ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep215

A Robust CNN Approach for BI-RADS Based Breast Cancer Detection

Abdul Samad[1]; Muhammed Kürşad Uçar[2]

[1]Electrical and Electronics Engineering, Faculty of Engineering,
Sakarya University, Sakarya 54187, Türkiye
[2]MKU Technology, Technopolis, Serdivan, Sakarya, Turkey

Publication Date: 2025/09/12

Abstract: Breast cancer remains one of the most prevalent cancers among women and a major cause of cancer-related mortality worldwide. Early and accurate detection is essential for reducing mortality rates, and mammography remains the most effective screening tool. This study proposes a convolutional neural network (CNN) framework for BI-RADS-based breast cancer classification using three publicly available datasets: CBIS-DDSM, INbreast, and KAU-BCMD. A comprehensive preprocessing pipeline, including noise reduction, contrast enhancement, and region-of-interest extraction, was applied, followed by data augmentation to improve generalization. The model was trained and optimized through grid search across multiple hyperparameter settings. The best configuration, with a learning rate of 0.001 and batch size of 32, achieved 92.28% test accuracy, with precision of 99.1% for BI-RADS4-5 cases and recall of 99.5% for BI-RADS 1 cases. These results demonstrate the potential of a custom CNN with robust preprocessing for BI-RADS based detection of breast cancer and highlight its clinical applicability for improving breast cancer detection.

Keywords: *Breast Cancer Detection, BI-RADS, Mammography, Convolutional Neural Networks, Deep Learning.*

How to Cite: Abdul Samad; Muhammed Kürşad Uçar (2025) A Robust CNN Approach for BI-RADS Based Breast Cancer Detection. *International Journal of Innovative Science and Research Technology*, 10(9), 348-354. https://doi.org/10.38124/ijisrt/25sep215

I. INTRODUCTION

Breast cancer is the most prevalent cancer among women and a major cause of cancer-related deaths worldwide. Its incidence continues to rise, with particularly high case numbers reported in Asia, Europe, and North America, highlighting its global health significance [1], [2]. In the United States, breast cancer remains the most diagnosed cancer among women the number of invasive cases is expected to increase from 287,850 in 2022 to 316,950 in 2025 [1]. Early detection remains the cornerstone of reducing mortality, as treatment outcomes are strongly tied to the stage at diagnosis [3].

Mammography has been the primary screening method, and its effectiveness has been enhanced through computer-aided detection (CAD) systems. While conventional CAD systems based on machine learning rely heavily on handcrafted features and large annotated datasets, these approaches often result in reduced generalizability and high false positive rates [4], [5]. Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful alternative by enabling automatic feature learning directly from images, leading to improved accuracy in medical image analysis [6], [7].

The aim of this research is to detect breast cancer based on BI-RADS categories. The Breast Imaging Reporting and Data System (BI-RADS), developed by the American College of Radiology [8], provides a standardized framework for interpreting and reporting mammographic findings. As summarize in Table 1, The BI-RADS categories range from 0 to 6. BI-RADS 0 indicates an incomplete assessment requiring further imaging. BI-RADS 1 represents negative findings, BI-RADS 2 denotes benign findings, and BI-RADS 3 indicates probably benign lesions (<2% risk) that require short-term follow-up. BI-RADS 4 refers to suspicious abnormalities with increasing malignancy risk (subdivided into 4A–4C), BI-RADS 5 reflects a high probability of malignancy (>95%), and BI-RADS 6 corresponds to biopsy-confirmed malignancy requiring treatment planning.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep215

Table 1 Breast Imaging Reporting and Data System (BI-RADS)

Category	Description	Likelihood of Cancer	Clinical Recommendation
0	Incomplete Assessment	N/A	Additional imaging required
1	Negative	No cancer detected	Routine screening
2	Benign	0%	Routine screening
3	Probably Benign	<2%	Short-interval follow-up
4A–C	Suspicious Abnormality	2–95%	Biopsy recommended
5	Highly Suggestive Malignancy	>95%	Immediate biopsy
6	Biopsy-Proven Malignancy	Confirmed	Treatment planning

In this study, two BI-RADS categories were selected: BI-RADS 1 (Negative), representing normal findings, and BI-RADS 4–5 (Suspicious/Malignant), which indicate cases requiring biopsy or further treatment.

