Chirurgia (2026) 121: 239-248
No. 3, May - June

http://dx.doi.org/10.21614/chirurgia.3254

AI-Powered Predictions of Breast Ductal in Situ Carcinoma Morphology and Surgical Outcomes

David Andraș[1,2], Radu Alexandru Ilieș[3]*, Victor Eșanu[1,2] and George Călin Dindelegan[1,2]

[1]1st Surgical Clinic, Department of General Surgery, Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
[2]Emergency County Hospital Cluj, First Surgical Unit, Cluj-Napoca, Romania
[3]Faculty of Medicine, Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

***Corresponding author:**
Radu Alexandru Ilieș, MD
Iuliu Hațieganu University of Medicine and Pharmacy Cluj, Faculty of Medicine
Victor Babeș Street, No 8
E-mail: ilies.radu.alexandru@elearn.umfcluj.ro

Abbreviations:
AI: Artificial Intelligence;
DCIS: Ductal Carcinoma In Situ;
NCCN: National Comprehensive Cancer Network;
ESMO: European Society for Medical Oncology;
NST: No Special Type;
NAT: Neoadjuvant Treatment;
SLNB: Sentinel Lymph Node Biopsy;
R0: Complete resection with negative margins;
PACS: Picture Archiving and Communication System;
DICOM: Digital Imaging and Communications in Medicine;
IDC: Invasive Ductal Carcinoma;
IMC: Invasive Mixed Carcinoma;
NOS: Not Otherwise Specified;
CI: Confidence Interval;
SD: Standard Deviation;
ROC: Receiver Operating Characteristic;
AUC: Area Under the Curve;
ER: Estrogen Receptor;
HER2: Human Epidermal Growth Factor Receptor 2;
ANOVA: Analysis of Variance;
LASSO: Least Absolute Shrinkage and Selection Operator;
H&E: Hematoxylin and Eosin (histological stain).

Received: 23.11.2025
Accepted: 30.01.2026

Rezumat

Predicția cu ajutorul inteligenței artificiale a caracteristicilor morfologice și a rezultatelor chirurgicale în carcinomul ductal in situ mamar

Introducere: Inteligența artificială (IA) este din ce în ce mai integrată în imagistica oncologică, însă capacitatea sa de a prezice caracteristici histopatologice detaliate pe baza mamografiei standard rămâne insuficient studiată. Acest studiu a avut ca scop evaluarea performanței unui model de limbaj extins (ChatGPT-4, Open AI, mai 2025) în prezicerea gradului nuclear, a subtipului arhitectural, a comedonecrozei şi a invaziei stromale din mamografia de control a piesei.
Materiale şi Metode: Am efectuat un studiu retrospectiv şi metodologic pe 29 de paciente cu CDIS confirmat histologic sau carcinom invaziv cu componente CDIS. Protocolul nostru clinic se bazează pe ghidurile NCCN/ESMO privind tratamentul carcinomului ductal in situ. Preoperator, toate pacientele cu boală confirmată au beneficiat de localizarea leziunii mamare cu ghidaj metalic. Pacientele au fost apoi supuse exciziei chirurgicale (chirurgie conservativă mamară urmată de mamografie de control a piesei pentru a confirma excizia macroscopică completă). Pentru fiecare dintre aceste cazuri, imaginile mamografice au fost analizate utilizând un model de IA conceput pentru a extrage şi procesa caracteristicile radiomice. Rapoartele modelului de IA au fost comparate cu rapoartele histopatologice, care au servit drept standard de aur. Performanța diagnostică a fost evaluată pentru patru parametri: gradul nuclear CDIS, subtipul arhitectural, comedonecroza şi invazia stromală. Au fost calculate acuratețea, sensibilitatea, specificitatea, precizia şi scorurile F1.
Rezultate: Dimensiunea leziunilor (măsurată mamografic) a variat de la 1,2 la 10,0 mm (medie ± SD: 4,46 ± 2,25 mm). Diagnosticele histopatologice au inclus CDIS pur (n = 17), carcinom NST invaziv cu CDIS (n = 10) şi histologii mixte (n = 2). Modelul de IA a atins o precizie de 65,5% pentru detectarea comedonecrozei (sensibilitate 75,0%, specificitate 53,8%) şi o precizie de 72,4% pentru detectarea invaziei stromale (specificitate 94,1%, sensibilitate 41,7%). Clasificarea gradului nuclear a corespuns histopatologiei în 20,7% din cazuri, în timp ce clasificarea subtipurilor arhitecturale a atins o concordanță de 17,2%. Predicțiile multi-clasă au arătat scoruri F1 scăzute pentru majoritatea categoriilor.
Concluzie: Deşi modelul de IA a demonstrat o utilitate acceptabilă pentru detectarea comedonecrozei şi excluderea invaziei stromale, s-a confruntat cu o serie de dificultăți în ceea ce priveşte gradarea nucleară şi clasificarea subtipurilor

arhitecturale. Deşi limitat de dimensiunea mică a eşantionului şi de imagistica 2D, acest studiu metodologic oferă o perspectivă asupra viitoarelor abordări bazate pe IA şi radiomică în caracterizarea tumorilor mamare.

Cuvinte cheie: radiomică; inteligență artificială; carcinom ductal in situ; mamografie; histopatologie

Abstract

Background/Objectives: Artificial intelligence (AI) is increasingly integrated into oncological imaging, but its ability to predict detailed histopathological features from standard mammography remains understudied in ductal carcinoma in situ (DCIS). This study aimed to evaluate the performance of a large language model (ChatGPT-4, Open AI, May 2025) in predicting nuclear grade, architectural subtype, comedo necrosis, and stromal invasion from specimen mammography.

