer tissue sections were evaluated manually by experienced pathologists. During the scoring process, a total of 1,000 tumor cells were counted across the tissue section by identifying nuclei showing Ki-67 positivity, and the Ki-67 proliferation index was calculated. This method, which is widely accepted in literature, is considered reliable for assessing tumor cell proliferative activity and predicting breast cancer prognosis. Counting was performed in the area of highest labeling intensity, known as the 'hot spot,' at 40× magnification.

Use of deep learning algorithms

A deep learning-based model was developed to facilitate faster, more objective, and reproducible evaluation of cell proliferation indices derived from Ki-67 immunohistochemistry (IHC) images, which are widely used in breast cancer diagnostics. For the task of cell segmentation, Cellpose—a general-purpose cell segmentation algorithm built on a convolutional neural network (CNN) architecture was employed [31]. Cellpose is capable of accurately distinguishing cellular structures such as nuclei and cytoplasm in biological images.

The images were examined by an expert pathologist, and 'hot-spot' areas with the highest labeling intensity were included in the analysis; stromal and inflammatory cells were excluded. Using the Cellpose interface, two separate models were trained: one configured to segment all cells and the other to detect only Ki-67–stained cells. During model training, all nuclei were manually annotated, and positive/negative classification was performed based on staining intensity and nuclear morphology. No pre-processing (e.g., color normalization, contrast enhancement) was applied. The hyperparameters used during training included a learning rate of 0.1, a weight decay of 0.0001, and 100 epochs. After training, both models were run for each image in the test dataset and the results were recorded. Validation of the test data was performed by the expert pathologist, after which the differences between the ground truth and the model predictions were calculated as percentages. Finally, statistical analyses were applied to demonstrate the reliability of these differences.

Model training and optimization

The model was trained using high-resolution digital images of Ki-67-stained tissue sections obtained from retrospectively confirmed cases of breast carcinoma. No preprocessing techniques (e.g., contrast enhancement, color normalization, noise reduction) were applied to maintain the visual fidelity of images as perceived by pathologists and to ensure fair comparison between manual and automated assessments.

Cellpose, an open-source segmentation algorithm, was deployed in a local Python environment on a personal computer. Leveraging its flexibility, two separate models were trained using both pre-trained weights and task-specific datasets:

1. A general model for detecting and counting all cells.

2. A specialized model for identifying and counting only Ki-67-positive cells (Figure 1).

The Ki-67 proliferation index for each image was calculated as the ratio between the outputs of these two models.

Training was conducted using a supervised learning approach on 100 Ki-67-stained histological images acquired at 40x magnification. All cells and Ki-67-positive nuclei in these images were manually annotated, and the resulting segmentation masks were formatted in .npy files for model input (Figure 2).

During training, model weights and hyperparameters were optimized to enhance segmentation performance. Training was executed on a personal computer equipped with an Intel i7-13650HX processor, 32 GB RAM, 512 GB SSD, and an NVIDIA RTX 4060 GPU (Figure 2).

Model validation with test dataset

Model validation was performed using an independent test dataset curated and verified by expert pathologists. This dataset comprised previously unseen samples not used during training. Accordingly, paraffin blocks from 50 histologically confirmed breast cancer cases were retrieved from the

https://doi.org/10.5455/annalsmedres.2025.08.244

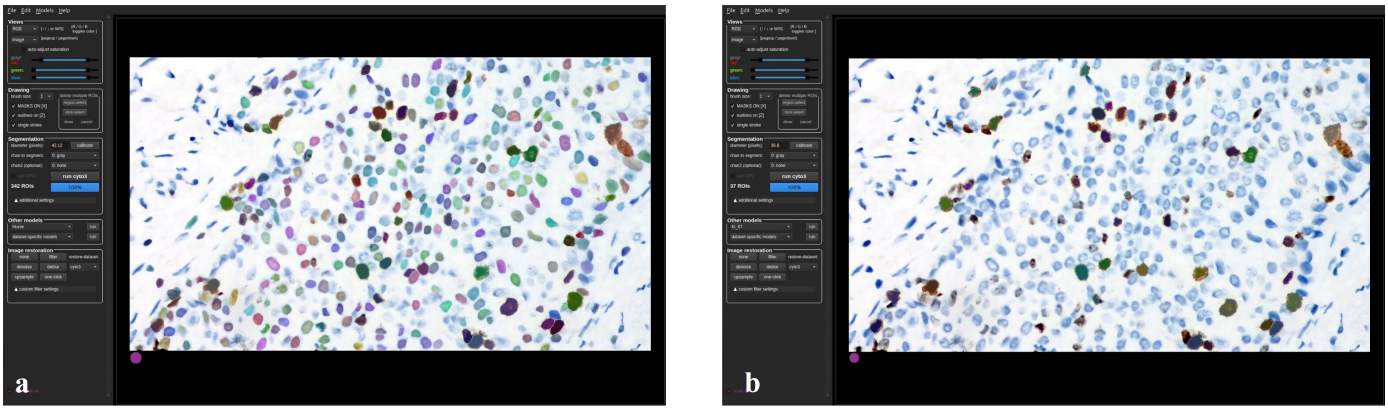

Figure 1. Screenshots of model outputs: (a) model detecting and counting all cells in the image, (b) specialized model detecting and counting only Ki-67–positive cells.

