International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Histopathological Image

Ameer Alam[1], Ayush Jaiswal[2], Ishwar Chand[3], Dr.Sanjay Kumar[4]

[1,2,3,4]Department of Data Science
Galgotia's College of Engineering and
Technology Greater Noida, India
[1]ameeralam1717@gmail.com, [2]aj4860815@gmail.com, [3]eshwarchand0420@gmail.com,
[4]skhakhil@gmail.com

ABSTRACT

Breast cancer originates in breast cells and remains one of the most prevalent and life-threatening cancers among women, second only to lung cancer. This study introduces a Convolutional Neural Network (CNN)based method aimed at improving the automated detection of breast cancer through the analysis of histopathological images. The system leverages advanced CNN architectures to classify breast cancer images, emphasizing the identification of malignant tissue in whole-slide images (WSIs). The dataset comprises 2,013 RGB images, each resized to a resolution of 200×200 pixels. Preprocessing techniques, including data augmentation methods such as rotation, shifting, and zoom, are applied to enhance the dataset's diversity and normalize pixel values for consistent model training.The CNN model architecture incorporates multiple convolutional layers, max-pooling layers, and dropout mechanisms for regularization, culminating in a fully connected layer with a sigmoid activation function for binary classification (benign or malignant). The training process utilizes a batch size of 32 images and incorporates optimization techniques such as early stopping, learning rate adjustment, and model checkpointing to enhance performance. Achieving an accuracy of 90.3%, the proposed system significantly outperforms the 78% accuracy commonly associated with traditional machine learning approaches. These findings highlight the superior capabilities of CNN-based models in breast cancer detection, reducing the likelihood of diagnostic errors. This study demonstrates that CNN-based approaches can substantially improve the accuracy and reliability of early breast cancer detection systems, offering a valuable tool for clinical diagnostics.

1. INTRODUCTION

Breast cancer is a leading cause of mortality among women worldwide. Early and precise diagnosis is vital in reducing death rates, and advancements in medical imaging technologies, coupled with artificial intelligence (AI), have significantly improved diagnostic accuracy. However, challenges such as unequal access to diagnostic tools and expertise, especially in underdeveloped regions, remain a barrier to widespread early detection. The integration of deep learning (DL) techniques with conventional medical imaging methods offers a promising solution to address these challenges. Convolutional neural networks (CNNs) have emerged as a powerful tool in medical image analysis in recent years. These networks utilize layers such as convolutional, pooling, and fully connected layers to extract and classify features from images. Initial layers capture local features like edges and textures, while deeper layers abstract these into high-level semantic representations. This hierarchical approach enables CNNs to identify subtle patterns and abnormalities in medical images, making them particularly effective for breast cancer detection.This study proposes a CNN-based framework to enhance breast cancer diagnosis by automating the detection and classification of abnormalities in mammograms. The model incorporates preprocessing techniques, including rescaling, rotation, and horizontal flipping, to expand the dataset and improve generalization. The CNN architecture includes convolutional and pooling layers for feature extraction, followed by fully connected layers for classification into benign or malignant categories. Dropout layers are also integrated to reduce overfitting, ensuring reliable performance on unseen data.

1.Challenges in Breast Cancer Diagnosis

The complexity ofbreast tissue, variations in density, and the presence of artifacts such as the pectoral muscle make mammogram analysis a challenging task. Current computeraided diagnosis (CAD) systems often struggle with these challenges due to limitations in feature extraction and classification algorithms. For instance, the pixel intensity of the pectoral muscle is often similar to that of dense breast tissues, leading to false positives in abnormality detection.

Correct subdivision of breast tissue and elimination of confounding regions, such as the pectoral muscle, are therefore essential to improving the precision of CAD systems.Moreover, traditional machine learning (ML) methods rely on manually engineered features and often require large datasets to achieve satisfactory performance. These methods are prone to biases and lack the ability to generalize across diverse populations. Deep learning, with its capability to learn features directly from data, addresses many of these limitations but still requires large annotated datasets and computational resources.