Materials and Methods: We conducted a retrospective and methodological study of 29 patients with histologically confirmed DCIS or invasive carcinoma with DCIS components. Our clinical protocol is based on NCCN/ESMO guidelines of treatment ductal carcinoma in situ. Preoperatively, all patients with confirmed disease benefited from wire guide localization of the breast lesion. Patients were then submitted to surgical excision (lumpectomy) and surgical specimen mammography to confirm the complete macroscopic excision. For each of these cases, mammographic specimens were analyzed using an AI model designed to extract and process radiomic features. AI-model reports were compared with histopathological reports which served as the gold standard. Diagnostic performance was evaluated for four parameters: DCIS nuclear grade, architectural subtype, comedo necrosis, and stromal invasion. Accuracy, sensitivity, specificity, precision, and F1 scores were computed.

Results: The size of mammographic lesions ranged from 1.2 to 10.0 mm (mean ± SD: 4.46 ± 2.25 mm). Histopathological diagnoses included pure DCIS (n = 17), invasive NST carcinoma with DCIS (n = 10), and mixed histologies (n = 2). The AI model achieved 65.5% accuracy for detecting comedo necrosis (sensitivity 75.0%, specificity 53.8%) and 72.4% accuracy for detecting stromal invasion (specificity 94.1%, sensitivity 41.7%). Nuclear grade classification matched histopathology in 20.7% of cases, while architectural subtype classification achieved 17.2% agreement. Multiclass predictions showed low F1 scores for most categories.

Conclusion: Although the AI model demonstrated acceptable utility for detection of comedo necrosis and excluding stromal invasion, it faced several difficulties regarding nuclear grading and architectural subtype classification. Although limited by the small sample size and 2D imaging, this methodological study provides an insight for future AI and radiomics approaches in breast tumor characterization.

Keywords: radiomics, artificial intelligence, ductal carcinoma in situ, mammography, histopathology

Introduction

Recently, artificial intelligence (AI) has emerged as a transformative feature across multiple fields, including healthcare, particularly directed towards large language models such as ChatGPT, which have become widely accessible to both medical professionals and the public, including patients (1,2). The strength of AI lies in its capability to execute complex cognitive tasks (problem-solving, decision-making, and even pattern recognition or predictions) using sophisticated algorithmic methods. By copying aspects of human reasoning and learning, AI systems can adapt to diverse data types and decision contexts, making them highly versatile tools in data-rich disciplines such as medicine (3-6).

Within the broader AI spectrum, machine learning and deep learning stand out for their capacity to autonomously identify patterns within large sets of data and continuously improve their performance through iterative training. Their application in radiomics (a field concentrating on the extraction of high-dimensional, quantitative features from medical images) has opened new directions for non-invasive characterization of the diseases. Radiomics allows for the quantification of subtle imaging patterns that might reflect underlying pathology or pathophysiology, often imperceptible to the human eye. When integrated with clinical and pathological data, these approaches have the potential to enhance the precision of diagnosis, support risk stratification, and improve individualized management in breast cancer evaluating response to therapy. This fusion of AI, radiomics, and imaging science holds significant

potential to advance various medical specialties, with radiology and diagnostics in the first position (7-9).

In a previous study, as part of the same institutional research project, we evaluated the performance of a large language model (ChatGPT-4, Open AI, May 2025 release) in analyzing intraoperative specimen mammograms for the evaluation of the status of surgical margin status in patients with early-stage breast cancer (9). Our model was able to localize the lesion, delineate resection margins, and estimate the minimal margin distance. Interestingly, it achieved high specificity for negative margins. When stratified by tumor invasiveness, the AI achieved an accuracy of 73.4% in invasive cancers (sensitivity 60.0%, specificity 89.8%) and 79.3% in non-invasive cancers (sensitivity 75.0%, specificity 80.0%), with no statistically significant differences between groups. In the neoadjuvant treatment (NAT) analysis, accuracy was 82.8% in the NAT group (sensitivity 42.9%, specificity 86.7%) and 85.4% in the non-NAT group (sensitivity 100%, specificity 84.6%). Although the NAT group showed lower sensitivity, differences were not statistically significant. Taken together, these patterns imply that model performance is more favorable in settings resembling DCIS biology and treatment pathways (non-invasive phenotype, typically no NAT). Because DCIS aligns with the non-invasive and predominantly non-NAT features (in which our model performed better), we hypothesize that a DCIS-only cohort will yield equal or improved accuracy and sensitivity for detecting positive margins compared with the mixed cohort. Consequently, we believe that this AI model might be capable of detecting certain morphological patterns which were associated with DCIS.

Ductal carcinoma in situ (DCIS) represents a non-invasive form of breast cancer that is characterized by the proliferation of malignant epithelial cells originating from the ductal-lobular system, without any extension beyond the basal membrane. Although it does not display stromal invasion, DCIS is clinically significant (because of its potential to progress to invasive carcinoma if left untreated). The disease presents with a wide variety of histopathological features, including variations in nuclear grade, architectural growth pattern, and the presence (or absence) of comedo necrosis (all of which have their own prognostic value and potential to influence treatment decisions) (10-12). Accurate preoperative characterization of these features can influence the selection of the appropriate surgical management, determine the need for adjuvant therapy, and estimate the risk of recurrence (13-15).

High-grade DCIS, particularly when associated with comedo necrosis, carries a substantial risk of harboring invasive carcinoma, even when preoperative biopsy shows only in situ disease. In many cases, final histopathology reveals incidental stromal or microinvasion, which may influence surgical management and prognosis (15-18).

