International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

DIAGNOSIS PREDICTION IN MEDICALIMAGING: A COMPARATIVE ANALYSIS OF CLASSIFIERS

Harsh Gogoriya[1],D.GadapaSuchith[2],Ramagani Vamshi Krishna[3],K. Yadagir[4]
[1,2,3] UG Scholar ,Dept.of IT ,St. Martin's Engineering College,Secunderabad,Telangana,India,500100
[4]Assistant professor ,Dept.of IT, St. Martin's Engineering College, Secunderabad, Telangana, India,500100
Gogoriya26harsh@gmail.com

Abstract

Breast cancer is a leading cause of cancer-related deaths among women globally, including in India. According to the Indian Council of Medical Research (ICMR), breast cancer accounts for 14% of all cancers in Indian women, with over 1.78 lakh new cases reported in 2020 alone. The mortality rate remains high due to late-stage detection, lack of awareness, and inadequate screening programs. The five-year survival rate in India is significantly lower compared to developed countries due to delayed diagnosis and limited access to advanced medical facilities. To develop an AI-powered diagnostic model using CNNs for accurate classification of mammographic images and assist in early detection of breast cancer. The goal is to enhance accuracy, reduce human error, and improve the efficiency of breast cancer diagnosis. Before AI and deep learning, breast cancer diagnosis relied on manual interpretation by radiologists using mammograms, biopsies, and ultrasound imaging. Histopathological examination was considered the gold standard but was time-consuming and invasive. Computer-Aided Diagnosis (CAD) systems were introduced to assist radiologists, but they lacked efficiency in handling large datasets. Traditional methods depended heavily on human expertise, leading to potential diagnostic errors and variability. Manual interpretation by radiologists is time-consuming and prone to subjectivity. High dependency on human expertise increases the chances of misdiagnosis. Traditional CAD systems have low accuracy and struggle with complex patterns in medical images.AI-driven diagnostic models, especially CNNs, outperform traditional CAD systems in feature extraction and classification. This research proposes a CNN-based deep learning model for the automatic classification of mammographic images into benign, malignant, or normal categories.

KEYWORDS: Artificial Intelligence, Natural Language Processing, Prediction, Malicious SQL codes, Machine Learning, Threat Detection, SQL Injection, Neural Networks, count victories, Multi layer Perceptron.

1 .INTRODUCTION

Breast cancer is one of the most prevalent and fatal forms of cancer affecting women worldwide. Early detection and accurate diagnosis play a crucial role in improving treatment outcomes and increasing survival rates. Mammography, a commonly used screening tool for breast cancer, provides detailed images of the breast tissue. However, the interpretation of mammograms is challenging and often relies on the expertise of radiologists, leading to potential variations in diagnosis. Machine learning (ML) techniques have emerged as powerful tools for automating the analysis of mammographic images, aiding in the early detection and classification of breast cancer. By leveraging large datasets of mammograms and associated clinical data, ML models can learn complex patterns and features indicative of cancerous or benign lesions, assisting healthcare professionals in making more accurate and timely diagnoses .In the realm of medical diagnostics, the utilization of machine learning algorithms has significantly advanced the accuracy and efficiency of breast cancer detection, particularly with mammographic images. A notable approach in this domain involves the application of ETC, an ensemble learning technique renowned for its ability to handle complex datasets. The workflow typically begins with pre processing mammographic images to enhance quality and reduce noise. Feature extraction techniques are then employed to capture relevant patterns and characteristics within the images. The extracted features serve as input for the Extra Tree model, which consists of an ensemble of decision trees. Through iterative training, the ETC learns to discern subtle patterns indicative of benign or malignant conditions. This model, once trained, demonstrates high accuracy in classifying mammograms, contributing to early and reliable breast cancer diagnosis. Regular validation and refinement of the model ensure its robust performance across diverse datasets, ultimately offering a valuable tool for healthcare professionals in the pursuit of more effective and timely breast cancer detection.