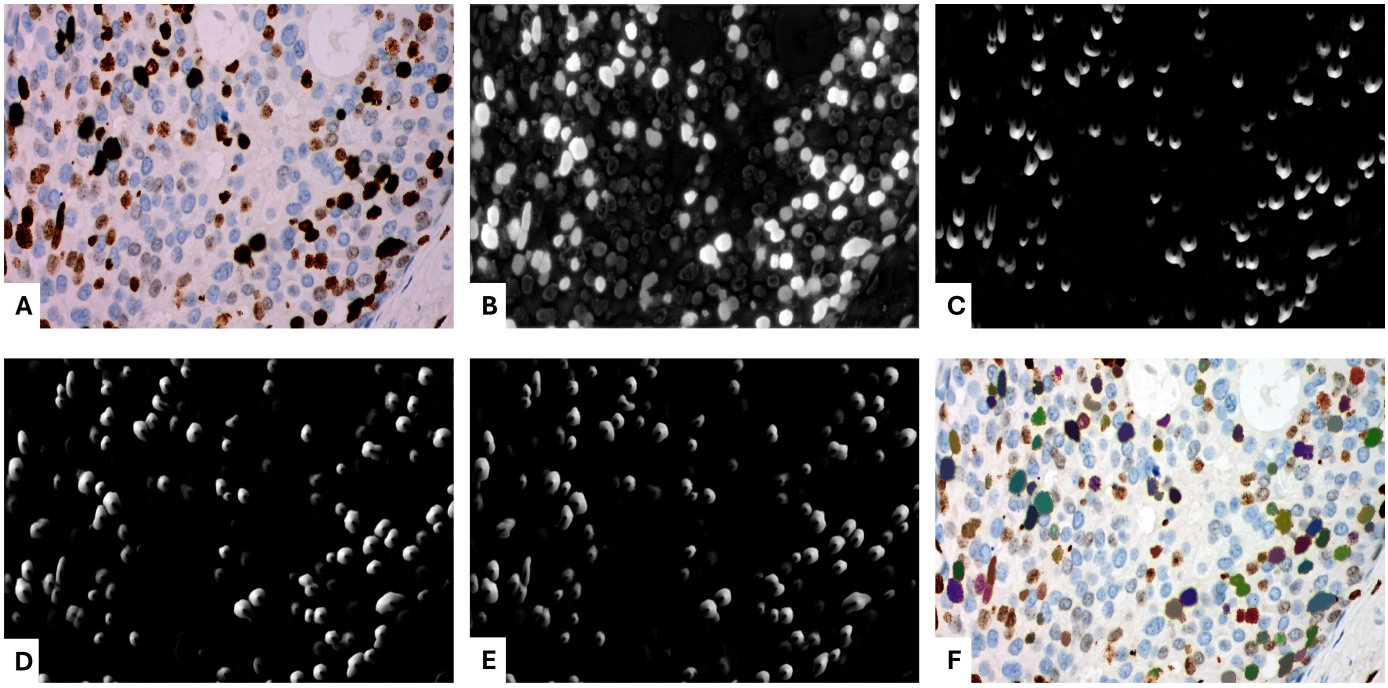

Figure 2. Diagram illustrating the workflow of the AI model from patch extraction to cell classification. (A) Original image; (B–E) segmentation steps including patch extraction and masking; (F) final output image showing the detection and separation of Ki-67–positive (brown-stained) and negative cells.

pathology archives of İzmir Bakırçay University Çiğli Training and Research Hospital, and tissue sections were prepared via microtomy. These sections were then subjected to Ki-67 immunohistochemical staining in the Histology and Pathology Laboratory of the university.

Immunohistochemical staining protocol

Paraffin-embedded tissue blocks from breast cancer patients were sectioned at 5 μm thickness using a microtome. Ki-67 immunostaining was performed using a primary antibody (Abcam, ab16667). Deparaffinization was achieved with xylene for 30 minutes, followed by hydration through graded alcohols. To block non-specific binding, sections were incubated in citrate buffer (pH 6.0) and 3% hydrogen peroxide, then treated with Ultra V Block. Subsequently, sections were incubated overnight with the Ki-67 primary antibody, followed by a 10-minute incubation with the secondary antibody. Signal development was achieved using streptavidin-peroxidase and diaminobenzidine (DAB). After chromogenic reaction, Mayer's hematoxylin was used as a counterstain. Coverslips were mounted with Entellan, and sections were visualized using a Zeiss Lab.A1 light microscope and photographed with the Zeiss AxioCam ERc 5s imaging system.