While our earlier investigation addressed assessment of the resection margins, this study particularly focuses on a different predictive task: evaluating if a radiomics component applied to mammogram of the operative specimen can predict key histopathological features of DCIS: grade, architectural subtype, presence of comedo necrosis, and presence of stromal invasion, using histo-pathology as the gold standard.

By limiting the analysis only to patients with DCIS component, either pure or associated with invasion, we aimed to explore whether AI-based radiomic signatures can accurately provide histo-pathological characterization in a subgroup where our previous results indicated an encouraging imaging-pathology correlation. Our goal is to generate complementary evidence within the same clinical research framework, with full methodological and ethical transparency to ensure compliance with publication ethics and to avoid redundancy with the earlier study.

Materials and Methods

Study Design and Patient Selection

This research was conducted as part of the institutional research project "Improving Postoperative Outcomes in Breast Surgery through the Use of Artificial Intelligence and Augmented Reality Programs", carried out at the First Surgical Department, Cluj County Emergency Clinical Hospital, Cluj-Napoca, Romania, in collaboration with the Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania. Designed retrospectively, this study was conducted in a surgical unit specialized in breast cancer, with the approval of the institutional review board and ethics committee (Approval No. 3387/31 January 2025). The present work was designed as a companion to our previously published study (9), which evaluated the performance of an artificial intelligence model in predicting surgical margin status using intraoperative specimen mammography (in patients with early-stage breast cancer). While that earlier analysis focused on margin assessment, the current study targets radiomic evaluation of DCIS and associated histopathological features. Both studies share the same research protocol, ethical considerations, and methodological framework, providing complementary perspectives within the same clinical research project.

We explicitly acknowledge the relationship between

the two studies to guarantee transparency and to avoid any issue of duplicate publication. Even though the present work shares certain conceptual and procedural elements with the earlier study, it addresses a distinct scientific topic (AI-based prediction of DCIS histopathological features from mammographic specimen images) rather than assessment of surgical margins. Neither the data, nor the figures have been duplicated from the earlier publication.

Data processing and analysis were conducted in accordance with the institutional and international standards for patient confidentiality, with all images fully anonymized prior to evaluation with the AI model, ChatGPT.

We extracted from our database patients with histopathological confirmation of DCIS component, who underwent breast lumpectomy at the Cluj County Emergency Hospital, Cluj-Napoca, Romania, between 2020 and 2025. All included patients had a preoperative Tru-Cut core needle biopsy confirming the presence of DCIS. The study population was selected based on this biopsy result, regardless of whether final surgical histopathology subsequently revealed incidental foci of stromal invasion or microinvasion. Such invasive components, when present, were considered incidental findings and did not alter the inclusion criteria. Only patients with complete clinical, imaging, surgical and histopathological data were included in our analysis.

Surgical Procedure and Imaging Protocol

Preoperative localization of the tumor was realized in each case using wire and/or clip marking. All surgical interventions were realized in concordance with ESMO and NCCN guidelines for the management of ductal carcinoma in situ. Following preoperative localization, patients underwent breast-conserving surgery (lumpectomy) with the intention of achieving negative surgical margins. In selected cases with high-risk histopathological features on preoperative biopsy (e.g., nuclear grade 3, comedo necrosis), sentinel lymph node biopsy (SLNB) was performed at the same time as the procedure. Immediately after excision, the specimen underwent mammographic assessment to verify macroscopic completeness of excision. Resection margins were considered clear (R0) when the minimal distance from tumor to ink was >2 mm, in accordance with current recommendations. Postoperatively, all the cases were discussed in a specialized, multidisciplinary tumor board, and adjuvant treatment (composed of radio-therapy and/or endocrine chemoprophylaxis) was administered according to individual risk profiles. Immediately after surgical excision, mammographic imaging of the lumpectomy specimens was achieved to verify complete removal of the tumor and to enable the breast radiologist to assess surgical margins intraoperatively, based on the radiographic findings.

For this study, we employed the ChatGPT-4 model (OpenAI, May 2025 release), accessed via a GPT-4 Plus subscription in a secure setting, with a stable internet connection. Image interpretation was carried out directly within the model interface, without any external software, and in a zero-shot configuration (no prior fine-tuning). Mammo-graphic images of the resected specimens were uploaded one at a time and evaluated using a predefined text prompt (which is available in *Appendix A*). The prompt incorporated essential metadata, including breast laterality, the tumor localization approach (e.g., wire or clip), and relevant intraoperative details. Based on this information and the image alone, ChatGPT was asked to predict the following: the nuclear grade of the tumor (low, intermediate, high, mixed or undefined), the architectural type of the lesion solid, comedo, cribriform, mixed (and specify which), unspecified/inconclusive. the presence of comedo necrosis (yes/no) and the presence of stromal invasion (yes/no).

Mammograms of the surgical specimens were retrospectively extracted from the hospital's Picture Archiving and Communication System (PACS), using DICOM files (between 2020 and 2025). All images had been obtained intraoperatively, within 10 to15 minutes following tumor excision, on a high-resolution digital mammography unit (GE Senographe Pristina® or equivalent). Standard craniocaudal and lateral projections were captured under mild compression, making use of a dedicated specimen radiography platform.

The extracted DICOM datasets contained full-resolution pixel information, with a native spatial resolution of approximately 70-100 microns per pixel. To obtain compatibility with the AI analysis environment, images were converted to JPEG or PNG format. In accordance with institutional protocol, orientation markers (radio-paque needles) were present in all cases. Image acquisition had been supervised by a breast radiologist to ensure optimal visualization.