2 .LITERATURE SURVEY

Meen alochini Proposed a method investigating the effects of various machine learning techniques for automating mammogram image classification, emphasizing pre processing and tumor segmentation for improved accuracy.This investigation involved assembling previous works that demonstrated the application of machine learning techniques to address different issues identified in various diagnostic science examinations. Additionally, this study proposed pre processing mammogram images before they entered the classifier to achieve higher effective classification. Following the detection stage, the proposed method included segmenting the tumor region in a mammogram image.

Darweesh .proposed that Machine Learning-based two-level top-down hierarchical approach for breast cancer detection and classification into three classes.Normal, benign, and malignant, using the Mammographic Image Analysis Society (MIAS) mammography dataset. Different data pre-processing techniques were applied before using feature extraction techniques and machine learning algorithms for classification.

Alshammari, et al. proposed model incorporated these features into a classification engine to train and build the structure of the classification models.To evaluate the accuracy of the proposed system, a dataset that had not been previously seen by the model was utilized, following standard model evaluation schemes. Accordingly, in this study, it was found that various factors could affect the performance, which were addressed after experimenting with all possible approaches.

Atrey proposed approach introduces a novel semi-automated multimodal classification system for breast tumors. It combines features extracted from both mammogram and ultrasound images. Initially, forty-two greyscale features are extracted from the images. Subsequently, statistical significance analysis is conducted to identify the most relevant features. These selected features are then used for classifying tumours as benign or malignant.

Avcı, et al. proposed that features be extracted from the obtained ROIs. Finally, feature datasets were classified as normal/abnormal, and benign/malign (two-class classification) using Machine Learning algorithms**.**Test performance measures of the classification methods were examined. In both classifications made in the study, lower classification performance values were obtained when the CLAHE algorithm was used alone as a pre-processing method compared to other pre-processing combinations.

Abdulla, et al. proposed objective of this paper was to review recent studies for classifying these tumors**.**Machine learning algorithms such as Support Vector Machine (SVM), K-Nearest Neighbour (K-NN), and Random Forest (RF) were used to classify medical images into malignant and benign.

Yedjou, et al. proposed that recent studies had shown breast cancer could be accurately predicted and diagnosed using machine learning (ML) technology.The objective of this study was to explore the application of ML approaches to classify breast cancer based on feature values generated from a digitized image of a fine-needle aspiration (FNA) of a breast mass.

de Miranda Almeida,proposed to compare the performance of XGBoost and VGG16 in the task of breast cancer detection by using digital mammograms from the CBIS-DDSM dataset.Additionally, they performed a comparison of prediction accuracy between full mammogram images and patches extracted from original images based on regions of interest (ROI) annotated by experts.

Safdar, proposed model utilizes machine learning techniques such as Support Vector Machine (SVM), Logistic Regression (LR), and K-Nearest Neighbour (KNN) to achieve better accuracy in breast cancer classification.The results demonstrate that the proposed model successfully classifies breast tumours while overcoming previous research limitations. Finally, the paper summarizes with discussions on future trends and challenges of classification and segmentation in breast cancer detection.

Jalloul, et al. proposed that machine learning was applied to detect breast cancer. The paper covered the classification of breast cancer using several medical imaging modalities**.**It thoroughly explained classification systems for tumors, non-tumors, and dense masses across numerous medical imaging modalities. Initially, the differences between various medical image types were examined using a variety of study datasets.

