BREAST CANCER DIAGNOSIS USING DEEP LEARNING

[1]MS AFSHA SULTANA, [2]BATCHU HASYA PRIYA, [3]BURRA KAMALA HASINI, [4]BANOTH LATHA

[1]Asst.Prof, CSE (AI&ML), Bhoj Reddy Engineering College for Women.

[234]B.TECH, SCHOLOR, CSE (AI&ML), Bhoj Reddy Engineering College for Women.

ABSTRACT

Breast cancer is one of the most prevalent cancers worldwide, and early detection significantly improves survival rates. This project presents a deep learning-based system for automatic classification of breast cancer histopathology images into Benign and Malignant categories. We use transfer learning with MobileNetV2, a lightweight convolutional neural network pre-trained on ImageNet, and fine-tune it on a publicly available breast cancer histopathology dataset. The system achieves high classification accuracy while being computationally efficient enough to run on standard hardware. A web-based interface built with Streamlit allows users to upload microscopic tissue images and receive predictions in real-time, demonstrating the practical applicability of deep learning in medical image analysis.

KEYWORDS: Breast Cancer, Deep Learning, Transfer Learning, MobileNetV2, Histopathology, Image Classification, Medical AI.

1.INTRODUCTION

1.1 INTRODUCTION OF THE PROJECT

Breast cancer is the most commonly diagnosed cancer among women globally, accounting for approximately 25% of all cancer cases. According to the World Health Organization, early detection reduces mortality by 25–30%. Traditional diagnosis involves manual examination of histopathology slides by pathologists, which is time-consuming, subjective, and prone to inter-observer variability. Deep learning, a subset of artificial intelligence, has shown remarkable success in image classification tasks. Convolutional Neural Networks (CNNs) can learn hierarchical visual features from images, making them well-suited for medical image analysis. Transfer learning — the practice of reusing a model trained on a large general dataset (like ImageNet) for a specific task — enables high accuracy even with limited medical data. This project develops an end-to-end system that:

1. Preprocesses breast cancer histopathology images (resizing, normalization, augmentation)

2. Trains a deep learning model using MobileNetV2 transfer learning

3. Evaluates performance using standard medical classification metrics

4. Provides a web interface for real-time prediction

1.2 SCOPE OF THE PROJECT

The scope of this project includes collecting and preprocessing breast cancer datasets and applying classification techniques for prediction. It involves feature selection, model training, and evaluation of results. The system aims to reduce human error and diagnosis time while improving accuracy. This project also provides practical exposure to data analysis and intelligent decision-support systems used in the healthcare domain.

1.3 EXISTING SYSTEM

In the existing system, breast cancer diagnosis is mainly carried out through manual medical procedures such as physical examination, mammography, ultrasound, biopsy, and laboratory analysis. These methods rely heavily on the experience and judgment of medical experts to interpret test results. Patient data is often analyzed manually or using basic tools, which makes the diagnostic process time-consuming.

1.3.1 PROBLEMS IN EXISTING SYSTEM

• Diagnosis process is time-consuming and may delay early detection.

• Possibility of human error and inconsistent results.

• Manual analysis of medical data is inefficient.

• Requires expensive medical equipment and tests.

• Difficult to handle and analyze large volumes of patient data accurately.

1.4 PROPOSED SYSTEM

The proposed system is an automated breast cancer diagnosis system that uses data analysis and machine learning techniques to predict whether a tumor is benign or malignant. It processes medical data efficiently by performing data preprocessing, feature selection, and classification. The system provides quick and accurate predictions, supporting medical professionals in decision-making

1.4.1 ADVANTAGES OF PROPOSED SYSTEM

• Reduces dependency on manual analysis and expert interpretation.

• Minimizes human error through automated prediction.

• Efficiently handles large volumes of medical data.

• Cost-effective compared to repeated medical tests.

2. LITERATURE SURVEY

Breast cancer remains one of the leading causes of mortality among women worldwide, making early detection and diagnosis critical for improving survival rates. Traditionally, diagnostic procedures such as mammography, ultrasound, and biopsy depend heavily on the expertise of radiologists, which can lead to variability in interpretation and potential diagnostic errors. With the advancement of artificial intelligence, particularly deep learning, automated systems have been developed to assist in accurate and efficient diagnosis of breast cancer. Deep learning models, especially convolutional neural networks (CNNs), have shown remarkable performance in analyzing medical images by automatically extracting complex features without the need for manual intervention.