For each patient, one representative image (most commonly the craniocaudal projection) was selected based on the best visualization of lesion. No manual annotations or segmentations were applied before the evaluation of the AI model.

Histopathological Evaluation and Statistical Analysis

All excised lesions underwent histopathological examination, performed by a specialized pathologist,

which served as the reference standard for both diagnosis and assessment of histopathological features. The pathological spectrum included various subtypes of breast cancer: DCIS in 18 cases, invasive ductal carcinoma of no special type (IDC NST) associated with DCIS in 9 cases, invasive mixed carcinoma (IMC) associated with DCIS in 1 case, respectively invasive mixed carcinoma, not otherwise specified (IMC NOS) associated with DCIS in 1 case.

All these histopathological findings were used for the validation of the AI-generated classifications and determine their clinical applicability.

To perform the comparison, the dataset contained paired values for four primary histo-pathological outcomes and their corresponding AI-predicted classifications: (1) DCIS grade (ordinal variable: low, intermediate, high), (2) architectural subtype (nominal variable with multiple categories, including cribriform, comedo, papillary, solid, mixed, unspecified/inconclusive), (3) comedo necrosis (binary variable: present/absent), and (4) invasion (binary variable: present/absent). The concordance between AI predictions and histo-pathological findings was evaluated using accuracy, sensitivity, specificity, precision (positive predictive value), and F1-score. For binary outcomes (comedo necrosis and stromal invasion), the corresponding confusion matrices were constructed, and 95% confidence intervals (CIs) for accuracy, sensitivity, specificity, and precision were calculated using the Clopper-Pearson exact method. Agreement for binary variables was further assessed using McNemar's test for paired proportions. For multiclass outcomes (DCIS nuclear grade and architectural subtype), per-class, macro-averaged, and weighted-average F1-scores were computed.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 29.0 and Microsoft Excel for Microsoft 365 (Version 2407).

Results

A total number of 29 patients with DCIS met the inclusion criteria. On histopathological examination (*Table 1*), pure DCIS was identified in 17 cases (58,62%). Additionally, 10 cases (34.48%) consisted of invasive carcinoma of no special type (NST) with an associated DCIS component, and 2 cases (6.9%) represented other mixed histologies, including combinations with cribriform-micropapillary or lobular elements.

Lesion size (measured on mammography) ranged between 1.2 and 10.0 mm (*Fig. 1*). The mean size was 4.46 mm (SD: 2.25 mm), the median was 4.1 mm, and the interquartile range was 2.7–6.3 mm. Lesions ≤ 3.0 mm were recorded in 8 cases (27.6%), lesions between 3.1 and 5.0 mm in 11 cases (37.9%), and lesions over 5.0 mm in 10 cases (34.5%).

Histopathological grading of the DCIS component (*Table 2*) revealed 4 low-grade cases (13.8%), 13 intermediate-grade cases (44.8%), 8 high-grade cases (27.6%), and 4 cases (13.8%) with mixed or undefined grades. Furthermore, the AI model assigned low grade DCIS in 2 cases (6.9%), intermediate in 10 cases (34.5%), high in 14 cases (48.3%), and mixed grades in 3 cases (10.3%). Exact agreement between the AI model and histopathology was observed in 6 cases (20.7%). Class-specific F1-scores were as follows: intermediate grade 0.52, low grade 0.00, high grade 0.00, mixed grade 0.00. The macro-average F1-score

Table 1. Histopathological assessment of the surgical specimens

Histopathological diagnosis	Count	Percentage (%)
Pure DCIS	17	58,62
DCIS + Invasive carcinoma NST (incidental)	10	34.48
Mixed histology (other types)	2	6.9

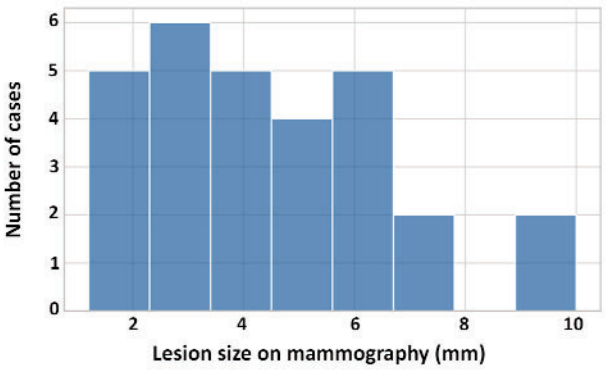

Figure 1. Distribution of lesion sizes on mammography. A histogram illustrating the distribution of lesion sizes (in mm) measured on mammographic images of the surgical specimens.

Table 2. DCIS nuclear grade distribution. Comparison between histopathology and AI-prediction.

Nuclear grade	Histopathology (n)	AI prediction (n)
Low grade	4	2
Intermediate grade	13	10
High grade	8	14
Mixed/undefined grade	4	3
Total	29	29

for DCIS grading was 0.05, and the weighted average was 0.23.

The architectural subtype distribution (*Table 3*) based on histopathological examination was: solid pattern in 4 cases (13.8%), comedo in 4 cases (13.8%), cribriform in 1 case (3.4%), solid combined with cribriform in 4 cases (13.8%), other mixed patterns in 13 cases (44.8%), and unspecified or inconclusive in 3 cases (10.3%). The AI model predicted solid pattern in 4 cases (13.8%), comedo in 8 cases (27.6%), cribriform in 5 cases (17.2%), cribriform plus papillary in 1 case (3.4%), comedo plus cribriform in 1 case (3.4%), solid plus cribriform in 1 case (3.4%), and undetermined in 2 cases (6.9%). Agreement between AI and histopathology was found in 5 cases (17.2%). Class-specific F1-scores: cribriform/papillary 1.00, solid 0.44, comedo 0.25, all other categories 0.00. The macro-average F1-score for architectural subtype classification was 0.10, and the weighted average was 0.13.