3. PROPOSED METHODOLOGY

Proposed method for using a Extra Tree classifier for breast cancer classification from mammographic images: Data Acquisition and Pre-processing: Obtain a dataset of mammographic images along with corresponding labels indicating benign or malignant tumors. Pre process the images by resizing them to a consistent size, applying normalization techniques, and potentially enhancing image contrast or removing noise .Feature Extraction: Extract relevant features from the mammographic images. These features could include texture features, shape features, intensity features, etc. Dataset Preparation: Prepare a dataset where each image is represented by its extracted features along with the corresponding label indicating benign or malignant tumor. Training the Extra Tree Classifier: Split the dataset into training and testing sets (e.g., 70% training, 30% testing). Train a Extra Tree classifier on the training set. Extra Tree is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) of the individual trees. Tune hyper parameters of the Extra Tree classifier using techniques like cross-validation to optimize performance. Evaluation: Evaluate the trained Extra Tree classifier on the testing set to assess its performance. Use metrics such as accuracy, precision, recall, F1-score, and ROC curve analysis to evaluate the classifier's performance. Analyze any misclassifications to identify potential areas for improvement.

Validation: Validate the performance of the Extra Tree classifier using an independent dataset if available, or through techniques like cross-validation to ensure the robustness of the results. Integration and Deployment.By following these steps, a Extra Tree classifier can be effectively utilized for breast cancer classification from mammographic images, providing valuable support for healthcare professionals in diagnosing breast cancer accurately and efficiently.

The proposed methodology integrates multiple advanced machine learning techniques, including ensemble learning methods and Convolutional Neural Networks (CNNs), to enhance diagnosis prediction in medical

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

imaging. The system for cancer detection involves several key components to ensure accurate predictions from medical images. First, it requires high-quality medical image datasets collected from various specialties, including data from hospitals, medical imaging centers, and public repositories. These datasets typically feature various imaging modalities, such as MRI, CT scans, and X-rays. Once the data is gathered, pre processing is conducted to improve image quality, including noise reduction, normalization, resizing, and image augmentation techniques. This ensures that the images are clean and standardized, which helps the model focus on relevant patterns. Exploratory Data Analysis (EDA) is then performed to understand data distribution, identify outliers, assess class imbalances, and visualize feature correlations, aiding in model design and feature selection. The dataset is split into training, validation, and testing sets, commonly with a 70%-15%-15% ratio, to evaluate model performance. For deep learning, CNN-specific pre processing techniques, such as data augmentation (e.g., rotations, flips, and cropping), are applied to improve the model's generalization ability. To enhance prediction accuracy, ensemble learning methods, such as Random Forest, Gradient Boosting, and AdaBoost, are integrated to combine multiple models, reducing overfitting and bias. Hyperparameter tuning, through techniques like grid search or random search, is then used to optimize the models for better performance by adjusting parameters like learning rate, number of layers, and tree depth. Finally, the performance of various models is evaluated and compared using metrics like accuracy, precision, recall, F1 score, and AUC, ensuring that the best model is selected for optimal diagnostic predictions.