Preparation of validation dataset for performance evaluation

Following IHC staining, the most intensely labeled nuclear regions ("hot spots") were selected from each sample, and 40x magnification digital images were acquired. These 50 images were reserved exclusively for the validation phase of the study.

https://doi.org/10.5455/annalsmedres.2025.08.244

Following segmentation and classification, the Ki-67 index values predicted by the model were compared against manually calculated values obtained by an experienced pathologist from the same images. This comparison served to assess the predictive accuracy of the model.

The validation phase was designed to evaluate the clinical reliability, validity, and applicability of the proposed model. The results demonstrated the model's predictive performance and highlighted the potential of AI-based systems in histopathological assessment.

Statistical analysis

The deep learning model's performance was evaluated using both regression and classification metrics. Prediction accuracy was quantified using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Pearson Correlation Coefficient (r), alongside a scatter plot to visualize the relationship between predicted and actual values. For binary classification, a clinical Ki-67 threshold of 20% was applied. A confusion matrix was generated to calculate Accuracy, Sensitivity (Recall), Precision, Specificity, and F1 Score, assessing the model's capacity to differentiate between low and high proliferative index groups.

RESULTS

Model performance and overall statistical evaluation

The performance of the Ki-67 proliferation index prediction model was evaluated using five fundamental regression metrics. All statistical calculations were performed using Microsoft Excel, and the selected metrics were used to assess the model's predictive capacity from multiple dimensions.

A quantitative summary of model performance is presented in Table 1.

The mean absolute error (MAE) was calculated as 8.69%, indicating an average deviation of ±8.69 percentage points in the model's predictions. The mean squared error (MSE) was found to be 169.07, while the root mean square error (RMSE) was 13.00%, suggesting that predicted values deviated from actual values by an average of 13 points, representing the standard deviation of the error distribution.

The coefficient of determination (R^2), calculated to evaluate the explanatory power of the model, was 0.540. This value indicates that the model explains 54% of the total variance in the Ki-67 proliferation index. Furthermore, the Pearson correlation coefficient (r) between the predicted and actual values was calculated as 0.753, indicating a statistically significant and strong positive correlation between the two variables.

Overall, the developed model demonstrated consistent predictive performance, particularly in low-to-moderate Ki-67 value ranges, and was capable of establishing a strong linear relationship with approximately 75% correlation strength.

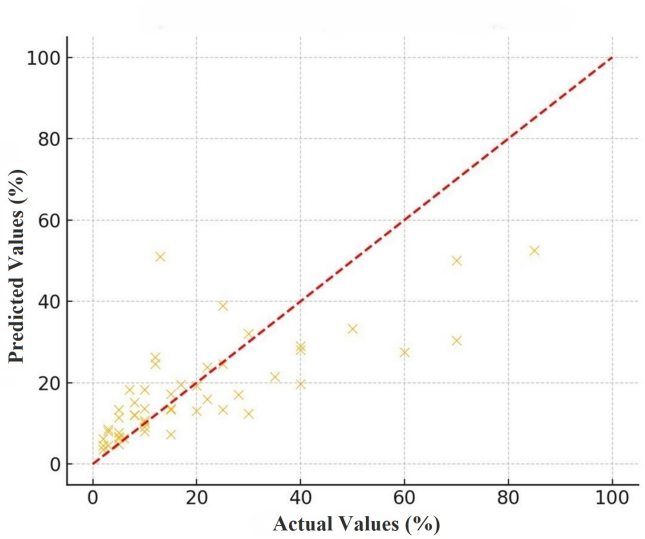

Figure 3. Visual analysis of model performance: relationship between actual Ki-67 index values and those predicted by the model.

Analysis of predicted and actual values

A visual analysis of the model's performance revealed the relationship between predicted and actual Ki-67 values. The results indicated that the model produced generally consistent predictions (Figure 3).

To further investigate model performance, the five most accurate and five most erroneous predictions were analyzed. The analysis of the top five predictions showed high accuracy, particularly in cases with low Ki-67 levels. Conversely, the most erroneous predictions revealed systematic issues in estimating high Ki-67 values (Figure 4).

Error analysis demonstrated that the model performed more accurately in the 10–30% prediction range, while substantial underestimation was observed for values ≥50% (Table 2).