Comedo necrosis was present in 16 cases (55.2%) and absent in 13 cases (44.8%), according to histopathology (*Fig. 2 A*). The AI model predicted comedo necrosis as present in 18 cases (62.1%) and absent in 11 cases (37.9%). This yielded 12 true positives, 4 false negatives, 6 false positives, and 7 true negatives. Performance metrics for comedo necrosis detection: accuracy 65.5% (95% CI: 0.46–0.82), sensitivity 75.0% (95% CI: 0.48–0.93), specificity 53.8% (95% CI: 0.25–0.81), precision 66.7% (95% CI: 0.41–0.87), and F1-score 0.71. McNemar's test did not reveal a statistically significant difference between AI and histopathology (p = 0.754).

Incidental stromal invasion was present in 12 cases (41.4%) and absent in 17 cases (58.6%) on histopathology (*Fig. 2 B*). The AI model predicted invasion in 6 cases (20.7%) and absence in 23 cases (79.3%). This resulted in 5 true positives, 7 false negatives, 1 false positive, and 16 true negatives. Corresponding performance metrics: accuracy 72.4% (95% CI: 0.53–0.87), sensitivity 41.7% (95% CI: 0.15–0.72), specificity 94.1% (95% CI: 0.71–1.00), precision 83.3% (95% CI: 0.36–1.00), and F1-score 0.56. McNemar's test indicated no statistically significant discordance (p = 0.070).

Table 3. Architectural subtype distribution. Comparison between Histopathology and AI-prediction.

Architectural subtype	Histopathology (n)	AI prediction (n)
Solid	4	4
Comedo	4	8
Cribriform	1	5
Solid + Cribriform	4	1
Cribriform + Papillary	0	1
Comedo + Cribriform	0	1
Other mixed patterns	13	0
Unspecified/inconclusive	3	0
Undetermined	0	2
Total	29	29

Discussion

Interpretation of the Main Results

The AI model showed moderate performance in detecting comedo necrosis, with a sensitivity of 75.0% and an F1-score of 0.71. These results point that the AI may be able to identify necrosis in most true-positive cases. Clinically, this is important when considering the prognostic significance of comedo necrosis in DCIS. Conversely, the specificity was relatively low (53.8%),

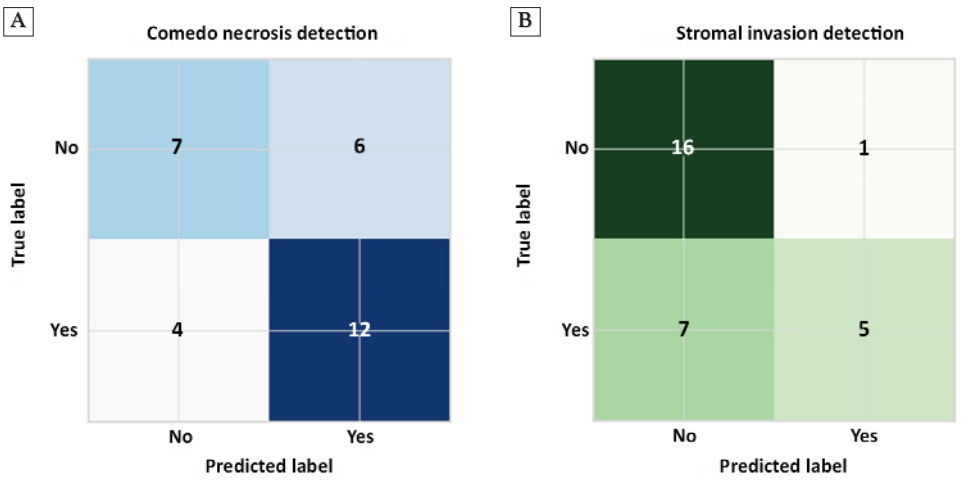

Figure 2. Confusion matrices of: **(A)** Comedo necrosis detection: accuracy 65.5%, sensitivity 75.0%, specificity 53.8%, precision 66.7%; **(B)** Detection of stromal invasion: accuracy 72.4%, sensitivity 41.7%, specificity 94.1%, precision 83.3%. Histopathological diagnosis (reference standard) was compared with AI predictions.

being responsible for a considerable number of false positives. This may be the cause of an overestimation of lesion aggressiveness. The overall accuracy of 65.5% slightly supports the use of such models as supportive diagnostic aids, but it highlights the need for further refinement (probably achieved via pre-trained models) to improve specificity.

Regarding the stromal invasion assessment, the AI model showed high specificity (94.1%), indicating a strong capability of detecting non-invasive cases. This feature has several applications, such as guiding decisions about the extent of surgery (and verifying surgical margins) and the potential need for additional treatment. However, its sensitivity was low (41.7%), with more than a half of the truly invasive cases being wrongly diagnosed. Even though the precision of the model was acceptable (83.3%), the overall F1-score of 0.56 shows a limited reliability in the detection of cases with stromal invasion. Thus, the model may be too conservative in predicting invasion and risks, likely to underestimate the severity of the disease in a category of patients.

The AI model's ability to precisely determine the nuclear grade of DCIS was limited. Only 6 out of 29 cases matched the histopathological grading, leading to an overall accuracy of 20.7%. Analysis of each category revealed that its performance was acceptable only for intermediate grade lesions, with an F1-score of 0.52. Our model completely failed to predict low-grade, high-grade, and mixed-grade cases, in each of these cases scoring an F1 of 0.00. The weighted average F1-score of 0.23 supports the fact that the AI program faces difficulties with this multiclass classification task.