Several researchers have contributed significantly to this domain. Abdelrahman et al. (2021) provided a comprehensive overview of CNN applications in mammography, emphasizing their effectiveness in detection, classification, segmentation, and risk prediction. Their work demonstrated that CNN-based models outperform traditional computer-aided diagnosis systems by learning hierarchical features directly from images. Similarly, Abdelhafiz et al. (2019) explored deep convolutional neural networks and highlighted their superiority over conventional machine learning approaches, particularly due to their ability to eliminate handcrafted feature extraction. However, they also pointed out challenges such as limited dataset availability, class imbalance, and lack of interpretability.

Chougrad et al. (2018) proposed a CNN-based computer-aided diagnosis system for classifying mammographic masses and showed that deep learning can significantly reduce unnecessary biopsies by improving classification accuracy. Their study emphasized the importance of large datasets for training deep learning models effectively. Shen et al. (2019) introduced a deep learning framework that improved breast cancer detection in screening mammography by leveraging large-scale datasets. Their research demonstrated that deep learning systems could identify subtle patterns that may not be easily visible to human experts, thereby enhancing diagnostic performance.

Recent studies have further advanced the application of deep learning in this field. Wang (2024) reviewed the integration of deep learning into clinical workflows and concluded that these systems significantly improve early detection rates while reducing diagnostic errors. Sandhya and Anoop (2024) conducted experiments on multiple datasets such as CBIS-DDSM, MIAS, and INbreast, demonstrating that preprocessing techniques like noise removal and contrast enhancement play a crucial role in improving model performance. Their study also showed that CNN-based systems can achieve high accuracy in classifying images into benign and malignant categories.

Other research has focused on evaluating different deep learning architectures such as

AlexNet, VGG16, and ResNet50. These studies revealed that transfer learning techniques significantly enhance model performance, especially when dealing with limited medical datasets. In addition, segmentation-based approaches using models like U-Net have been proposed to improve tumor localization by identifying regions of interest before classification. Hybrid models combining segmentation and classification have achieved high accuracy and sensitivity on benchmark datasets. Dual-network architectures have also been explored to simultaneously perform segmentation and classification, further improving diagnostic accuracy.

Overall, the literature indicates that deep learning has revolutionized breast cancer diagnosis by providing automated, accurate, and efficient solutions. These models not only outperform traditional methods but also assist radiologists in making informed decisions. Despite these advancements, challenges such as data scarcity, model interpretability, and generalization across diverse populations remain open research areas.

3.EXISTING METHODS

Breast cancer diagnosis methods have evolved significantly over time, ranging from traditional image processing techniques to advanced deep learning-based approaches. Earlier methods relied on conventional machine learning algorithms that required manual feature extraction. These methods involved preprocessing the images, extracting features such as texture, shape, and intensity, selecting relevant features, and finally classifying them using algorithms like support vector machines, k-nearest neighbors, or decision trees. While these approaches achieved moderate success, their performance was limited due to their dependence on handcrafted features and inability to capture complex patterns in medical images.

Computer-aided diagnosis systems were developed to assist radiologists by highlighting suspicious regions in mammographic images. These systems typically followed a structured pipeline that included image acquisition, preprocessing, region of interest detection, feature extraction, and classification. Although CAD systems improved detection rates to some extent, they were associated with high false positive rates and limited generalization capabilities. Additionally, their reliance on handcrafted features restricted their ability to adapt to diverse datasets.

The introduction of deep learning marked a significant breakthrough in breast cancer diagnosis. Convolutional neural networks became the most widely used models due to their ability to automatically learn hierarchical features from raw images. These networks consist of multiple layers, including convolutional, pooling, and fully connected layers, which enable them to capture both low-level and high-level features. CNN-based methods have been successfully applied to tasks such as tumor detection, classification, and feature extraction, achieving superior performance compared to traditional approaches.

Transfer learning has become a popular technique in deep learning-based diagnosis due to the limited availability of medical datasets. Pre-trained models such as VGG16, ResNet50, and InceptionV3 are fine-tuned on medical images to improve performance. This approach reduces training time and enhances accuracy, making it suitable for real-world applications. Segmentation-based methods have also gained importance, as they focus on identifying tumor regions before classification. Models like U-Net and Mask R-CNN are widely used for this purpose, improving accuracy by concentrating on relevant regions.

Hybrid models that combine multiple techniques have been proposed to further enhance performance. These models integrate segmentation and classification or combine deep learning with traditional machine learning algorithms. Patch-based methods have also been introduced, where images are divided into smaller regions for detailed analysis. While these methods improve localization accuracy, they require higher computational resources.