Class-based error analysis

Given the clinical significance of Ki-67 values, a class-based performance analysis was performed. For this purpose, true Ki-67 values were dichotomized using a 20% threshold, separating the cases into two groups: low proliferation (<20%) and high proliferation (≥20%). This threshold aligns with the commonly used luminal subtype classification in the literature and is critical for tumor aggressiveness and prognosis evaluation. For each class, MAE and RMSE were calculated.

The error levels for low Ki-67 values (<20%) were significantly lower, suggesting that the model had better predictive accuracy for this group. In contrast, prediction errors increased in the high Ki-67 group (≥20%), with the MAE reaching 12.32 (Table 3).

Trend and subgroup-based evaluation of model predictions

The Ki-67 prediction model was evaluated using 50 paired observations, yielding mean values of 19.86 for real and 17.92

https://doi.org/10.5455/annalsmedres.2025.08.244

Table 1. General evaluation of model performance metrics.

Metric	Value	Description
Mean Absolute Error (MAE)	8.69%	Average prediction error
Mean Squared Error (MSE)	169.07	Weighting of large errors
Root Mean Squared Error (RMSE)	13.00%	Measure of standard deviation
R^2 Score	0.540	Proportion of explained variance
Correlation Coefficient (r)	0.753	Strength of linear relationship

Table 2. The actual values, model predictions, and absolute error percentages of the five best and five worst predictions of the artificial intelligence model regarding the Ki-67 index.

Best 5 Predictions	Actual Value (%)	Predicted Value (%)	Absolute Error (%)
a	10	10.14	+0.14
b	5	4.83	-0.17
c	6	6.35	+0.35
d	25	24.64	-0.36
e	10	10.76	+0.76
Worst 5 Predictions	**Actual Value (%)**	**Predicted Value (%)**	**Absolute Error (%)**
f	70	30.42	-39.58
g	13	51.04	+38.04
h	85	52.50	-32.50
i	60	27.58	-32.42
j	40	19.71	-20.29

Table 3. Class-based performance analysis of the model according to Ki-67 proliferation levels. The model demonstrated low error rates and high accuracy in the <20% class, whereas its performance declined in the ≥20% class.

Class	Sample Size	MAE	RMSE	Performance Evaluation
<20%	31	5.31	8.79	High performance
≥20%	19	12.32	14.19	Moderate performance

Table 4. Performance metrics used in the binary classification analysis of the Ki-67 proliferation index.

Metric	Value	Definition	Description
Accuracy	0.80	Proportion of correctly classified samples	Overall success rate
Error Rate	0.20	Proportion of misclassified samples	Overall error rate
Sensitivity (Recall)	0.80	Proportion of correctly detected <20 cases	Detection of low Ki-67
Precision	0.90	Proportion of correct <20 predictions	Reliability of positive predictions
Specificity	0.80	Proportion of correctly detected ≥20 cases	Detection of high Ki-67
F1 Score	0.84	Harmonic mean of sensitivity and precision	Balanced performance measure

for predicted data. The results demonstrated a strong correlation (r = 0.753) and no significant overall bias according to the paired t-test (mean difference = 1.94, p = 0.296), suggesting that the model's predictions were statistically consistent with the actual values.

Subgroup analysis revealed that the model's predictive accuracy varied across Ki-67 value ranges. In the low-proliferation group (<20%, n = 31), prediction errors were small (MAE ≈ 5.3, RMSE ≈ 8.8) and statistically aligned with real values (p ≈ 0.003), indicating that the model performed reliably in this clinically relevant range. For the moderate-high group (≥20%, n = 19), errors increased (MAE ≈ 12.3, RMSE ≈ 14.2) but still followed the correct statistical trend, showing no major deviation in direction. However, in high-proliferation subgroups (≥50%, n = 5 and ≥70%, n = 3), significant underestimation was observed, with higher error magnitudes (MAE ≈ 28.2–30.7; RMSE ≈ 29.5–31.7) and corresponding p-values of ≈0.003 and 0.03.

These findings indicate that while the model captured the correct overall trend between predicted and real Ki-67 values, its accuracy decreased in extreme ranges, primarily due to the limited number of high-proliferation samples. The reduced sample size in these subgroups likely contributed to increased error variance and lower statistical stability. Overall, the model performed well for low-to-moderate Ki-67 indices, demonstrating reliable predictive capacity, but tended to underestimate values in clinically critical high-proliferation cases.