II. LITERATURE REVIEW

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide. Early and accurate detection is therefore a critical clinical priority. Traditional machine learning algorithms such as Decision Trees, Random Forests, and Support Vector Machines (SVM) have been widely applied to this task, achieving moderate success in differentiating between benign and malignant lesions. However, these approaches typically rely on hand-crafted features, which often fail to capture the complex patterns present in mammographic images.

With recent advances in deep learning, Convolutional Neural Networks (CNNs) have become the dominant in disease detection, outperforming conventional methods due to their ability to automatically learn discriminative and hierarchical image features. Transfer learning has shown strong potential in medical imaging applications, as models originally trained on large-scale datasets such as ImageNet (e.g., VGGNet, ResNet, Inception, DenseNet, MobileNet, and EfficientNet) can be fine-tuned for breast cancer detection tasks. This strategy not only accelerates training but also consistently improves classification performance [9], [10].

In research [11] applied a CNN-based binary classification approach and achieved an accuracy of 0.929, highlighting the efficacy of deep models for lesion detection. Similarly, another study [5] on the DDSM-400 and CBIS-DDSM datasets reported that ResNet-101 achieved an accuracy of 78.5% and an AUC of 0.859 in binary classification. More recently, in another research [12] introduced a hybrid feature-fusion framework for three-class classification of breast lesions such as malignant, benign, and normal, achieving accuracies above 97% across MIAS, CBIS-DDSM, and INbreast, thus demonstrating robustness across heterogeneous datasets. In BI-RADS-oriented research [13], fine-tuned InceptionResNetV2 on the RSNA and Vindir datasets, reporting 91% precision and an F1-score of 0.80 for BI-RADS 0 category.

A critical gap in the literature, however, is the tendency to merge BI-RADS categories, particularly BI-RADS 1 (negative) and BI-RADS 2 (benign). Although this simplification can balance datasets, it risks obscuring clinically distinct diagnostic meanings and may increase the likelihood of false-negative assessments. Our research addresses this limitation by explicitly distinguishing between BI-RADS categories in the classification process, thereby aligning computational outputs more closely with clinical decision-making.

This study investigates BI-RADS-based breast cancer detection using three publicly available mammography datasets: CBIS-DDSM, INbreast, and KAU-BCMD. The target categories are BI-RADS 1 (Negative) and BI-RADS 4–5 (Suspicious/Malignant). A rigorous preprocessing pipeline was applied, including noise reduction, Contrast Limited Adaptive Histogram Equalization (CLAHE) for contrast enhancement, and region-of-interest (ROI) extraction to isolate diagnostically relevant regions. To enhance generalization and mitigate overfitting, data augmentation was employed, with the datasets partitioned into 80% training, 10% validation, and 10% testing.

A designed CNN model was implemented in Python to perform classification. By integrating domain-specific preprocessing with deep learning, the study aims to provide a robust framework for BI-RADS categories-based mammogram classification that preserves clinically meaningful distinctions and improves early breast cancer detection.

III. MATERIAL AND METHOD

In this research, Figure 1 illustrates the approach used for the classification of breast cancer into two categories: BI-RADS 1 (Negative) and BI-RADS 4/5 (malignant). To improve robustness and generalizability, multiple datasets were combined to increase the overall data size. Data preprocessing was applied to all mammograms, including noise reduction and contrast enhancement, to improve image quality. Furthermore, data augmentation techniques were employed to expand the dataset and enhance generalization for more reliable breast cancer detection. The Python programming language was used to develop a Convolutional Neural Network (CNN)-based architecture, which was subsequently utilized to perform the classification task.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep215

