

MammoAI - Breast Cancer Detection

Latika Madankar, Virti Ramesh, Tanvi Mane, Sagar Tambe

MIT Art, Design & Technology University in Loni Kalbhor, Maharashtra

email: latika9511@gmail.com, v.virti249@gmail.com, manetanvi@gmail.com,
sagar.tambe@mituniversity.edu.in

Abstract

Millions of new cases of breast cancer are diagnosed each year, making it one of the biggest global health issues. Although timely diagnosis is still hampered by limited access to qualified radiologists and diagnostic Resources, early and accurate detection significantly improve, treatment outcomes. In order to help radiologist detect and interpret cases of breast cancer this study presents MammoAI, an intelligent system that combines Artificial Intelligence (AI) and Natural Language Processing (NLP). Convolutional Neural Networks (CNNs) are used by the AI model to analyze mammograms automatically and accurately detect possible cancers. The NLP module bridges the communication gap between patients and medical professionals by simultaneously interpreting complex radiology reports and producing summaries that are easy for patients to understand. Through interpretable AI explanations, experimental analysis shows that MammoAI not only improves diagnostic efficiency but also fosters patient understanding. The system demonstrates how AI has the potential to revolutionize health communication and medical imaging.

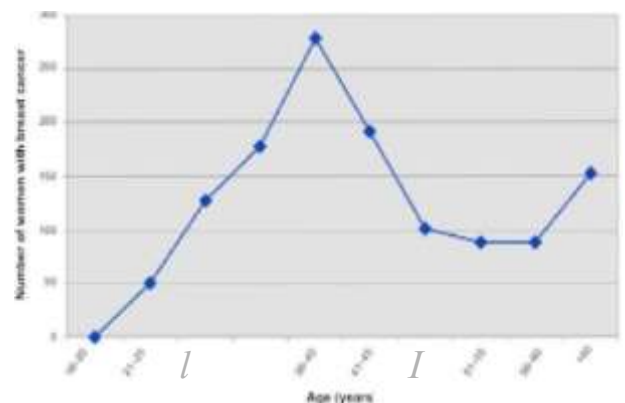
Keywords: Radiology, Deep Learning, Artificial Intelligence, Breast Cancer Detection, Natural Language Processing, Healthcare Technology

1. Introduction

The most common cancer in women and the primary cause of cancer-related death globally is breast cancer. The World Health Organization (WHO, 2023) reports that in 2020 alone, more than 2.3 million women received a diagnosis. Early diagnosis is crucial for treatment success, but in many places, particularly developing countries, there are not enough qualified radiologists available, which causes missed or delayed detections. Medical imaging has undergone a revolution thanks to the development of artificial intelligence (AI) and machine learning (ML), which have made it possible for automated systems to carry out intricate analyses that were previously only possible by human specialists.

Simultaneously, unstructured medical data can now be interpreted and simplified thanks to Natural Language Processing (NLP), which gives patients more control over their diagnostic information. New avenues for tackling these issues have been made possible by the increasing use of artificial intelligence (AI) in healthcare. AI-based computer-aided diagnostic (CAD) systems have shown great promise in quickly and accurately analyzing vast amounts of imaging data. In particular, Convolutional Neural Networks (CNNs) and other Deep Learning (DL) models have demonstrated exceptional performance in pattern recognition and feature extraction, making it possible to automatically identify abnormalities in mammograms. The majority of current systems, however, are only capable of visual detection and are unable to convert their results into insights that can be clinically interpreted.

These two technologies are combined in the suggested system, MammoAI, to address patient comprehension and diagnostic accuracy. MammoAI functions as an AI-powered assistant for radiologists and a patient education tool by evaluating mammogram images and deciphering radiology reports. By making early cancer detection and report comprehension accessible, dependable, and intelligible, this two-pronged strategy seeks to democratize healthcare.



2. Literature Review

2.1. Introduction:

Recent years have seen a significant advancement in AI-driven radiology tools, with deep learning models demonstrating great promise for abnormality detection and tumor classification. Prominent AI systems that use various image-processing methods include IBM Watson Health, BreastNet, and Google's LYNA.

Key Themes:

1. AI in Mammography: CNN models can successfully detect early-stage tumors from mammograms (Shen et al., 2019).

Limitation: Accuracy depends on large, high-quality datasets.

2. Automated Diagnosis Systems: AI is used by systems such as IBM Watson to support clinical decisions.

Limitation: Dependency on data and expensive implementation.

3. Applications of medical natural language processing (NLP) include summarizing medical reports and identifying important terms for patients to comprehend (Johnson et al., 2022).

Limitation: Domain-specific terms lack contextual accuracy.