https://doi.org/10.5455/annalsmedres.2025.08.244

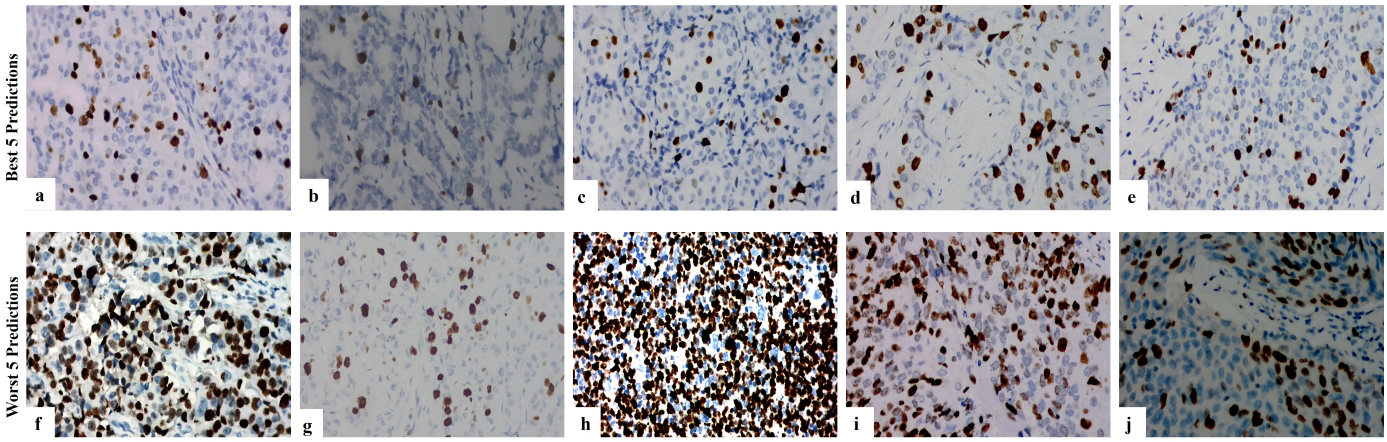

Figure 4. Example images of the five best (a, b, c, d, e) and five worst (f, g, h, i, j) predictions of the AI model for the Ki-67 index. (The actual values, model predictions, and absolute error percentages for these examples are presented in Table 2).

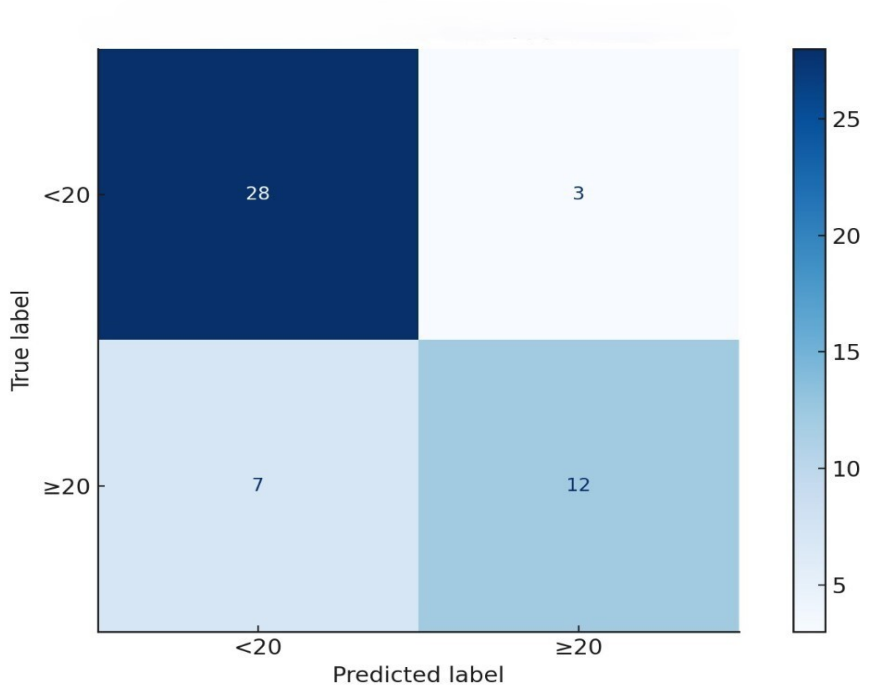

Figure 5. Confusion matrix of the binary classification analysis using the 20% threshold for Ki-67.

Confusion matrix and classification performance

To evaluate the model's performance in binary classification, confusion matrix–based performance metrics were calculated. Classification was performed using the 20% Ki-67 threshold to divide cases into two categories (Figure 5).

The model achieved an accuracy of 0.80, indicating that 80% of predictions were correct overall. Sensitivity was also 0.80, meaning that the model correctly identified 80% of low Ki-67 cases. Consequently, 20% of high Ki-67 cases were missed. The precision was calculated as 0.90, showing that among the cases predicted to have <20% Ki-67, 90% were indeed true positives. This high precision underscores the model's reliability in estimating low proliferation indices (Table 4).