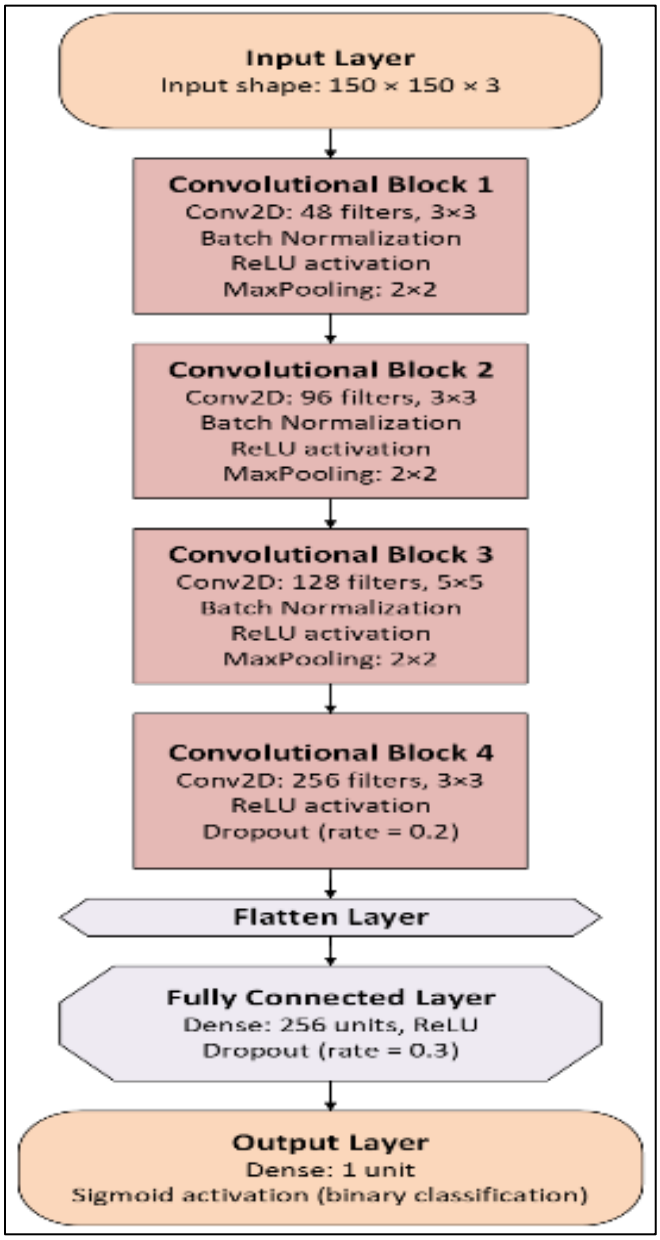

Fig 1 Flow Diagram

IV. DATASET

In this study, three open-source datasets have been used. INbreast, CBIS-DDSM, and KAU-BCMD. All the datasets contain the BIRADS (Breast imaging reporting and data system) information. The goal of this study is to perform Birads based breast cancer detection. All the datasets used in this study were preprocessed using the noise filtration, ROI (Region of interest) cropping and other enhancement methods have been applied.

➢ *INbreast:*

The INbreast dataset [14] developed at Centro Hospitalar de S. João in Porto, Portugal, contains 410 images from 115 cases with high-quality annotations provided by expert radiologists. It covers a wide range labeled with BI-RADS categories. For this study, images corresponding to BI-RADS 1, BI-RADS 4, and BI-RADS 5 were utilized.

➢ *CBIS-DDSM:*

The CBIS-DDSM dataset [15] is an enhanced version of the original DDSM, distributed in modern DICOM format. It consists of 1,696 mass cases. From this dataset, cases belonging to BI-RADS 1, 4, and 5 were selected.

➢ *KAU-BCMD:*

The KAU-BCMD dataset [16] was developed at King Abdulaziz University between 2019 and 2020. Each categorized using the BI-RADS system. For this research, cases belonging to BI-RADS 1, 4, and 5 were included.

For binary classification, BI-RADS 1 (Negative) was contrasted with BI-RADS 4 and 5, which were merged into a malignant class, reflecting clinical practice where BI-RADS 4–5 indicate high cancer probability. In total, the study used 1,935 normal cases and 1,294 malignant cases across the combined datasets.

The class distribution is summarized in Table 2.

Table 2 Class Distribution of Datasets

Dataset	BI-RADS 1	BI-RADS 4	BI-RADS 5	Total Selected Images
INbreast	67	43	49	159
CBIS-DDSM	3	702	374	1,079
KAU-BCMD [3]	1,865	102	24	1,991
Total	1,935	847	447	3,229

➢ *Preprocessing*

Preprocessing mammogram images is a critical step in developing reliable breast cancer detection systems. In this study, a multi-stage pipeline was implemented to enhance image quality, suppress noise, and crop significant regions of interest (ROI). The steps are as follows:

- Images were converted to grayscale to ensure consistency, as mammographic features are primarily intensity-based.
- Gaussian blur (5×5 kernel) was applied to reduce noise while preserving structural details.
- Otsu's automatic thresholding method was used to separate breast tissue from the background.
- Contour detection was performed to identify breast boundaries, with the largest contour selected as the ROI.
- A bounding rectangle was computed around the largest contour to crop out irrelevant background areas.
- A 5×5 median filter was applied to further reduce noise while preserving edges.
- Histogram equalization was performed to enhance contrast and highlight subtle variations such as masses or microcalcifications.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep215

This pipeline effectively reduced artifacts, improved tissue visibility, and prepared the mammograms for subsequent classification tasks. The process is illustrated in Figure 2.

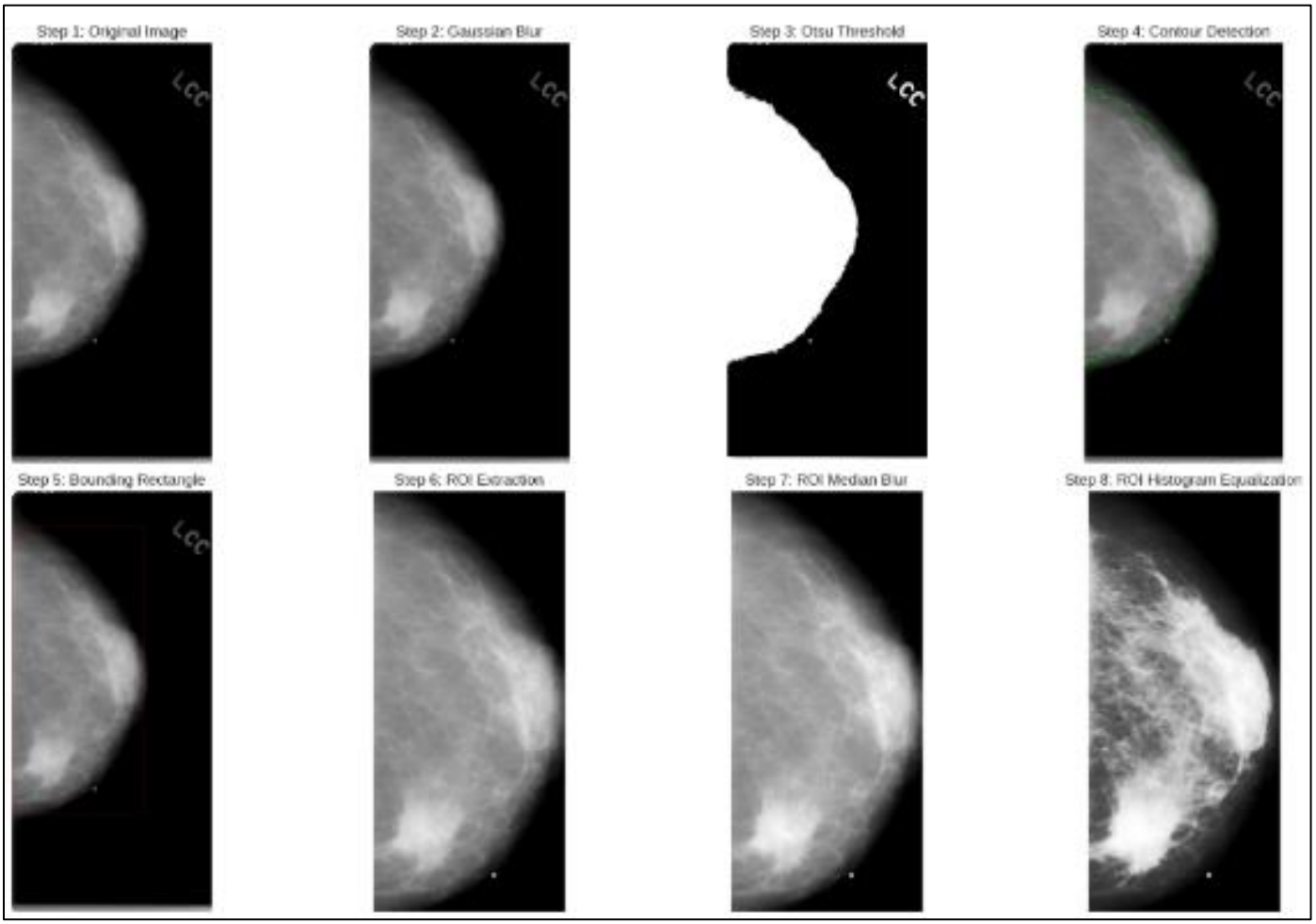

Fig 2 Data Preprocessing

➢ *Proposed Model*

A research-modified Convolutional Neural Network (CNN) was implemented for binary breast cancer classification, distinguishing BI-RADS 1 (normal) from BI-RADS 4/5 (malignant) cases. The network was designed with four convolutional blocks of increasing depth, integrating modern architectural improvements for robust feature learning.