The AI model also showed a poor performance in identifying the architectural subtype of DCIS, with an accuracy of only 17.2%. Even if its performance was acceptable for some isolated classes such as cribriform (F1-score of 1.00), most patterns (particularly mixed or less distinct forms) were not predicted with accuracy. The macro-average F1-score of 0.10 and the weighted F1-score of 0.13 highlight the fact that the model lacks robustness across the wide morphological spectrum of DCIS. This further supports its difficulty in detecting architectural heterogeneity based on 2D specimen mammograms, especially in the cases where histological subtypes overlap or coexist within the same lesion.

Alignment with the Existing Literature

To begin with, our observations on incidental stromal invasion in non-invasive disease align with published evidence that DCIS, particularly high-grade lesions with comedo necrosis, is frequently associated with microinvasion on final histo-pathology despite a purely in situ diagnosis at biopsy (15,16). Such microinvasive foci have been linked to more aggressive clinicopathological features and worse prognostic outcomes compared with pure DCIS (16,17). Moreover, long-term follow-up data indicate that microinvasive breast cancer can have distinct treatment responses and survival patterns, highlighting the importance of tailoring adjuvant therapies in this subgroup (16-18).

Together, these findings underscore the clinical relevance of preoperative identification of occult invasion in DCIS and support the exploration of AI-based tools for margin status prediction in this context.

Yu et al. conducted a multicenter retrospective study that involved 1,088 patients to create and validate an MRI radiomics approach for predicting axillary lymph node status in the preoperative period, in early-stage invasive breast cancer. Radiomic features from both the lymph nodes and the primary tumor were extracted using a random forest algorithm and combined into a signature of the tumor with a support vector machine classifier. This model achieved areas under the ROC curve of 0.88 in the training cohort, 0.87 in an external validation cohort, and 0.87 in a prospective-retrospective validation cohort. When radiomics was combined with clinicopathologic and molecular subtype information into a multiomic signature, predictive performance further improved, achieving AUC values up to 0.93. In patients who received neoadjuvant treatment, a few changes were noticed in certain radiomic features, particularly gray-level dependence matrix measures, which were linked to alterations inside the tumor microenvironment such as immune cell composition, expression of long non-coding RNAs, and DNA methylation patterns (found via blood sampling) (15). The findings indicate that MRI radiomics, when combined with clinical and molecular data, is able to provide a reliable and non-invasive method for evaluating lymph node metastasis preoperatively and might offer insight into underlying tumor biology.

Mao et al. performed a multicenter retrospective study including 304 patients with ER-positive, lymph node-negative invasive breast cancer to develop a mammography-based radiomics model for predicting risk of recurrence as defined by Oncotype DX. Radiomic features were extracted using preoperative mammograms, reduced through correlation analysis, ANOVA, and LASSO selection, and combined into a radiomics signature. A multivariate logistic regression model which integrated the radiomics signature with clinical factors (tumor grade and HER2 status)

achieved a strong predictive performance, with AUCs equal to 0.92 in the training set, 0.88 in the internal test set, respectively 0.84 in the external test set. Decision curve analysis indicated a greater clinical benefit of the model over "treat all" or "treat none" approaches (across a wide variety of threshold probabilities) (19). Hence, the combination of mammography radiomics with clinical variables was found to provide a non-invasive, preoperative tool for estimating recurrence risk in such a patient subgroup.

Wu et al. retrospectively analyzed data extracted from 611 breast cancer patients across multiple hospitals to develop a multimodality deep learning radiomics nomogram (DLRN) for preoperative prediction of malignancy. The model integrated clinical information with ultrasound brightness mode (B-mode) and color Doppler flow imaging characteristics. Patients were divided into a primary cohort, a validation cohort, and two external test cohorts. The DLRN achieved high discriminatory performance, with AUCs of 0.983, 0.972, 0.897, respectively 0.993 in the cohorts, consistently outperforming the three unimodal models. Decision curve analysis indicated a strong potential for clinical benefit (21). The integration of deep learning radiomics into multimodality ultrasound might be able to enhance preoperative malignancy assessment and to support computer-aided diagnosis in breast cancer.

Other applications of AI-based radiomics in the field of breast cancer extend beyond DCIS and include features like prognostic prediction, molecular subtyping, and the assessment of intratumoral heterogeneity. For instance, a multicentric study by Yu et al. described an interpretable deep-learning radiomics model relying on MRI for predicting recurrence-free survival in non-metastatic, invasive breast cancer, achieving high AUCs across multiple moments in time and stratifying patients into different risk groups, with radiomic features in relation to specific long non-coding RNAs (22). Peng et al. applied MRI radiomics to split tumors into the updated HER2 categories, reporting AUCs between 0.75 and 0.89, and showing that the addition of pathological predictors improved classification performance (23). In another multicenter investigation, Chen et al. made use of habitat-based MRI radiomics to calculate intra-tumoral heterogeneity for HER2 classification, achieving AUCs up to 0.94, especially for improving discrimination between HER2-low and HER2-zero cancers (24). Moreover, Huang et al. developed a deep learning radiopathomics model combining preoperative ultrasound images and biopsy whole-slide H&E images, which significantly outperformed single-variable models in distinguishing luminal from non-luminal early-stage breast cancers (AUC 0.929 internal, 0.900 external) (25). Generally, such studies underscore the versatility of AI in breast cancer radiomics, demonstrating its ability to non-invasively predict prognosis, update molecular subtyping, and use multimodal data to enhance preoperative decision-making.