DISCUSSION

The deep learning-based model developed in this study yielded promising results for the objective, reproducible, and rapid assessment of the Ki-67 proliferation index in breast cancer samples. The model demonstrated acceptable diagnostic sensitivity, particularly in low-proliferation cases (<20%), with a mean absolute error (MAE) of 5.31% and a high F1 score of 0.84.

Conventional pathological assessment of the Ki-67 index via manual counting is subject to interobserver variability and human error, as previously reported in the literature [32]. Furthermore, this variability particularly at clinically relevant thresholds such as 10–30% can significantly influence treat-

https://doi.org/10.5455/annalsmedres.2025.08.244

ment decisions [33]. In this context, the potential of digital image analysis and AI-assisted systems to improve objectivity and promote standardization in pathological evaluation is noteworthy [34].

Recent studies have demonstrated the efficacy of digital pathology and deep learning approaches for similar tasks. For instance, Cireşan et al. (2013) achieved high accuracy in mitotic figure detection using a CNN-based model [35]. Similarly, Veta et al. (2019), through the TUPAC16 challenge, showcased the effectiveness of machine learning algorithms in predicting breast tumor proliferation [34].

The sensitivity of the model in binary molecular subtype classification using the 20% Ki-67 threshold was found to be 80%, suggesting that some highly proliferative tumors may be missed. Given the clinical importance of accurately identifying high-risk cases, the incorporation of secondary control mechanisms such as pathologist review is recommended before integration into clinical workflows.

The observed high accuracy of the model in cases with low proliferation aligns with previous findings. Matsumoto et al. (2025) reported that their deep learning model demonstrated strong agreement (q = 0.961) with manual scoring in 494 breast cancer cases and showed low error rates in low Ki-67 samples. This supports our results, particularly the low MAE of 5.31% for <20% Ki-67 values [36]. Likewise, Dy et al. (2024) showed that AI-assisted scoring significantly reduced error rates (from 5.9% to 2.1%), improved interobserver agreement (ICC: 0.70 → 0.92), and decreased evaluation time in a study involving 90 international pathologists. Notably, AI applications were especially effective in standardizing assessments within the 5–30% Ki-67 range, thereby reducing variability [37].

These findings support the low error rates observed in our study for low Ki-67 values.

Additionally, error analysis of cases with Ki-67 positivity ≥20% revealed that the model's limitations were not solely due to class imbalance in the training dataset, but also to morphological and staining-related segmentation challenges. Analysis of the five most erroneous predictions revealed two major issues. First, in regions with 85–90% immunopositive nuclei, intense DAB staining and nuclear clustering impaired the model's ability to differentiate individual nuclei, leading to systematic underestimation of cell counts. This finding is consistent with prior reports indicating diminished segmentation performance in densely packed cell regions [38, 39]

Second, some immunonegative nuclei were too faintly stained (pale blue) to be recognized by the segmentation algorithm. This issue is commonly observed in threshold-based or traditional intensity-mapping segmentation approaches [31]. Moreover, our model intentionally excluded any preprocessing such as color filtering to preserve the visual appearance of slides as perceived by pathologists. While this approach ensured objectivity in comparative evaluation, studies suggest that integrating color space optimization may enhance classification accuracy [40].

Limitations

Although the model developed in this study demonstrated high accuracy in low-proliferation cases, a decline in performance was observed in high Ki-67 values (≥20%). This limitation is not solely attributable to class imbalance in the training dataset but is also associated with the absence of preprocessing steps (e.g., color normalization, contrast adjustment) and the challenges posed by intense DAB staining and nuclear clustering, which complicated segmentation. In addition, the very faint staining of some immunonegative nuclei restricted the model's discriminative capacity. These methodological and technical constraints may limit the generalizability of the findings and their applicability to clinical practice.

CONCLUSION

This study demonstrated that a deep learning-based model can provide rapid, objective, and reproducible assessments of the Ki-67 proliferation index in breast carcinoma samples. The model's high accuracy in low-proliferation cases supports its potential role as a clinical decision-support tool. However, segmentation errors observed in highly proliferative samples underscore the need for advanced image processing techniques to improve the model's sensitivity to morphological variability.

Future work will aim to integrate multi-channel color normalization, histogram equalization, and adaptive thresholding techniques into the segmentation pipeline. Additionally, data augmentation strategies will be employed to ensure a more balanced representation of complex and challenging images in the training dataset. These enhancements are expected to improve overall model performance and facilitate the broader and more reliable application of AI-assisted systems in histopathological diagnostics.

Acknowledgement: We would like to thank the Department of Pathology at İzmir Bakırçay University Çiğli Training and Research Hospital for providing the opportunity to conduct our study.

Ethics Committee Approval: This study was conducted in accordance with the ethical principles stated in the Declaration of Helsinki. The research protocol was reviewed and approved by the Non-Interventional Clinical Research Ethics Committee of İzmir Bakırçay University (Approval No: 1357; Date: 13.12.2023). All procedures were carried out in line with ethical standards, and patient confidentiality was strictly protected.