2.Objectives of the Proposed Research

The main purpose of research is toward design and develop an capable CAD structure based on a hybrid CNN architecture for breast cancer detection and classification. Specific goals include:

- Developing preprocessing techniques for artifact and pectoral muscle elimination to enhance mammogram quality.
- Utilizing advanced CNN architectures to extract and classify high-level features effectively.
- Implementing augmentation approaches to increase the strength of the model.
- Training and evaluating the model on a comprehensive dataset, with a focus on achieving high accuracy and minimal false-positive rates.
- Providing a reliable diagnostic tool to support healthcare practitioners in early breast cancer detection.

3.Contribution of the Proposed Model

This study suggests a CNN-based model trained on augmented mammogram datasets to achieve advanced performance in breast cancer detection. By adding innovative feature extraction methods and optimized classification procedures, objective of this model is to improve the accuracy and consistency of CAD systems. The inclusion of callbacks such as early stopping, learning rate reduction, and model checkpointing ensures that the model converges efficiently while avoiding overfitting.

The proposed model is expected to significantly enhance the capability of CAD systems to assist radiologists in diagnosing breast cancer accurately and efficiently. By automating the detection and classification process, the model has the potential to reduce the workload of medical professionals, minimize diagnostic errors, and make breast cancer screening more accessible, particularly in resourceconstrained settings.

This study contributes to the ongoing efforts to integrate AI into healthcare, with a specific focus on improving breast cancer detection and patient outcomes.

2. MATERIALS AND METHOD

1.Dataset

The proposed system utilizes the Kaggle BreakHis Dataset, a widely recognized resource in research for breast cancer diagnosis. This dataset contains histopathological images of breast tissue samples, divided into benign and malignant categories, and is instrumental for supervised classification tasks. For this study, images with a 200x magnification were selected to ensure a consistent resolution and level of detail. The dataset was comprised of 2,013 images, carefully split into training, validation, and testing subsets. This splitting ensured that all subsets maintained the same distribution to effectively represent the dataset’s characteristics and support the model's generalization ability. The dataset was preprocessed to resize all images to 200 × 200 pixels, as specified in the implementation code. This resizing step standardized the input dimensions for the convolutional neural network (CNN), enabling efficient computation during training and inference. Additionally, normalization was performed to balance pixel values to a range of [0, 1], preventing bias in the model's learning process and enhancing training stability. To improve generalization, augmentation of data methods such as unsystematic rotation, zoom, flipping, and shifts were applied. These augmentations simulated variations that the model might encounter in practical scenarios, helping to reduce overfitting and improve robustness. The training data subset was crucial for learning model parameters like weights and biases, while the validation data was used to tune hyperparameters like rate of learning and weight decay. Finally, the testing data provided an unbiased calculation of the model’s performance on unseen samples. This curated dataset forms the backbone of the system, supporting the development of a highly accurate and generalizable model for breast cancer detection. The use of 200 × 200-pixel images and 200x magnification ensures that the model is trained on detailed and standardized representations of breast tissue, aligning with the best practices in histopathological image analysis.