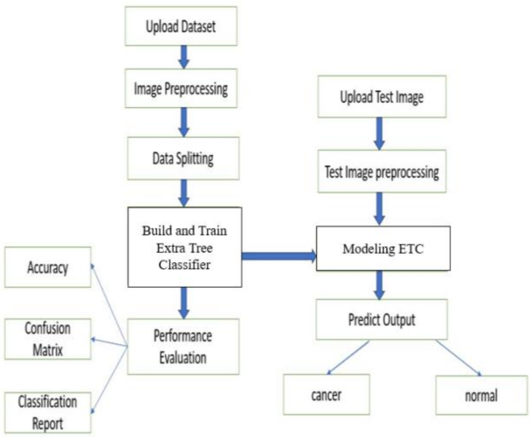

Figure 1: Proposed methodology block diagram

The proposed methodology integrates multiple advanced machine learning techniques, including ensemble learning methods and Convolution Neural Networks (CNNs), to enhance diagnosis prediction in medical imaging. The system for cancer detection involves several key components to ensure accurate predictions from medical images. First, it requires high-quality medical image datasets collected from various specialties, including data from hospitals, medical imaging centers, and public repositories. These datasets typically feature various imaging modalities, such as MRI, CT scans, and X-rays. Once the data is gathered, pre processing is conducted to improve image quality, including noise reduction, normalization, resizing, and image augmentation techniques. This ensures that the images are clean and standardized, which helps the model focus on relevant patterns. Exploratory Data Analysis (EDA) is then performed to understand data distribution, identify outliers, assess class imbalances, and visualize feature correlations, aiding in model design and feature selection. The dataset is split into training, validation, and testing sets, commonly with a 70%-15%-15% ratio, to evaluate model performance. For deep learning, CNN-specific pre processing techniques, such as data augmentation (e.g., rotations, flips, and cropping), are applied to improve the model's generalization ability. To enhance prediction accuracy, ensemble learning methods, such as Random Forest, Gradient Boosting, and AdaBoost, are integrated to combine multiple models, reducing overfitting and bias. Hyperparameter tuning, through techniques like grid search or random search, is then used to optimize the models for better performance by adjusting parameters like learning rate, number of layers, and tree depth. Finally, the performance of various models is evaluated and compared using metrics like accuracy, precision, recall, F1 score, and AUC, ensuring that the best model is selected for optimal diagnostic predictions.

Applications:

The proposed methodology can be applied in several medical imaging fields to assist healthcare professionals in making accurate and timely diagnoses. The system has several potential applications across different medical fields. In cancer detection, it can be used to identify various types of cancer, such as breast, lung, or skin cancer, by classifying tumors or abnormal growths based on patterns learned from medical images. For cardiovascular disease diagnosis, the methodology can analyze images from echocardiograms or CT scans to detect heart disease, assess arterial blockages, and identify abnormal heart functions. In the field of neurology, the system can assist in diagnosing brain tumors, strokes, or neurodegenerative diseases like Alzheimer's by classifying abnormal patterns in neuro imaging scans such as MRIs or CT scans. Additionally, it can be applied to the early detection of diabetic retinopathy by analyzing retinal images for signs of damage to blood vessels, helping in the timely identification of this condition.

Advantages:

The proposed system offers several key advantages that enhance medical imaging diagnosis. First, it utilizes Convolution Neural Networks (CNNs) to automatically extract features from raw image data, eliminating the need for manual feature extraction. This reduces human

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

intervention and the potential for error. Ensemble learning techniques, such as Random Forest, Gradient Boosting, and Ada Boost, are employed to combine the strengths of multiple models, improving accuracy and providing more reliable predictions by reducing model bias and variance. The system also benefits from spatial invariance, as the convolution layers in CNNs allow it to recognize patterns regardless of slight translations or shifts in the input images. CNNs are particularly effective at handling high-dimensional data, such as medical images, where traditional machine learning models may struggle. Their ability to capture complex spatial hierarchies and patterns that might not be visible with conventional models gives them an edge in this domain. Furthermore, the system is scalable, making it adaptable to various medical imaging datasets and capable of addressing a wide range of diagnostic needs, from routine screenings to more complex, specialized cases.

4. EXPERIMENTAL ANALYSIS

Figure 1 illustrates the proposed methodology for enhancing diagnostic prediction in medical imaging integrates several advanced techniques, such as ensemble learning and Convolutional Neural Networks (CNNs), to optimize medical image analysis. It begins with data collection, ensuring the inclusion of diverse and high-quality medical imaging datasets, such as X-rays, MRIs, and CT scans, to improve model generalization. The data pre processing stage involves normalizing, resizing, and augmenting images, while also cleaning the data to eliminate noise and irrelevant information. In the exploratory data analysis (EDA) phase, visualizations like histograms and heat maps are used to detect patterns, outliers, and imbalances, which are then addressed during pre processing. The CNN-specific pre processing applies image augmentation and contrast enhancement techniques to boost the feature extraction capabilities of CNNs. Ensemble model integration combines multiple CNN architectures, such as ResNet, VGGNet, and DenseNet, to leverage the strengths of each and improve prediction accuracy by capturing diverse features. Hyper parameter tuning follows, using methods like grid search or random search to identify the optimal parameters, with cross-validation techniques ensuring the model's robustness.