- Block 1: 48 filters (3×3), ReLU activation, Batch Normalization, and 2×2 max pooling.
- Block 2: 96 filters (3×3), ReLU activation, Batch Normalization, and 2×2 max pooling.
- Block 3: 128 filters (5×5), ReLU activation, Batch Normalization, and 2×2 max pooling.
- Block 4: 256 filters (3×3), ReLU activation, followed by Dropout (0.2).

The flattened feature maps were passed to a fully connected dense layer of 256 neurons (ReLU), followed by Dropout (0.3). The final output layer was a single sigmoid unit, producing probability values for binary classification. This architecture combined Batch Normalization for training stability, Dropout for regularization, and carefully selected filter sizes to capture both fine-grained and high-level features.

➢ *Hyperparameter Optimization Via Grid Search*

Hyperparameters, such as learning rate and batch size, play a critical role in the training performance of deep learning models. Unlike model parameters that are learned during training, hyperparameters must be chosen before training starts, and their values strongly influence model convergence, generalization, and stability. Grid search is a systematic optimization technique that explores combinations of predefined hyperparameter values to identify the configuration that yields the best performance.

In this research, a grid search was conducted to optimize the CNN training process. Three learning rates (0.0002, 0.001, and 0.01) and three batch sizes (16, 32, and 64) were evaluated in all possible combinations. Each configuration was trained using the Adam optimizer with binary cross-entropy loss for up to 15 epochs. The optimal hyperparameter configuration was selected based on the highest accuracy, ensuring reliable classification of both classes.

➢ *Data Augmentation*

Data augmentation is a widely used technique which is used to expand the size and diversity of training datasets [17]. By applying geometric and intensity-based transformations,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep215

additional samples are generated that preserve diagnostic features while introducing variability. This helps to prevent overfitting and improves the ability of deep learning models to generalize to unseen data.

In this study, data augmentation was applied exclusively to the training dataset, while validation and test sets were only rescaled for unbiased evaluation. The augmentation operations used are summarized in Table 3, including horizontal flipping, spatial shifts, and intensity rescaling. These transformations preserved the diagnostic integrity of mammograms while enabling CNN to learn more robust feature representations.

Table 3 Augmentation Techniques

Technique	Description	Parameters
Rescaling	Normalized pixel intensities	1/255
Horizontal Flip	Random left–right flip	50% probability
Width Shift	Horizontal translation	±10% of image size
Height Shift	Vertical translation	±10% of image size
Fill Mode	Filled empty regions after shifting	Nearest neighbor

➢ *Performance Evaluation Criteria*

To evaluate the performance of the proposed model, several key metrics were employed, including Accuracy, sensitivity, specificity, Precision, Recall, F1-score, and the Confusion Matrix. These metrics provide a comprehensive assessment of the model's ability to distinguish between classes.

V. RESULT

The convolutional neural network (CNN) was assessed for binary classification of mammographic images into BI-RADS 1 (Normal) and BI-RADS 4–5 (malignant). The independent test set comprised 650 images, with 390 benign and 260 malignant cases. A systematic grid search across nine learning rate and batch size combinations revealed substantial variability in model performance, underscoring the critical influence of hyperparameter selection. The detailed results of all experiments are summarized in Table 4.