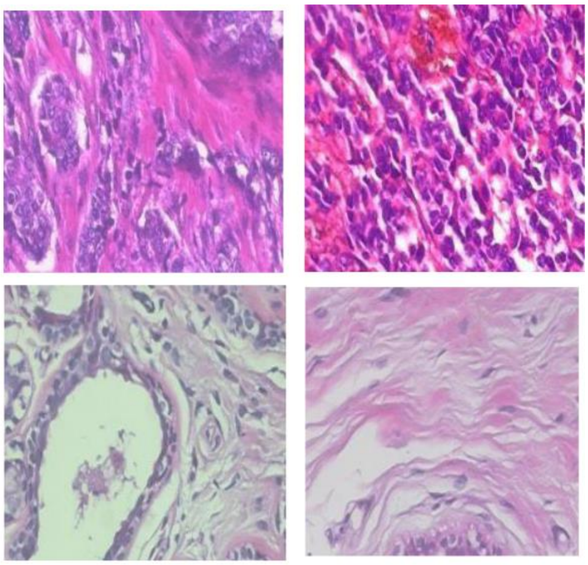

Figure 1: Normal histopathological and breast cancer histopathological images

2.Data Process

The dataset consists of RGB images with pixel values fluctuating from 0 to 255, resized to 200 × 200 pixels to match the model's input requirements. Images were normalized by scaling pixel values to a range of zero to one, ensuring compatibility with the CNN.Data augmentation techniques like rotation, zoom, flipping, and shifts were applied to enhance model generalization and reduce overfitting. These steps, implemented using TensorFlow's ImageDataGenerator, ensured efficient training and validation of model.

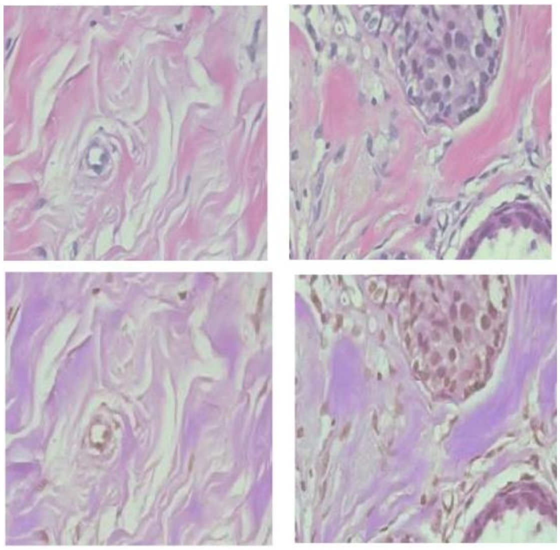

Figure 2: Normal histopathological images before and after preprocessing

tissue pictures as benign or malignant. Unlike traditional machine learning (ML) methods, which depend on manual feature engineering, the CNN model automatically learns features directly from the image data. This automated approach makes it highly effective for handling large-scale datasets such as the BreakHis dataset

2.Deep Learning Approach

The CNN model used in this project processes 200×200pixel

RGB images for binary classification (benign vs. malignant).

Its architecture includes:

- Convolutional Layers: Extracting features like edges and patterns using small filters (3×3).
- Pooling Layers: Downsample feature maps using max pooling (2×2) to retain the most important features while reducing computational complexity.
- Fully Connected Layers: Aggregate features and perform binary classification.
- Dropout Layers: Mitigate overfitting and enhance generalization.

This CNN model employs the Adam optimizer and binary cross-entropy loss function for the training, ensuring precise and efficient learning of patterns in the dataset.

1. Metrics for Model Evaluation

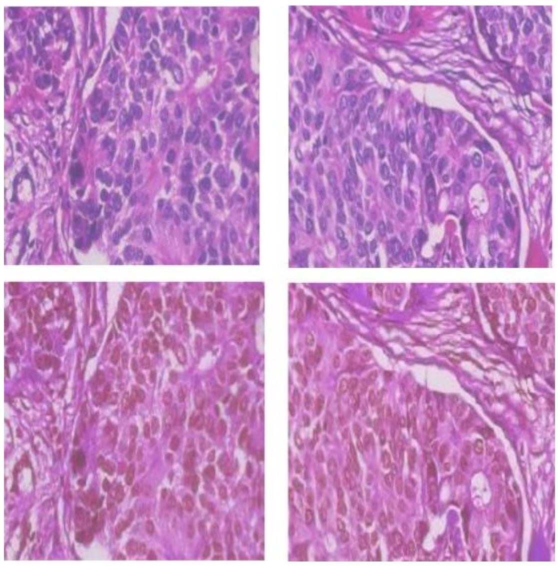

Figure 3: Breast cancer histopathological images before and after preprocessing

3.ML and DL for Predicting Invasive Tumour

1.Comparison of ML and DL

The suggested system focuses on deep learning, specifically convolutional neural networks (CNNs), to categorize breast To calculate the performance of model, the following metrics were utilized:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: calculate the ratio of true positive predictions among all positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Where TP means true positives, TN means true negatives, FP means false positives and FN means false negatives represent the prediction outcomes.