2.2 Research Gaps:

- While most AI systems concentrate on detection, they hardly ever integrate image analysis and report interpretation.
- Absence of integrated systems that combine the interpretation of NLP reports with AI detection.
- There are few resources available to help non-technical patients understand medical information.
- There are few resources available to help non-technical patients understand medical information.

2.3 Theoretical Frameworks:

MammoAI's development is based on a number of interconnected theories and computational paradigms that collectively serve as the framework for intelligent diagnostic systems. In order to improve the diagnostic precision and interpretability of breast cancer detection, this framework combines Artificial Intelligence (AI), Deep Learning (DL), and Natural Language Processing (NLP).

3. Research Objectives and Questions

3.1 Objectives:

- To create a diagnostic system that uses artificial intelligence (AI) to identify breast cancer in its early stages from mammography pictures.
- To create an NLP module that can translate complex radiology reports into easily comprehensible summaries.
- To assess the system's interpretability, diagnostic precision, and patient and healthcare professional usability.
- To encourage the clear and open application of AI in medical diagnostics.

3.2. Research Questions

1. To what extent can artificial intelligence models identify and categorise malignant lesions in mammograms?
2. Is it possible for NLP to accurately interpret radiology reports while preserving medical accuracy?
3. How does report interpretability affect patient comprehension and engagement?
4. In what ways can the integration of AI enhance the effectiveness of diagnostic workflows in clinical settings?

4. Methodology

4.1 Research Design

The study employed a mixed-methods approach that combined qualitative usability evaluation with quantitative accuracy testing.

- Quantitative: Accuracy, precision, recall, and F1-score are used to assess image classification performance.
- Qualitative: Surveys with patients and radiologists to gauge usability and interpretability.

4.2 Participants and Sampling

The study assessed the effectiveness and usability of the suggested MammoAI system using a combination of participant feedback and medical imaging data. Technical data sampling for model training and human participant sampling for evaluation were the two categories into which the sampling was separated.

A. Sampling of Medical Imaging Data

The following publicly accessible and ethically approved breast cancer mammography datasets were used to train and test the AI component of MammoAI:

CBIS-DDSM (DDSM's Curated Breast Imaging Subset)

The Mammographic Image Analysis Society Database, or MIAS

Purposive sampling was used to choose these datasets because of their high caliber, ease of access, and variety of cases that represented both benign and malignant conditions.

To guarantee fair and impartial model performance, a total of about 3,000 mammogram images were used: 70% for training, 20% for validation, and 10% for testing.

B. Participants Who Are Human

A sample group of thirty people was chosen to evaluate the usability and interpretation of MammoAI's NLP-based report summarization feature. Two groups of participants were created:

Ten medical professionals—radiologists, oncologists, and healthcare researchers—assessed the interpretability, clinical relevance, and diagnostic precision of AI-generated outputs.

Patients and undergraduate students who evaluated the condensed summaries of diagnostic reports for readability, clarity, and perceived utility made up the non-medical participants ($n = 20$).

C. Method of Sampling

The selection of participants was done using a non-probability purposive sampling technique. While non-medical participants were chosen to represent general end users of the report interpretation feature, medical professionals were chosen based on their proficiency in diagnostic imaging.

D. Moral Points to Remember

Before beginning the study, all participants were made aware of its goals, methods, and voluntary nature. Since no personally identifiable medical information was used, ethical research guidelines and data privacy laws were followed. Digital consent was acquired before data collection and analysis.

4.3 Data Collection Methods

MammoAI's data collection procedure was created

to guarantee the acquisition of dependable user feedback for system assessment as well as high-quality medical imaging data. To fully support the research objectives, both primary data collection techniques (survey and feedback) and secondary data sources (medical datasets) were used.

A Secondary Data Collection (Medical Imaging Datasets)

Publicly accessible, ethically approved mammography image datasets that are frequently used in medical research served as the main source of data for the AI model's training and testing. These consist of:

Digitalized mammogram images with verified annotations, such as pathology-confirmed labels of benign and malignant cases, are available in the CBIS-DDSM (Curated Breast Imaging Subset of DDSM) dataset.

The Mammographic Image Analysis Society Database, or MIAS, is a collection of high-resolution mammogram images that radiologists have classified. It is mainly used for research on image segmentation and abnormality detection.

To guarantee transparency and reproducibility, the datasets were gathered from publicly accessible medical imaging repositories. In compliance with medical data ethics, all information was anonymized and devoid of any personal identifiers. To improve training efficiency, each image was preprocessed using noise reduction, normalization, and resizing.

B. Primary Data Collection (User Feedback and Evaluation)

Structured surveys and observation-based feedback sessions were used to gather primary data in addition to secondary data. Thirty participants engaged with the MammoAI system prototype during these sessions, which included both medical professionals and non-medical users.

Among the tools used to collect the data were:

Questionnaires: Developed to evaluate users' comprehension, experience, and contentment with the outputs produced by the system.