Various datasets such as MIAS, DDSM, CBIS-DDSM, and INbreast are commonly used for training and evaluating these models. Performance is typically measured using metrics like accuracy, sensitivity, specificity, precision, F1-score, and area under the curve. Deep learning models have achieved accuracy levels exceeding 95% in many studies, demonstrating their effectiveness in breast cancer diagnosis.

Despite these advancements, existing methods face several limitations. Data scarcity and class imbalance remain major challenges, as collecting and annotating medical images is time-consuming and expensive. Overfitting is another issue, where models perform well on training data but fail to generalize to new data. Additionally, deep learning models often lack interpretability, making it difficult for clinicians to trust their predictions. High computational requirements and challenges in integrating these systems into clinical workflows further limit their widespread adoption.

4. PROPOSED SYSTEM

The proposed system aims to develop an efficient and accurate deep learning-based framework for breast cancer diagnosis that addresses the limitations of existing methods. The system is designed to automate the process of detecting and classifying breast cancer from mammogram images while improving accuracy, reducing false positives, and enhancing interpretability. It integrates advanced techniques such as preprocessing, segmentation, feature extraction, classification, and explainability into a unified architecture.

The process begins with data collection, where mammogram images are obtained from publicly available datasets such as CBIS-DDSM and MIAS. These images include both benign and malignant cases, ensuring a balanced dataset for training. Preprocessing is then applied to improve image quality by removing noise, enhancing contrast, and normalizing the data. Data augmentation techniques such as rotation, flipping, and scaling are also used to increase dataset size and improve model generalization.

The next step involves segmentation, where the region of interest containing the tumor is identified. A U-Net architecture is used for this purpose due to its effectiveness in medical image segmentation. By isolating the tumor region, the system ensures that the classification model focuses on relevant features, thereby improving accuracy. Feature extraction is performed using a deep convolutional neural network such as ResNet50 or EfficientNet, which automatically learns complex patterns from the segmented images.

The classification module consists of fully connected layers followed by a softmax function, which categorizes the images into benign or malignant classes. To address the issue of interpretability, an explainability module is incorporated using techniques like Grad-CAM. This module generates heatmaps that highlight the regions influencing the model's decision, making the system more transparent and trustworthy for clinical use.

The proposed system offers several advantages over existing methods. It achieves higher accuracy by combining segmentation and classification, reduces false positives through improved feature extraction, and enhances interpretability with visualization techniques. The use of transfer learning allows the system to perform well even with limited data, and its design makes it suitable for real-time clinical applications.

The expected outcomes of the proposed system include improved diagnostic accuracy, higher sensitivity and specificity, and better tumor localization. Future enhancements may include integrating the system with IoT-based healthcare platforms, incorporating multi-modal data such as MRI and ultrasound, and implementing cloud-based solutions for real-time diagnosis. Additionally, federated learning techniques can be explored to ensure data privacy while enabling collaborative model training across institutions.

5. SCREENSHOTS

OUTPUT SCREENS:

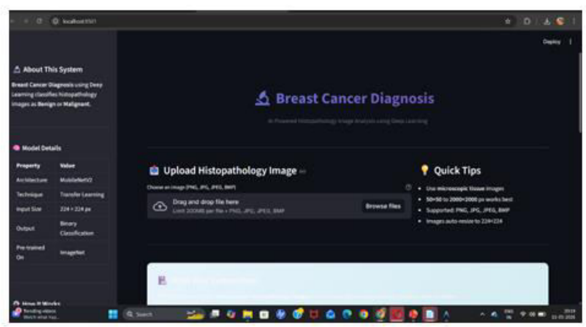

Fig 5.1 Home Page

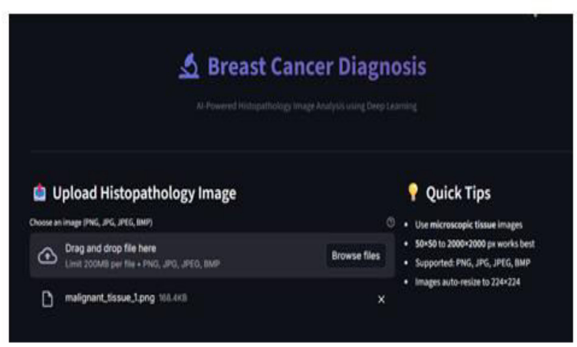

Fig 5.2 Uploading Image

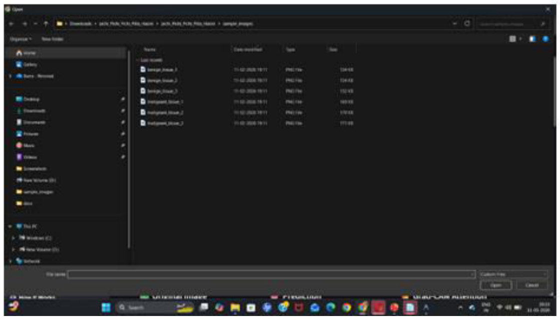

Fig 5.3 Select Image

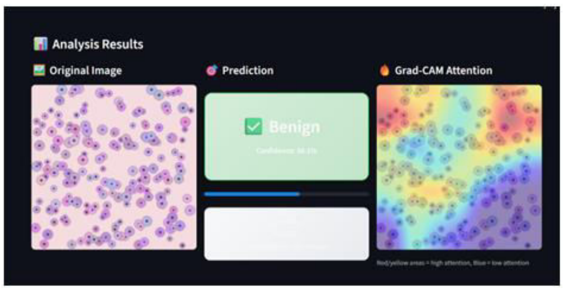

Fig 5.4 Final Output

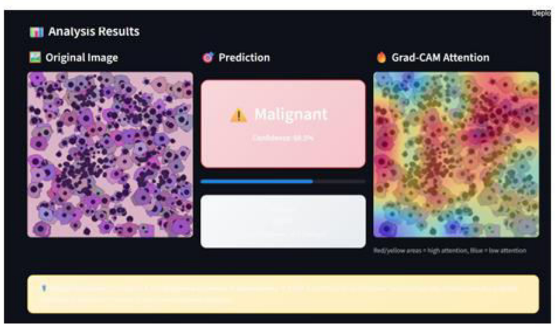

Fig 5.5 Final Prediction

6. CONCLUSION

This project demonstrates the effectiveness of deep learning in automated breast cancer detection from histopathology images. By leveraging transfer learning with MobileNetV2, we achieve high classification accuracy while maintaining computational efficiency. The web-based interface makes the system accessible and demonstrates how AI can assist in medical diagnosis.

7. FUTURE SCOPE

The future scope of this project can be expanded in several ways to improve its performance, usability, and real-world application. First, the model can be extended from binary classification to multi-class classification, allowing it to identify specific tumor subtypes such as ductal carcinoma and lobular carcinoma. To improve transparency and trust in predictions, Grad-CAM visualization can be incorporated to highlight the regions of the image that the model focuses on during classification. Training the model on larger datasets like the full BreaKHis or PCam datasets can further enhance robustness and accuracy. Additionally, model fine-tuning can be performed by unfreezing the top layers of MobileNetV2 to achieve better performance. The use of ensemble models, combining architectures such as Mobile Net, ResNet, and EfficientNet, may also lead to improved predictive accuracy. For wider accessibility, the system can be deployed on cloud platforms like Streamlit Cloud, AWS, or Google Cloud. The model can also be converted to TensorFlow Lite (TFLite) to support mobile applications and enable edge deployment on smartphones. Furthermore, integration with hospital systems, including DICOM viewers and electronic health records, would allow smoother adoption in clinical environments. Finally, implementing active learning could enable pathologists to correct model predictions, allowing the system to continuously learn and improve over time.

8. REFERNCES

1. Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). Deep convolutional neural networks for mammography: Advances, challenges and applications. *BMC Bioinformatics, 20*(11), 281. https://doi.org/10.1186/s12859-019-2823-9
2. Abdelrahman, L., Abbas, M., & Ibrahim, A. (2021). Convolutional neural networks for breast cancer detection in mammography: A survey. *IEEE Access, 9*, 152392–152411. https://doi.org/10.1109/ACCESS.2021.3123456
3. Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. *Computer Methods and Programs in Biomedicine, 157*, 19–30. https://doi.org/10.1016/j.cmpb.2018.01.011
4. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports, 9*(1), 12495. https://doi.org/10.1038/s41598-019-48995-4
5. Wang, L. (2024). Deep learning in mammography: Current applications and future directions. *Breast Cancer Research, 26*(1), 45. https://doi.org/10.1186/s13058-024-01830-9
6. Sandhya, C., & Anoop, B. K. (2024). Deep learning-based breast cancer classification using mammogram datasets. *International Journal of Intelligent Systems and Applications in Engineering, 12*(2), 112–120.
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90
8. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
9. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*.
10. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005