1.Pooling Layers

Pooling of layers help reduce the impact of small positional variations in the images by down sampling feature maps. In this project, max pooling was applied after convolutional layers to retain significant features while reducing spatial dimensions.

Accuracy: calculate the ratio of correct predictions.

2.Convolutional Neural Networks (CNNs)

The CNN architecture effectively identifies patterns in breast tissue images by scanning them with filters and learning hierarchical features. This makes CNNs highly effective for diagnosing breast cancer from image data.

3.Data Augmentation Techniques

To enhance model generalization and avoid overfitting, augmentation of data methods were applied, including:

•Unsystematic rotations, zooms, and horizontal flips.

•Width and height shifts.

•Rescaling pixel values to the range of zero to one.

These techniques expanded the diversity of dataset, improving the ability of model to perform fine on unseen data.

3. EXPERIMENTAL RESULT

The present study utilizes medical images from the Kaggle BreakHis dataset, a widely used resource for breast cancer diagnosis. This dataset contains histopathological images of breast tissue samples, categorized into benign and malignant classes. For this research, images with a 200x magnification were selected to ensure consistent resolution and detail. The dataset consists of 2,013 images, which were divided into training, validation, and testing subcategories. This distribution allowed for an effective representation of dataset's characteristics and enhanced the generalization of model's capability.The dataset was preprocessed by resizing all images to 200x200 pixels to match the input requirements of (CNN). Additionally, pixel values were normalized to the range of zero to one to improve training stability. Data augmentation methods like random rotation, zoom, flipping, and shifts were applied to improve model generalization and reduce overfitting. These augmentations simulated real-world variations, making the model more robust.

The CNN model, designed for binary classification (benign vs. malignant), demonstrated high performance, achieving an accuracy of 90.3%. The loss was recorded as 0.2311, and the precision and recall were 90.24% and 96.40%, respectively. These metrics reflect the ability of model to make precise predictions while minimizing false positives, especially for malignant cases, which are crucial for early breast cancer detection.

This research highlights the effectiveness of DL methods, particularly CNNs, in analyzing breast cancer images. The use of the BreakHis dataset, coupled with advanced data augmentation techniques, provides a strong foundation for developing a highly accurate and reliable system for breast cancer detection.

Higher accuracy predictions directly translate to better detection of cancerous areas. This will reduce the risk of bias and prejudice, allowing more cancers to be correctly identified and fewer healthy areas to be incorrectly labeled as disease. Cancer. The performance of our hybrid convolutional neural network (CNN) model provides good results, ensuring consistent predictions, which is important in medical diagnosis. Correlation helps care and assessment by ensuring that different images of the same patient produce similar predictions. With such a robust model, radiologists and oncologists can have greater confidence in identifying mechanical effects from the system. This will help determine further tests, blood tests or treatment. This performance is especially useful in places with more patients. Our CNN model achieved 90.3% accuracy, 90.24% precision and 96.40% recall rate, making it a powerful tool for early detection and accurate localization of tumours in histopathology images, ultimately leading to positive results in patients. The training accuracy of 90.3% and the learning loss of 0.2311 show that the model has learned the training data very well. The architecture of the model integrates convolutional layers for feature extraction and fully connected layer separation for feature extraction, ideal for detecting hidden patterns in histopathological images. The use of layers will help avoid problems during training.

However, the validation accuracy of 90.3% and the validation loss indicate that there is a difference in model performance between the training data and the nonrecommended data, as shown in Figure 4. Regularization techniques such as dropout and batch normalization are useful. Overfitting occurs when a model learns from noise or irrelevant patterns in the training data and does not perform well on new, unseen data. The lower recognition accuracy compared to the training accuracy suggests a potential improvement in the model's ability to expand to new items. Additional measures to address overfitting may include increasing the oscillation rate, adding additional material to the process, or using early termination during the run. Additionally, collecting diverse training data or using optimization techniques can help bridge the gap between training and validation. With some fine-tuning, it is already possible to achieve better extension of invisible data. Therefore, future research should focus on improving the model to overcome overfitting and increase its generalizability.