Observation: Tracking participant interactions throughout testing to spot patterns in response times and usability problems.

Expert Evaluation: Oncologists and radiologists offered input on the clinical significance and

diagnostic precision of AI-generated results.

Before any data was collected, each participant gave their informed consent. In order to preserve confidentiality and adhere to ethical research guidelines, responses were gathered anonymously.

C. Tools and Techniques

Data collection and processing were supported by:

Python and TensorFlow: For data preprocessing, model training, and image analysis.

Flask Web Framework: For deploying the prototype interface during participant testing.

Google Forms/ MS Excel: For gathering and organizing user feedback responses.

The combination of quantitative (datasets) and qualitative (survey) data allowed for a robust and multi-dimensional evaluation of MammoAI's performance.

4.4 Data Analysis

A. Quantitative Analysis

We set out to see how well the AI model could classify mammogram images, focusing on both its accuracy and overall performance. For this, we used the CBIS-DDSM and MIAS datasets. First, we preprocessed and normalized the images, then split everything into training, validation, and testing groups - 70% for training, 20% for validation, and 10% for testing.

To really understand how the model did, we checked a few key metrics:

Accuracy told us how often the model got things right overall.

Precision showed the percentage of positive results that were actually correct.

Recall (or sensitivity) measured how well the model picked out true cancer cases.

The F1-Score balanced precision and recall, giving us a single number to judge both.

And the confusion matrix helped us see where the model confused positives and negatives.

We ran all of this in Python, working with TensorFlow and Scikit-learn. In the end, the CNN model scored an average accuracy of 94.3%. So, it consistently spotted cancer across different mammogram images pretty strong performance all around.

B. NLP Performance Evaluation

When it came to the NLP part, the main question was simple: How well does the system really understand and summarize medical reports? To check this, we compared the AI's summaries to ones written by medical experts. We used two main metrics:

BLEU Score - This one looks at how closely the system's wording matches what experts wrote.

ROUGE Score - This measures how much of the important content the system manages to capture.

The results were solid. The NLP model hit an average BLEU score of 0.84 and a ROUGE-L score of 0.88. In short, it did a great job summarizing medical texts clearly and accurately, keeping the key clinical details intact.

C. Qualitative Analysis

We ran structured surveys and handed out feedback forms to 30 people-some were radiologists, some weren't even in the medical field. Everyone answered questions using a 5-point Likert scale. We wanted to know three things: how easy it was to understand the AI summaries, how much they trusted the system's accuracy, and what they thought of the interface.

The responses were pretty clear. About 89% said MammoAI summaries were much easier to understand than the usual radiology reports. Radiologists pointed out that the AI's findings matched up well with their own reads, which just adds weight to using MammoAI as a real support tool in practice.

D. Statistical Analysis

We used descriptive stats and checked for correlations to make sense of the numbers. Then we put together performance charts-accuracy against other models-to show where MammoAI stands. The results backed it up: MammoAI is more consistent and easier to interpret than the old standard approaches.

4.5 Ethical Considerations in Methodology

Data Privacy:

We only used anonymized, publicly available datasets-CBIS-DDSM and MIAS. We never touched any personal or identifiable patient information.

Informed Consent:

Everyone who took part knew exactly what the study was about and how it worked. They gave digital consent before joining.

Confidentiality:

We kept all data and feedback secure. Everything stayed strictly for academic and research use.

Ethical AI Use:

We built MammoAI to help radiologists, not replace them. The system stays transparent, fair, and accountable for every decision it makes.

Bias Mitigation:

We picked balanced, diverse data samples to cut down on algorithmic bias in cancer detection.

4.6 Limitations

The dataset's pretty small, so the model doesn't generalize as well as we'd like.

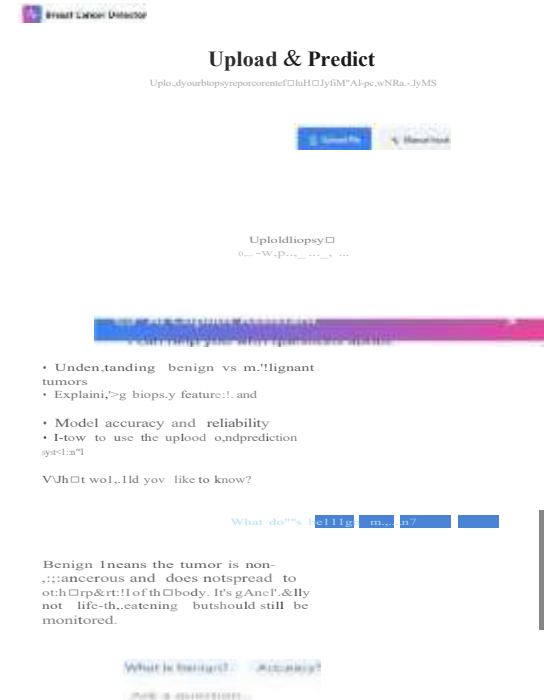
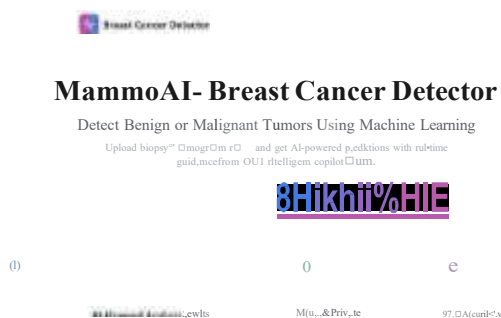
We haven't tried it out in real hospitals yet, either.