Table 4 Performance Evaluation Results

Exp	Learning Rate	Batch Size	Test Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	0.0002	16	84.77	91.28	68.46	78.24
2	0.0002	32	90.77	89.37	87.31	88.33
3	0.0002	64	90.15	97.57	77.31	86.27
4	0.001	16	92.77	98.63	83.08	90.19
5	**0.001**	**32**	**92.92**	**99.08**	**83.08**	**90.38**
6	0.001	64	65.08	100.00	12.69	22.53
7	0.01	16	86.00	100.00	65.00	78.79
8	0.01	32	90.46	98.53	77.31	86.64
9	0.01	64	69.08	94.03	24.23	38.53

The best-performing configuration was achieved with a learning rate of 0.001 and batch size of 32, yielding a test accuracy of 92.28%. For BI-RADS 1 (Normal) cases, the model attained a precision of 89.8% and recall of 99.5%, while for B-IRADS 4-5 (malignant) cases, it achieved a precision of 99.1% and recall of 83.1%. The corresponding confusion matrix in Figure 3 indicated 388 true negatives, 216 true positives, 2 false positives, and 44 false negatives, demonstrating a favorable balance between minimizing false positives and maintaining clinically meaningful sensitivity.

Other experimental configurations produced strong results but were less balanced than the best model. With a learning rate of 0.001 and batch size of 16, the model reached an accuracy of 92.77% and very high precision 98.6% for BI-RADS 4-5 cases, but its recall was slightly lower, meaning more BI-RADS 4-5 lesions were missed. Using a learning rate of 0.0002 and batch size of 32 increased recall to 87.3%, but precision dropped to 89.4%, leading to more false positives. In contrast, models trained with a batch size of 64 proved unstable, and in one case, with a learning rate of 0.001 and batch size of 64, performance collapsed to 65.1% accuracy with BI-RADS 4-5 recall as low as 12.7%, failing to detect most BI-RADS 4-5 cases. This outcome underscores the importance of careful hyperparameter selection for ensuring both accuracy and clinical reliability.

The experimental analysis establishes that the optimal configuration of a 0.001 learning rate with a batch size of 32 provides the most robust and clinically relevant performance.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep215