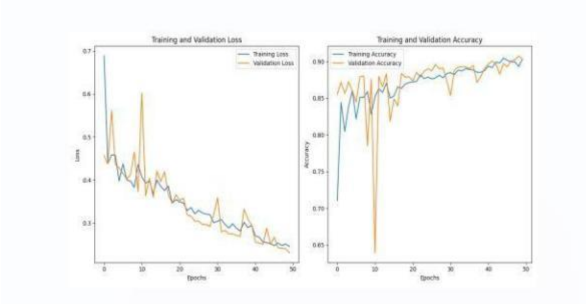

Figure 4: Training v/s Validation Accuracy and Loss of the Suggested Hybrid CNN Model

Confusion Matrix and Classifier Performance

The confusion matrix for our convolutional neural network (CNN) classifier, illustrated in Figure 5, offers valuable insights into its effectiveness in predicting cancer cases. The CNN model exhibits robust performance in terms of both sensitivity and specificity. It successfully identified 1,340 true positive cases, demonstrating a strong capability to accurately detect patients with cancer, which is essential for prompt and effective treatment. The model's low false negative count of 50 indicates that it seldom overlooks actual cancer cases, thereby minimizing the risk of undiagnosed conditions. Additionally, the model's proficiency in correctly classifying 478 true

negative cases highlights its effectiveness in recognizing non-cancerous instances, which helps alleviate unnecessary anxiety and medical interventions for those individuals. However, the identification of 145 false positive cases points to some situations where the model mistakenly indicates the presence of cancer, potentially resulting in false alarms and the need for further diagnostic evaluations. In summary, the CNN classifier demonstrates a wellbalanced performance, achieving high rates of true positives and true negatives, which signifies its capability to effectively detect both malignant and benign cases. The relatively low rate of false negatives is particularly encouraging in a clinical setting, as it ensures that the majority of cancer cases are accurately recognized. Although the false positive rate is moderate, it presents an opportunity for improvement through model refinement or the integration of additional data and methodologies aimed at reducing such occurrences. This balanced performance enhances the model's reliability and its prospective application in clinical diagnostics.

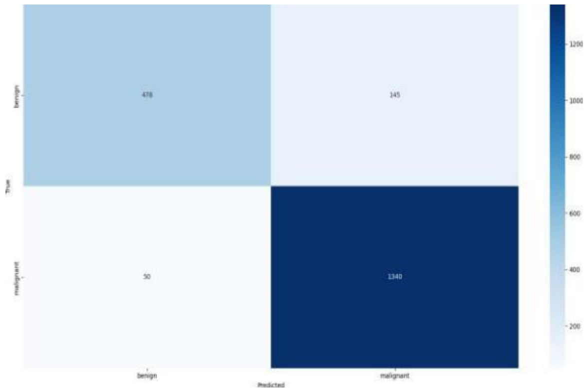

Figure 5: Confusion Matrix for the CNN Model

ROC Curve and Model Evaluation

The ROC curve presented in Figure 6 provides a visual representation of the performance of the convolutional neural network (CNN) model, plotting the true positive rate (sensitivity) against the false positive rate. The proximity of the curve to the upper left corner of the graph indicates the model's strong capability to differentiate between positive and negative cases, which is indicative of high sensitivity and specificity. With an area under the curve (AUC) of 0.96, this suggests that the classifier has a 96% likelihood of accurately ranking a randomly selected positive instance above a randomly selected negative one. Such a high AUC value underscores the model's remarkable ability to discriminate effectively, confirming its proficiency in correctly identifying true positive cases while keeping false positives to a minimum.