Informed Consent: Due to the retrospective nature of the study, the ethics committee waived the requirement for informed consent.

Peer-review: Externally peer-reviewed.

https://doi.org/10.5455/annalsmedres.2025.08.244

Conflict of Interest: The authors declare no conflict of interest.

Author Contributions: Concept/Design: SCK, AFD; Data Acquisition: HA; Data Analysis and Interpretation (including AI-based analyses): SCK, BK, HA, AG, AFD; Drafting Manuscript: SCK, BK, HA; Critical Revision of Manuscript: AFD, SCK, AG; Final Approval and Accountability: SCK, BK, HA, AG, AFD; Technical or Material Support: SCK, BK, HA, AG, AFD; Supervision: SCK, AFD.

Financial Disclosure: This study was supported by the İzmir Bakırçay University Scientific Research Projects Coordination Unit (Project No: BBAP.2023.014).

Artificial Intelligence Disclosure: The authors declare that an artificial intelligence–based tool (ChatGPT, version 5.2) was used solely for language editing to improve the clarity and readability of the manuscript. The AI tool did not contribute to the study design, data analysis, interpretation of results, or scientific conclusions. All content was critically reviewed, verified, and approved by the authors, who retain full responsibility for the manuscript.

REFERENCES

1. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Semin Cancer Biol.* 2021;72:214-225. doi: 10.1016/j.semcancer.2020.06.002.
2. Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: How special are they? *Mol Oncol.* 2010;4(3):192-208. doi: 10.1016/j.molonc.2010.04.004.
3. Tan PH, Ellis I, Allison K, et al. The 2019 World Health Organization classification of tumours of the breast. *Histopathology.* 2020;77(2):181-5. doi: 10.1111/his.14091.
4. Moghbel M, Ooi CY, Ismail N, Hau YW, Memari N. A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. Artificial Intelligence Review. 2020;53(3):1873-1918. doi: 10.1007/s10462-019-09721-8.
5. Moghbel M, Mashohor S. A review of computer assisted detection/diagnosis (CAD) in breast thermography for breast cancer detection. *Artificial Intelligence Review.* 2013;39(4):305-13. doi: 10.1007/s10462-011-9274-2.
6. Murtaza G, Shuib L, Wahab AWA, et al. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artificial Intelligence Review.* 2020;53(3):1655-720. doi: 10.1007/s10462-019-09716-5.
7. Domingues I, Pereira G, Martins P, Duarte H, Santos J, Abreu PH. Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET. *Artificial Intelligence Review.* 2020;53:4093-160. doi: 10.1007/s10462-019-09788-3.
8. Kozegar E, Soryani M, Behnam H, Salamati M, Tan T. Computer aided detection in automated 3-D breast ultrasound images: a survey. *Artificial Intelligence Review.* 2020;53(3):1919-41. doi: 10.1007/s10462-019-09722-7.
9. Saha M, Chakraborty C, Racoceanu D. Efficient deep learning model for mitosis detection using breast histopathology images. *Comput Med Imaging Graph.* 2018;64:29-40. doi: 10.1016/j.compmedimag.2017.12.001.
10. Cheng HD, Cai XP, Chen XW, Hu LM, Lou XL. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition.* 2003;36(12):2967-91. doi: 10.1016/S0031-3203(03)00192-4.
11. Cheng HD, Shi XJ, Min R, Hu LM, Cai XR, Du HN. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition.* 2006;39(4):646-68. doi: 10.1016/j.patcog.2005.07.006.
12. Suh YJ, Jung J, Cho BJ. Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning. *J Pers Med.* 2020;10(4):211. doi: 10.3390/jpm10040211.
13. Rindi G. Nomenclature and classification of neuroendocrine neoplasms of the digestive system. WHO classification of tumours of the digestive system. 2010:13-4.
14. Yamaguchi T, Fujimori T, Tomita S, et al. Clinical validation of the gastrointestinal NET grading system: Ki67 index criteria of the WHO 2010 classification is appropriate to predict metastasis or recurrence. *Diagn Pathol.* 2013;8:65. doi: 10.1186/1746-1596-8-65.
15. Tellez D, Balkenhol M, Otte-Höller I, et al. Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks. *IEEE Transactions on Medical Imaging.* 2018;37(9):2126-36. doi: 10.1109/TMI.2018.2820199.
16. Romo-Bucheli D, Janowczyk A, Gilmore H, Romero E, Madabhushi A. A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. *Cytometry A.* 2017;91(6):566-73. doi: 10.1002/cyto.a.23065.
17. Lu C, Romo-Bucheli D, Wang X, et al. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Lab Invest.* 2018;98(11):1438-48. doi: 10.1038/s41374-018-0095-7.
18. Kuş Silav Z. Meme Kanserinde Ki67 İndeks Ölçümlerinin Manuel ve Dijital Yöntemler Açısından Kıyaslanması. *İstanbul Gelişim Üniversitesi Sağlık Bilimleri Dergisi.* 2023.
19. Biswal A. Top 10 deep learning algorithms You should Know in 2021. https://www simplilearn com/tutorials/deep-learning-tutorial/deep-learning-algorithm, accessed. 2021;5:26.
20. Datwani S, Khan H, Niazi MKK, Parwani AV, Li Z. Artificial intelligence in breast pathology: Overview and recent updates. *Hum Pathol.* 2025;162:105819. doi: 10.1016/j.humpath.2025.105819.
21. Jiang B, Bao L, He S, Chen X, Jin Z, Ye Y. Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis. *Breast Cancer Res.* 2024;26(1):137. doi: 10.1186/s13058-024-01895-6.
22. Soliman A, Li Z, Parwani AV. Artificial intelligence's impact on breast cancer pathology: a literature review. *Diagn Pathol.* 2024;19(1):38. doi: 10.1186/s13000-024-01453-w.
23. Valieris R, Martins L, Defelicibus A, et al. Weakly-supervised deep learning models enable HER2-low prediction from H &E stained slides. *Breast Cancer Res.* 2024;26(1):124. doi: 10.1186/s13058-024-01863-0.
24. Chaurasia AK, Toohey PW, Bennett MT, Harris HC, Hewitt AW. Automated quantification of Ki-67 expression in breast cancer from H&E-stained slides using a transformer-based regression model. *Breast Cancer Res.* 2025;27(1):206. doi: 10.1186/s13058-025-02149-9.
25. Goyal M, Marotti JD, Workman AA, et al. A multi-model approach integrating whole-slide imaging and clinicopathologic features to predict breast cancer recurrence risk. *NPJ Breast Cancer.* 2024;10(1):93. doi: 10.1038/s41523-024-00700-z.
26. Mehta N, Devarakonda MV. Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey? *J Allergy Clin Immunol.* 2018;141(6):2019-21.e1. doi: 10.1016/j.jaci.2018.02.025.
27. Li S, Deng YQ, Zhu ZL, Hua HL, Tao ZZ. A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging. *Diagnostics (Basel).* 2021;11(9):1523. doi: 10.3390/diagnostics11091523.
28. Castiglioni I, Rundo L, Codari M, et al. AI applications to medical images: From machine learning to deep learning. *Phys Med.* 2021;83:9-24. doi: 10.1016/j.ejmp.2021.02.006.
29. Şensu S, Erdoğan N, Gürbüz YS. Patolojide dijital çağ ve yapay Zekâ: temel bilgiler. Türkiye Klinikleri Tip Bilimleri Dergisi. 2020;40(1):104-12. doi: 10.5336/medsci.2019-72835.