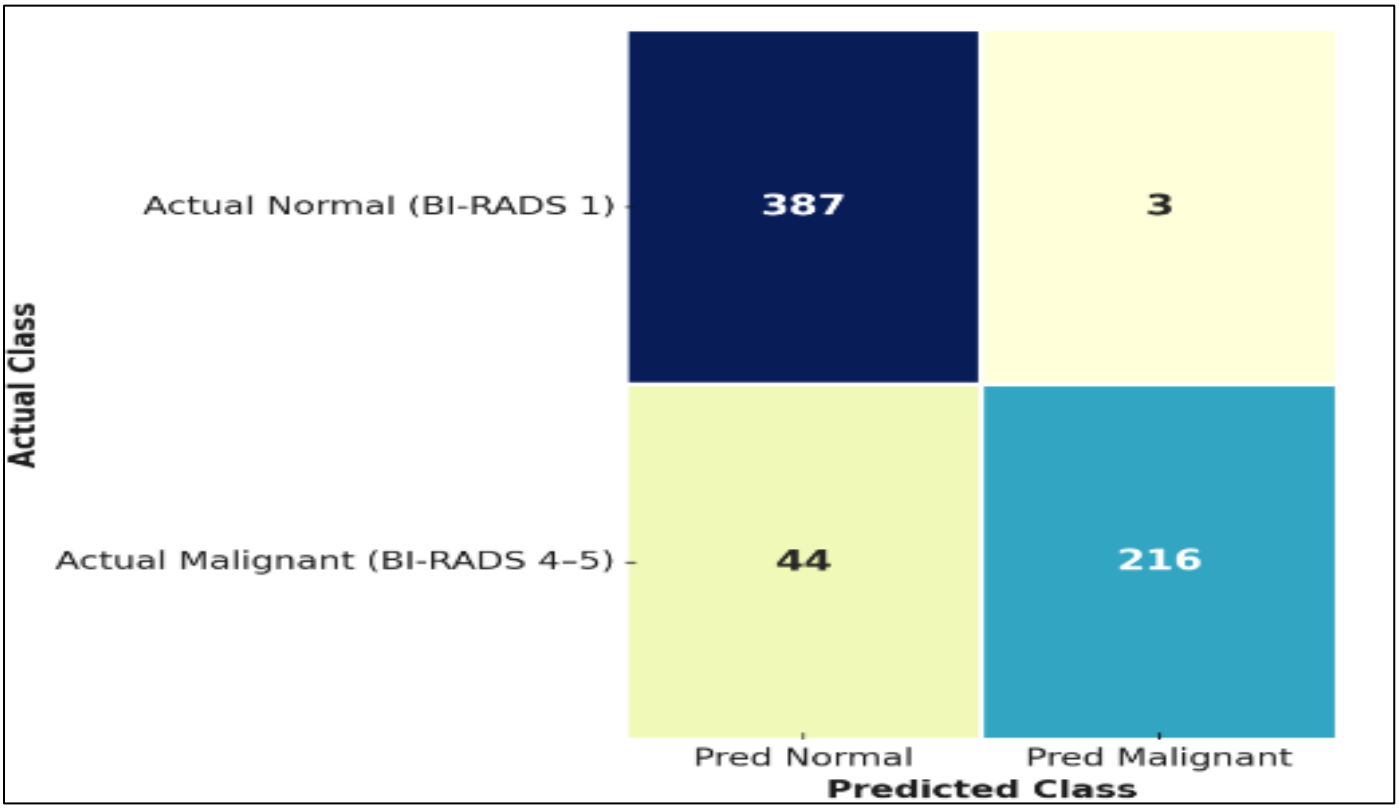

Fig 3 Confusion Metrix of Best Model

VI. DISCUSSION

The results of this study demonstrate that the proposed CNN-based model achieved strong performance in classifying mammographic images into BI-RADS 1 (Negative) and BI-RADS 4–5 (Suspicious/Malignant). The optimal configuration, obtained with a learning rate of 0.001 and batch size of 32, delivered test accuracy of 92.28%, with particularly high precision 99.1% for BI-RADS 4-5 cases and 99.5% for Negative cases. These outcomes suggest that the model effectively balances the trade-off between sensitivity and specificity, which is critical for clinical adoption. By minimizing false positives while maintaining robust detection of BI-RADS 4-5 lesions, the framework provides diagnostic outputs that align with clinical expectations.

When compared with existing literature, the proposed approach demonstrates competitive results. For example, a CNN-based binary classification study reported an accuracy of 92.9% [11], while another transfer learning approach using ResNet-101 on the DDSM-400 and CBIS-DDSM datasets achieved an accuracy of 78.5% with an AUC of 0.859 [5]. More recent work on hybrid feature-fusion frameworks for three-class classification reported accuracies of 98.7% on MIAS, 97.7% on CBIS-DDSM, and 98.8% on INbreast [18]. In BI-RADS-focused research, InceptionResNetV2 achieved 91% precision and an F1-score of 0.80 on the RSNA and Vindir datasets [13]. Although the performance of our model is slightly lower than some multi-class methods, the results highlight the effectiveness of a custom CNN designed specifically for BI-RADS based binary classification, particularly when combined with robust preprocessing and augmentation strategies.

The findings of this study have important implications. Unlike many works that merge BI-RADS 1 and 2 for class balancing, this study preserves their distinction, thereby ensuring greater clinical relevance and reducing the risk of false negatives. Future research should focus on integrating attention mechanisms and transformer-based architectures to further enhance robustness, as well as extending the framework to cover a wider range of BI-RADS categories for more comprehensive diagnostic assessment.

VII. CONCLUSION

This study presented a CNN-based framework for classifying mammographic images into BI-RADS 1 (Negative) and BI-RADS 4–5 (Suspicious/Malignant) categories. The model, trained on combined datasets with rigorous preprocessing and augmentation, achieved strong performance, with the best configuration yielding 92.28% accuracy and a favorable balance between precision and recall. By preserving the distinction between BI-RADS 1 and 2, the study ensures greater clinical relevance and reduces the likelihood of false negatives compared to approaches that merge these categories. While the model generalized well across three datasets. Future research should focus on further BI-RADS categories for a more comprehensive diagnostic tool.

ACKNOWLEDGMENT

This study did not require ethical approval since it used a publicly accessible dataset obtained from an open-source repository.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep215

REFERENCES

[1]. R. L. Siegel, T. B. Kratzer, A. N. Giaquinto, H. Sung, and A. Jemal, "Cancer statistics, 2025," CA Cancer J Clin, vol. 75, no. 1, pp. 10–45, Jan. 2025, doi: 10.3322/CAAC.21871,.