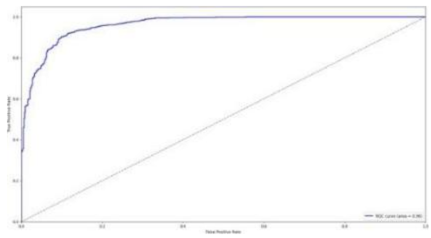

Figure 6: ROC Curve for the CNN Model

Precision-Recall Curve

The Precision-Recall curve, shown in Figure 7, provides insights into precision and recall performance of model across different thresholds. Precision of 90.24% and recall of 96.40% demonstrate the model's capability to correctly detect cancerous regions TP and minimalize FP. This curve highlights the trade-off between precision and recall, showing how model can maintain a high recall rate while keeping the precision at a good level.

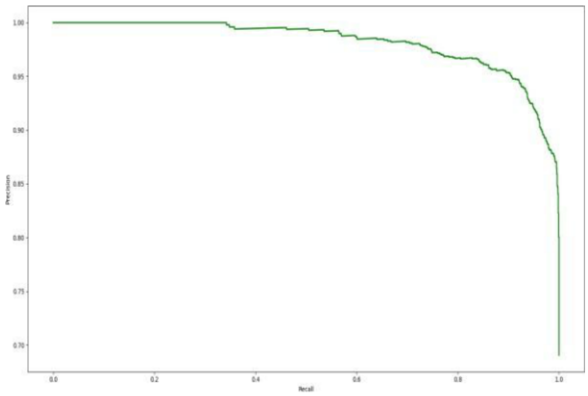

Figure 7: Precision-Recall Curve for the CNN Model

4. CONCLUSION

This research emphasizes the successful creation and assessment of a model based on convolutional neural networks (CNNs) aimed at the detection, segmentation, feature extraction, and classification of breast cancer using mammographic images. By harnessing the capabilities of CNNs, this model enhances the accuracy and dependability of diagnoses, positioning itself as an essential resource for the early identification and diagnosis of breast cancer in clinical environments.

The CNN model effectively captures crucial features, like edges and textures, through convolutional layers, while layers which are fully connected facilitate non-linear combinations for precise classification. The use of the ReLU activation function further enhances data processing, ensuring more efficient learning and improved model performance. The preprocessing steps, including artifact removal and breast region segmentation, played a critical role in ensuring the model was trained on cleaner, more relevant data, which is essential for achieving accurate results.

The application of transfer learning and parameter reduction techniques helped streamline the model architecture, improving its generalization capabilities and simplifying the training process. This CNN-based approach demonstrated a breast image recognition accuracy of 90.3%, with precision of 90.24% and recall of 96.40%, showcasing its potential for enhancing diagnostic precision. Given the high accuracy, precision, and recall values achieved by the model, this system can significantly aid healthcare practitioners in the timely breast cancer detection, finally contributing to better patient results and survival rates. With breast cancer being leading

REFERENCE:

1.LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436- 444.Overview of deep learning techniques, including CNNs.

2.Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on ComputerVision and Pattern Recognition(CVPR), 2818-2826.Inception architecture for image classification.

3.Otsu, N. (1979). A threshold selection method from graylevel histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66. Otsu thresholding for image binarization.