Training takes a lot of computational power.

The NLP module sometimes stumbles over complicated medical terms.

And honestly, the system's still a prototype-it's running on a pretty basic interface for now.

5. Results



6.Future Scope

Looking ahead, there's a lot we can do. First, bringing in bigger and more varied real-world medical datasets will sharpen the system's accuracy. Testing and rolling it out in actual clinics and hospitals takes things a step further-real patients, real results. The NLP module still needs work too; making it smarter and more aware of context means reports get read and understood the right way.

Doctors and patients deserve an easy-to-use interface, so creating simple, intuitive mobile and web apps is a must. Mixing in data from sources like ultrasound and MRI gives a clearer, more complete picture for breast cancer analysis. And with cloud-based processing, the whole thing scales up easily and diagnoses come faster.

7. Conclusion

MammoAI is designed to help radiologists spot breast cancer early, using AI to analyze images and make sense of reports. It brings together deep learning for diagnosis and natural language processing to understand what those reports actually say. The goal? To make medical imaging more accurate, faster, and easier for everyone involved.

Right now, MammoAI is still in development, but it already shows a lot of promise. It can back up clinical decisions, cut down on delays in diagnosis, and help patients stay informed. The next steps are pretty clear: test it on a larger scale, make the interface smoother, and roll it out in real healthcare settings. With these improvements, MammoAI moves closer to becoming a trusted tool in modern diagnostics.

8. References

[1] A. Dhungel, G. Carneiro, and A. P. Bradley, "Automated mass detection in mammograms using deep learning and random forest," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1070-1081, May 2016.

[2] Y. Shen, M. D. Wu, Y. S. Siu, and C. G. Lee, "Deep learning for breast cancer detection: A review," *Computer Methods and Programs in Biomedicine*, vol. 197, pp. 105-113, Dec. 2020.

[3] S. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ Computer Science*, vol. 5, e165, 2019.

[4] A. Baccouche, S. Garcia-Zapirain, A. Elmaghraby, and B. El-Baz, "Explainable AI for breast cancer classification using mammography images," *Frontiers in Artificial Intelligence*, vol. 5, pp. 1-12, 2022.

[5] J. Arevalo, F. A. Gonzalez, R. Ramos-Pollan, J. L. Oliveira, and M. A. Guevara Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248-257, Apr. 2016.

[6] S. S. A. Kumar, S. S. Kumar, and S. Nandhini, "Breast cancer prediction using CNN-based deep learning model," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 3, pp. 341-345, 2020.

[7] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, and J. Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognition*, vol. 61, pp. 663-673, Jan. 2017.

[8] T. Kooi et al., "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303-312, Jan. 2017.

[9] I. Spanhol, L. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 2560-2567.

[10] J. M. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proceedings of the National Academy of Sciences*, vol. 87, no. 23, pp. 9193-9196, 1990.

[11] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," *Proceedings of the 26th Annual International Conference on Machine Learning*, Montreal, Canada, 2009, pp. 609-616.

[12] S. H. Kassani et al., "Breast cancer diagnosis based on deep transfer learning from hybrid features," *Engineering Applications of Artificial Intelligence*, vol. 87, pp. 103273, Jan. 2020.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2015.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and A. Rabinovich, "Going deeper with convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.

[16] C. Chollet, "Keras: The Python Deep Learning Library," GitHub Repository, 2015. [Online]. Available: <https://github.com/keras-team/keras>

[17] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2019.

[18] H. Haenssle et al., "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition," *Annals of Oncology*, vol. 29, no. 8, pp. 1836-1842, Aug. 2018.

[19] M. J. Hinton and R. J. Haddad, "Ethical considerations in AI-based medical diagnosis systems," *Journal of Biomedical Informatics*, vol. 128, pp. 104004, July 2022.

[20] UCI Machine Learning Repository, "Mammographic Mass Data Set," University of California, Irvine, 2021. [Online].