Highlights and Limitations of the Present Research

This research represents one of the few existing studies that focus specifically on AI-based predictions of multiple histopathological features of DCIS, derived from specimen mammography, conducted within a clearly defined and approved research project. By concentrating on a well-defined DCIS cohort, selected from the subgroup where our previous margin-assessment study indicated favorable AI performance (9), we aimed to maximize the likelihood of obtaining meaningful imaging-pathology correlations. The study also addressed four distinct histopathological features (comedo necrosis, stromal invasion, nuclear grade, and architectural pattern), providing a general assessment of the model's versatility. The design allowed for a direct, paired comparison between AI predictions and histopathological exam, offering details regarding how the model aligns with clinical reality and where it diverges. Moreover, by integrating these evaluations into a multi-categorical and binary task framework, the generated evidence could lead to future model optimization for specific diagnostic objectives rather than general-purpose use.

Despite these strengths, several limitations should be noted. The model's performance was inconsistent, showing more reliability in simpler binary tasks compared to complex classifications that imply discrimination between subtle morphological subtypes. Such variability suggests the fact that the current feature extraction could be insufficiently sensitive to the imaging characteristics that correspond to nuclear grade or architectural heterogeneity in DCIS. The relatively small sample size and class imbalances, particularly for certain architectural subtypes and other rare grades, have contributed to reduced accuracy and precision. Additionally, the model's reliance on 2D specimen mammograms alone limits the capture of 3D morphological complexity, potentially diminishing its predictive accuracy. Finally, the model remains a proof of concept, requiring further validation in larger, multi-institutional datasets, probably with pre-training models on domain-specific images, before any validation into clinical workflows.

The small sample size, as result of the narrow inclusion criteria that we adopted, limits the statistical power of the present analysis and drawing definitive

conclusions. Our intention was not to establish conclusive clinical evidence, but rather to conduct a primarily methodological exploration which aimed to open a pathway toward future studies. By focusing on a DCIS subgroup in which our previous AI-based margin assessment study (9) showed promising results, we wanted to investigate if similar performance could be achieved for histopathological feature prediction. This approach allowed us to work with a restricted cohort and to test the feasibility of correlating imaging and histopathological features through radiomics, artificial intelligence, and large language models such as ChatGPT-4. The retrospective design of the study, combined with the use of a specific, pre-selected subgroup, inevitably influenced the findings, but the methodology reflects our aim to introduce an innovative approach and encourage further research to focus on AI-driven radiomic analysis into breast cancer diagnosis.

Conclusions

All in all, in this study, the AI model demonstrated limited agreement with histopathology for DCIS nuclear grading and architectural subtype classification, with low macro-average F1-scores. Its performance was moderate for detecting comedo necrosis and stromal invasion, showing higher specificity than sensitivity in the latter. These findings suggest that, while current AI performance is not sufficient for accurate stand-alone characterization of DCIS subtypes and grades, it holds potential for specific morphological assessments. Further optimization in extended training datasets, and combination with multimodal imaging might enhance diagnostic accuracy and clinical applicability.

Author's Contributions

Conceptualization, D.A., R.A.I., V.E. and G.C.D.; methodology, D.A., R.A.I. and V.E.; formal analysis R.A.I., V.E.; investigation, D.A., R.A.I. and V.E.; data curation, D.A. and R.A.I.; writing - original draft preparation, R.A.I. and V.E.; writing - review and editing, D.A. and G.C.D.; supervision, G.C.D. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The current work was conducted as part of a clinical study entitled "Improving Postoperative Outcomes in Breast Surgery through the Use of Artificial Intelligence and Augmented Reality Programs", conducted at Cluj County Emergency Clinical Hospital, 1st Surgery Department, Cluj-Napoca, Romania with the support of the Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

Institutional Review Board Statement

This study was realized in accordance with the Declaration of Helsinki and approved by the Institutional Review Board and Ethics Committee of Cluj County Emergency Hospital (Approval No. 3387/31 January 2025).

Informed Consent Statement

Informed consent was obtained from all participants involved in the study.

Data Availability Statement

The raw data supporting the conclusions of this article will be made available by the authors on request.

Appendix A

Prompt for ChatGPT (Instructions section): "A mammography image of the operative breast lesion specimen is provided. The lesion was localized preoperatively by the radiologist using a percutaneous guide wire. The resection specimen was subsequently oriented with radiopaque marking needles: the shorter needle indicates the 3 o'clock surgical margin, while the longer needle indicates the 12 o'clock margin. Based on the provided image, please assess the following parameters:

- Nuclear grade: low, intermediate, high, mixed, or undefined.
- Architectural type: solid, comedo, cribriform, papillary, mixed (specify subtypes), or unspecified / inconclusive.
- Presence of comedo necrosis: present or absent.
- Presence of stromal invasion: present or absent."