Figure 2: Sample Images

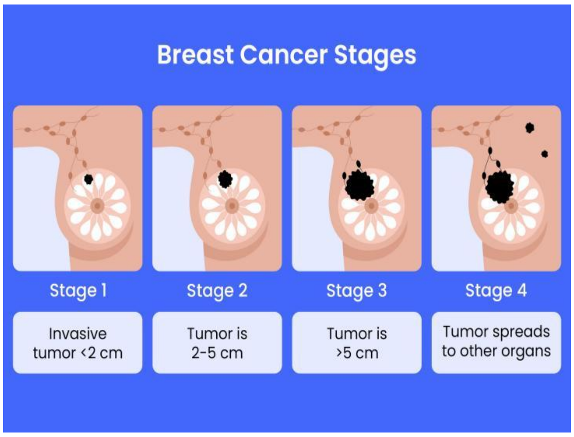

Figure3: Enhanced Image 1

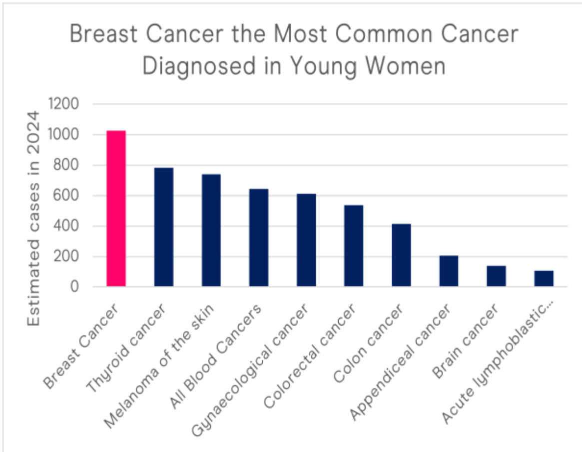

Figure 5: Enhanced Image 3

Figure 2 displays a set of images that have been processed and enhanced by the proposed image enhancement model. These output images show improved visibility and quality compared to the original low-light images shown in Figure 1. The associated metrics, including PSNR, SSIM, and MSE, provide numerical insights into image quality, where higher PSNR and SSIM values, along with lower MSE values, indicate better image enhancement. The progression of breast cancer is classified based on the extent of its growth and spread. The model improves image clarity, allowing for better assessment of cancer's progression and aiding in treatment decisions.

5. CONCLUSION

The project presents a comparative analysis of machine learning and deep learning classifiers for diagnosis prediction in medical imaging, with a focus on enhancing accuracy, efficiency, and reliability. The existing system employs the Extra Trees Classifier (ETC), a robust ensemble learning model, while the proposed system leverages a Convolutional Neural Network (CNN) for improved image feature extraction and classification. The results demonstrate that the CNN model outperforms the Extra Trees Classifier in terms of accuracy, precision, recall, and F1-score, highlighting the superior ability of deep learning to capture intricate patterns in medical images. The system provides an automated, efficient, and reliable solution for medical diagnosis, reducing manual workload and human error in the detection of diseases from medical images.

Furthermore, the Graphical User Interface (GUI) facilitates ease of use, allowing healthcare professionals to upload medical images, process them, and receive instant classification results. The system's ability to visualize confusion matrices and performance metrics aids in model interpretability, making it a valuable tool in clinical decision-making.

However ,there are several constraints that must be addressed to ensure the system's practical implementation and reliability. First, the performance of the system heavily depends on the quality and diversity of the dataset used for training and evaluation. A heterogeneous dataset with diverse patient demographics, including age, gender, and medical history, is critical to ensure robustness and prevent over fitting to specific patient populations. Additionally, the system assumes that medical images are adequately pre processed, which involves resolution standardization, noise removal, and normalization. Inconsistent input data can affect the model's performance, so an automated prep rocessing pipeline is crucial for handling these issues.