https://doi.org/10.5455/annalsmedres.2025.08.244

30. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88. doi: 10.1016/j.media.2017.07.005.
31. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods.* 2021;18(1):100-6. doi: 10.1038/s41592-020-01018-x.
32. Polley M-YC, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst.* 2013;105(24):1897-906. doi: 10.1093/jnci/djt306.
33. Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst.* 2011;103(22):1656-64. doi: 10.1093/jnci/djr393.
34. Veta M, Heng YJ, Stathonikos N, et al. Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. *Med Image Anal.* 2019;54:111-21. doi: 10.1016/j.media.2019.05.008.
35. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J, editors. Mitosis detection in breast cancer histology images with deep neural networks. International conference on medical image computing and computer-assisted intervention; 2013: Springer.
36. Matsumoto H, Miyata R, Tsuruta Y, et al. Ki-67 evaluation using deep-learning model-assisted digital image analysis in breast cancer. *Histopathology.* 2025;86(3):460-71. doi: 10.1111/his.15356.
37. Dy A, Nguyen NNJ, Meyer J, et al. AI improves accuracy, agreement and efficiency of pathologists for Ki67 assessments in breast cancer. *Sci Rep.* 2024;14(1):1283. doi: 10.1038/s41598-024-51723-2.
38. Graham S, Vu QD, Raza SEA, et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal.* 2019;58:101563. doi: 10.1016/j.media.2019.101563.
39. Trullo R, Bui Q-A, Tang Q, Olfati-Saber R, editors. Image translation based nuclei segmentation for immunohistochemistry images. MICCAI Workshop on Deep Generative Models. 2022. doi: 10.48550/arXiv.2208.06202.
40. Gonzalez RC. Digital image processing: Pearson education india; 2009.

https://doi.org/10.5455/annalsmedres.2025.08.244