References

1. Zhi Z, Zhao J, Li Q, Li Q, Xu M, Zuo Y, et al. Evolving Perceptions and Attitudes to

Adopting Generative AI in Professional Settings: Multicenter Longitudinal Qualitative Study of Senior Chinese Hospital Leaders. J Med Internet Res. 2025; 27:e75531.
2. Omar M, Nadkarni GN, Klang E, Glicksberg BS. Large language models in medicine: A review of current clinical trials across healthcare applications. PLOS Digit Health. 2024;3(11):e0000662.
3. Cevik J, Seth I, Rozen WM. Transforming breast reconstruction: The pioneering role of artificial intelligence in preoperative planning. Gland Surg. 2023;12:1271-1275.
4. Gommers JJJ, Verboom SD, Duvivier KM, van Rooden CJ, van Raamt AF, Houwers JB, et al. Influence of AI Decision Support on Radiologists' Performance and Visual Search in Screening Mammography. Radiology. 2025;316(1):e243688.
5. Ng AY, Oberije CJG, Ambrózay É, Szabó E, Serfőző O, Karpati E, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. Nat Med. 2023;29(12):3044-3049.
6. Park EK, Kwak S, Lee W, Choi JS, Kooi T, Kim EK. Impact of AI for Digital Breast Tomosynthesis on Breast Cancer Detection and Interpretation Time. Radiol Artif Intell. 2024;6(3):e230318.
7. Gu J, Tong T, Xu D, Cheng F, Fang C, He C, et al. Deep learning radiomics of ultrasonography for comprehensively predicting tumor and axillary lymph node status after neoadjuvant chemotherapy in breast cancer patients: A multicenter study. Cancer. 2023;129(3):356-366.
8. Wu X, Xia Y, Lou X, Huang K, Wu L, Gao C. Decoding breast cancer imaging trends: the role of AI and radiomics through bibliometric insights. Breast Cancer Res. 2025; 27(1):29.
9. Andras D, Ilies RA, Esanu V, Agoston S, Marginean Jumate TF, Dindelegan GC. Artificial Intelligence as a Potential Tool for Predicting Surgical Margin Status in Early Breast Cancer Using Mammographic Specimen Images. Diagnostics (Basel). 2025;15(10):1276.
10. Badve SS, Gökmen-Polar Y. Ductal carcinoma in situ of breast: update 2019. Pathology. 2019;51(6):563-569.
11. Wang J, Li B, Luo M, Huang J, Zhang K, Zheng S, et al. Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance. Signal Transduct Target Ther. 2024;9(1):83.
12. Sanati S. Morphologic and Molecular Features of Breast Ductal Carcinoma in Situ. Am J Pathol. 2019;189(5):946-955.
13. Perez AA, Balabram D, Salles Mde A, Gobbi H. Ductal carcinoma in situ of the breast: correlation between histopathological features and age of patients. Diagn Pathol. 2014;9:227.
14. Wang SY, Shamliyan T, Virnig BA, Kane R. Tumor characteristics as predictors of local recurrence after treatment of ductal carcinoma in situ: a meta-analysis. Breast Cancer Res Treat. 2011;127(1):1-14.
15. Magnoni F, Bianchi B, Corso G, Alloggio EA, Di Silvestre S, Abruzzese G, et al. Ductal Carcinoma In Situ (DCIS) and Microinvasive DCIS: Role of Surgery in Early Diagnosis of Breast Cancer. Healthcare (Basel). 2023;11(9):1324.
16. Shaaban AM, Hilton B, Clements K, Dodwell D, Sharma N, Kirwan C, et al. The presentation, management and outcome of patients with ductal carcinoma in situ (DCIS) with microinvasion (invasion $\leq$ 1mm in size)-results from the UK Sloane Project. Br J Cancer. 2022;127(12):2125-2132.
17. Shiino S, Quinn C, Ball G, Syed BM, Kurozumi S, Tsuda H, et al. Prognostic significance of microinvasion with ductal carcinoma in situ of the breast: a meta-analysis. Breast Cancer Res Treat. 2023;197(2):245-254.
18. Goldberg M, Parpia S, Rakovitch E, Chang L, Bowen J, Lukka H, et al. Long-term outcomes and effects of hypofractionated radiotherapy in microinvasive breast cancer: Analysis from a randomized trial. Breast. 2023;68:189-193.
19. Yu Y, Tan Y, Xie C, Hu Q, Ouyang J, Chen Y, et al. Development and Validation of a Preoperative Magnetic Resonance Imaging Radiomics-Based Signature to Predict Axillary Lymph Node Metastasis and Disease-Free Survival in Patients With Early-Stage Breast Cancer. JAMA Netw Open. 2020;3(12):e2028086.
20. Mao N, Yin P, Zhang H, Zhang K, Song X, Xing D, et al. Mammography-based radiomics for predicting the risk of breast cancer recurrence: a multicenter study. Br J Radiol. 2021;94(1127):20210348.
21. Wu P, Jiang Y, Xing H, Song W, Cui X, Wu XL, et al. Multimodality deep learning radiomics nomogram for preoperative prediction of malignancy of breast cancer: a multicenter study. Phys Med Biol. 2023;68(17).
22. Yu Y, Ren W, He Z, Chen Y, Tan Y, Mao L, et al. Machine learning radiomics of magnetic resonance imaging predicts recurrence-free survival after surgery and correlation of LncRNAs in patients with breast cancer: a multicenter cohort study. Breast Cancer Res. 2023;25(1):132.
23. Peng Y, Zhang X, Qiu Y, Li B, Yang Z, Huang J, et al. Development and Validation of MRI Radiomics Models to Differentiate HER2-Zero, -Low, and -Positive Breast Cancer. AJR Am J Roentgenol. 2024;222(4):e2330603.
24. Chen H, Liu Y, Zhao J, Jia X, Chai F, Peng Y, et al. Quantification of intratumoral heterogeneity using habitat-based MRI radiomics to identify HER2-positive, -low and -zero breast cancers: a multicenter study. Breast Cancer Res. 2024 Nov 22;26(1):160. Erratum in: Breast Cancer Res. 2024;26(1):173.
25. Huang Y, Yao Z, Li L, Mao R, Huang W, Hu Z, et al. Deep learning radiopathomics based on preoperative US images and biopsy whole slide images can distinguish between luminal and non-luminal tumors in early-stage breast cancers. EBioMedicine. 2023; 94:104706.