Given that deep learning models, particularly CNNs, require significant computational resources, deploying the system may require high-performance GPUs for both training and inference. This could impose a constraint on the deployment in resource-constrained environments, making cloud-based or distributed computing solutions necessary for large-scale implementation. The system should also optimize for real-time processing, especially when handling high-resolution images. Optimizing the model for low latency without sacrificing accuracy is essential for practical use in clinical settings.

While the CNN model offers improved accuracy, it remains a "black-box" approach, which can limit its interpretability. Incorporating explainability tools, such as Grad-CAM or LIME, would enhance trust in the model's predictions, particularly in high-stakes medical decision-making. Moreover, the system must be evaluated for its ability to generalize across unseen or out-of-distribution images, such as those from different hospitals or imaging devices. Rigorous testing and fine-tuning are necessary to mitigate performance degradation in these cases.

The system’s ability to scale across multiple medical domains, including radiology, dermatology, and pathology, will also be a challenge. It is critical to ensure that the system can accommodate various disease types and imaging modalities such as X-rays, MRIs, and CT scans. Moreover, the system should include continuous monitoring and retraining to maintain its relevance as new medical imaging techniques and diseases emerge. Furthermore, to facilitate widespread adoption, the system should be cost-effective, ensuring accessibility across healthcare settings with varying budgets.

By addressing these constraints, the proposed system has the potential to revolutionize the way medical diagnoses are made, offering a reliable, interpretable, and accessible tool for healthcare professionals, while also ensuring fairness, privacy, and ethical responsibility.

REFERENCES

[1] Meenalochini, G., and S. Ramkumar. "Survey of machine learning algorithms for breast cancer detection using mammogram images." Materials Today: Proceedings 37 (2021): 2738-2743.

[2] Darweesh, M. Saeed, Mostafa Adel, Ahmed Anwar, Omar Farag, Ahmed Kotb, Mohamed Adel, Ayman Tawfik, and Hassan Mostafa. "Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images." Cogent Engineering 8, no. 1 (2021): 196832.

[3] Alshammari, Maha M., Afnan Almuhanna, and Jamal Alhiyafi. "Mammography image-based diagnosis of breast cancer using machine learning: a pilot study." Sensors 22, no. 1 (2021): 203.

[4] Atrey, Kushangi, Bikesh Kumar Singh, and Narendra Kuber Bodhey. "Multimodal classification of breast cancer using feature level fusion of mammogram and ultrasound images in machine learning paradigm." Multimedia Tools and Applications (2023): 1-22.

[5] Avcı, Hanife, and Jale Karakaya. "A Novel Medical Image Enhancement Algorithm for Breast Cancer Detection on Mammography Images Using Machine Learning." Diagnostics 13, no. 3 (2023): 348.

[6] Abdulla, Srwa Hasan, Ali Makki Sagheer, and Hadi Veisi. "Breast cancer classification using machine learning techniques: A review." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 14 (2021): 1970-1979.

[7] Yedjou, Clement G., Solange S. Tchounwou, Richard A. Aló, Rashid Elhag, BereKet Mochona, and Lekan Latinwo. "Application of machine learning algorithms in breast cancer diagnosis and classification." International journal of science academic research 2, no. 1 (2021): 3081.

[8] de Miranda Almeida, Rhaylander Mendes, Dehua Chen, Agnaldo Lopes da Silva Filho, and Wladmir Cardoso Brandao. "Machine Learning Algorithms for Breast Cancer Detection in Mammography Images: A Comparative Study." In ICEIS (1), pp. 660-667. 2021.

[9] Safdar, Sadia, Muhammad Rizwan, Thippa Reddy Gadekallu, Abdul Rehman Javed, Mohammad Khalid Imam Rahmani, Khurram Jawad, and Surbhi Bhatia. "Bio-imaging-based machine learning algorithm for breast cancer detection." Diagnostics 12, no. 5 (2022): 1134.

[10] Jalloul, Reem, H. K. Chethan, and Ramez Alkhatib. "A review of machine learning

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

techniques for the classification and detection of breast cancer from medical images." Diagnostics 13, no. 14 (2023): 2460.