[2]. S. Boumaraf, X. Liu, C. Ferkous, and X. Ma, "A New Computer-Aided Diagnosis System with Modified Genetic Feature Selection for BI-RADS Classification of Breast Masses in Mammograms," Biomed Res Int, vol. 2020, 2020, doi: 10.1155/2020/7695207,.

[3]. V. Sureshkumar, R. S. N. Prasad, S. Balasubramaniam, D. Jagannathan, J. Daniel, and S. Dhanasekaran, "Breast Cancer Detection and Analytics Using Hybrid CNN and Extreme Learning Machine," J Pers Med, vol. 14, no. 8, Aug. 2024, doi: 10.3390/JPM14080792,.

[4]. G. Litjens et al., "A survey on deep learning in medical image analysis," Med Image Anal, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[5]. L. Tsochatzidis, L. Costaridou, and I. Pratikakis, "Deep learning for breast cancer diagnosis from mammograms — A comparative study," J Imaging, vol. 5, no. 3, Mar. 2019, doi: 10.3390/JIMAGING5030037,.

[6]. G. K. Thakur, A. Thakur, S. Kulkarni, N. Khan, and S. Khan, "Deep Learning Approaches for Medical Image Analysis and Diagnosis," Cureus, vol. 16, no. 5, p. e59507, May 2024, doi: 10.7759/CUREUS.59507.

[7]. H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique," IEEE Trans Med Imaging, vol. 35, no. 5, pp. 1153–1159, May 2016, doi: 10.1109/TMI.2016.2553401.

[8]. A. Arian, K. Dinas, G. C. Pratilas, and S. Alipour, "The Breast Imaging-Reporting and Data System (BI-RADS) Made Easy," Iranian Journal of Radiology, vol. 19, no. 1, Jan. 2022, doi: 10.5812/IRANJRADIOL-121155.

[9]. J. F. Lazo, S. Moccia, E. Frontoni, and E. de Momi, "Comparison of different CNNs for breast tumor classification from ultrasound images," Convegno Nazionale di Bioingegneria, pp. 560–563, Dec. 2020, Accessed: Aug. 27, 2025. [Online]. Available: https://arxiv.org/pdf/2012.14517

[10]. P. Ghafariasl, M. Zeinalnezhad, and S. Chang, "Fine-tuning pre-trained networks with emphasis on image segmentation: A multi-network approach for enhanced breast cancer detection," Eng Appl Artif Intell, vol. 139, p. 109666, Jan. 2025, doi: 10.1016/J.ENGAPPAI.2024.109666.

[11]. D. Lévy and A. Jain, "Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks".

[12]. S. Chakravarthy et al., "Multi-class Breast Cancer Classification Using CNN Features Hybridization," International Journal of Computational Intelligence Systems, vol. 17, no. 1, pp. 1–19, Dec. 2024, doi: 10.1007/S44196-024-00593-7/TABLES/9.

[13]. A. Tekin, B. Toktay, A. C. Günay, H. Yazgan, N. G. İnan, and O. Kocadağlı, "Bi-Rads Classification in Mammography using Deep Learning," Güncel Ekonometrik ve İstatistiksel Uygulamalar ile Akademik Çalışmalar, Nov. 2024, doi: 10.58830/OZGUR.PUB518.C2134.

[14]. "INbreast Dataset." Accessed: Aug. 27, 2025. [Online]. Available: https://www.kaggle.com/datasets/ramanathansp20/inbreast-dataset

[15]. "CBIS-DDSM: Breast Cancer Image Dataset." Accessed: Aug. 27, 2025. [Online]. Available: https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset

[16]. "KAU-BCMD (Mamography dataset)." Accessed: Aug. 27, 2025. [Online]. Available: https://www.kaggle.com/datasets/orvile/kau-bcmd-mamography-dataset?select=King+Abdulaziz+University+Mammogram+Dataset

[17]. T. Kumar, R. Brennan, and M. Bendechache, "Image Data Augmentation Approaches: A Comprehensive Survey and Future directions", Accessed: Aug. 28, 2025. [Online]. Available: https://www.baeldung.com/cs/ml-underfitting-overfitting

[18]. S. Chakravarthy et al., "Multi-class Breast Cancer Classification Using CNN Features Hybridization," International Journal of Computational Intelligence Systems, vol. 17, no. 1, pp. 1–19, Dec. 2024, doi: 10.1007/S44196-024-00593-7/TABLES/9.