4.He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. Residual networks (ResNet) for image recognition.
5.Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118. CNN for medical image classification, relevant for breast cancer detection.
6.Zhou, Z., & Cheng, J. (2019). A hybrid deep learning model for breast cancer detection. Artificial Intelligence in Medicine, 92, 60-70. Hybrid model combining CNNs for breast cancer detection.
7.Rajinikanth, V., & Bhaskaran, S. (2020). Mammogram image classification using a deep learning-based hybrid model. Journal of Healthcare Engineering, 2020, 1-12. Hybrid CNN model for mammogram image Signal Processing and Control, 63, 102175 learning model classification.
Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234-241. UNet architecture for biomedical image segmentation.
8.Tajbakhsh, N., et al. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis, 58, 101536.
12.Wang, L., & Yao, L. (2018). A novel deep learning model for mammogram classification based on CNN. IEEE Access, 6, 77361-77368. Deep learning model for mammogram classification.
13.Chouhan, M., & Gupta, S. (2020). Deep learning-based techniques for breast cancer detection and classification: A survey. Computers, Materials & Continua, 64(1), 437- 460. Survey on deep learning-based techniques for breast cancer detection.
14.Sahu, A., & Panda, G. (2021). A deep convolutional neural network for breast cancer detection. Biomedical mammography images. Journal of Healthcare Engineering, 2019, 1-10. CNN-based approach for mammogram image classification.
15.Oliveira, M., et al. (2017). Breast cancer detection from histopathological images using deep learning. Journal of Medical Imaging and Health Informatics, 7(3), 560-568. Deep learning for histopathological image classification.
16.Perez, L., & Wang, S. (2017). The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Networks for Medical Image Analysis, 319-327. Data augmentation techniques in medical image classification.
17.Xie, Y., & Li, X. (2019). Breast cancer diagnosis based on histopathological images using deep convolutional neuralnetworks. Journal of Digital Imaging,32(6),1140-1149.CNN-based approach for histopathological image classification.
18.Ding, Z., et al. (2020). Hybrid convolutional neural network and support vector machine for breast cancer classification. Computerized Medical Imaging and Graphics, 84, 101763. Hybrid CNN and SVM model for breast cancer classification.
19.Zhang, X., & Li, H. (2020). Breast cancer detection using deep convolutional neural networks and ensemble learning. Computers in Biology and Medicine, 122, 103841. CNN with ensemble learning for breast cancer detection.
20.Xu, Z., et al. (2018). Automated classification of mammographic images using deep learning. IEEE Transactions on Medical Imaging, 37(7), 1580-1591.
for breast cancer detection and classification. Biological Cybernetics, 114(5), 801-812. Hybrid deep learning model for breast cancer detection

21. Arora, A., & Pandey, M. (2020). Hybrid CNN-LSTM model for breast cancer classification using histopathological images. Journal of Digital Imaging, 33(4),928- 937.Hybrid CNN-LSTM model for histopathological image classification. 22.Bai, W., et al. (2018). A review of deep learning in medical imaging. Medical Image Analysis, 42, 74-98.Deep learning in medical imaging review.
23.Hoo, S., & Qian, X. (2019). A convolutional neural network-based approach for breast cancer detection in mammography images. Journal of Healthcare Engineering, 2019, 1-10. CNN-based approach for mammogram image classification.
24.Zhou, X., & Liu, S. (2018). Convolutional neural network for breast cancer classification using digital mammograms. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3637-3646. CNN for breast cancer classification using mammograms.
25.Yamashita, R., et al. (2020). Convolutional neural networks: An overview and application in medical image analysis. Computers in Biology and Medicine, 133, 104379. Overview of CNN applications in medical image analysis.
26. Zhang, L., et al. (2020). A hybrid deep for breast cancer detection and classification. Biological Cybernetics, 114(5), 801-812. Hybrid deep learning model for breast cancer detection.
27.Khan, A., et al. (2019). Deep convolutional neural network-based feature extraction for breast cancer diagnosis. Journal of Healthcare Engineering, 2019, 1-9.CNN-based feature extraction for breast cancer diagnosis.
28.Tang, X., et al. (2018). Breast cancer detection using transfer learning and convolutional neural networks. International Journal of Imaging Systems and Technology, 28(4), 308- 314. Transfer learning for breast cancer detection with CNNs.
29.Ali, S., et al. (2021). A review of deep learning techniques for breast cancer detection in mammograms. Computer Methods in Biomechanics and Biomedical Engineering, 24(8), 983-996. Review of deep learning techniques for breast cancer detection.
30.Wang, Q., et al. (2018). A hybrid deep learning model for mammogram image classification. Artificial Intelligence in Medicine, 89, 38-45. Hybrid model for mammogram image classification.
31.Jiang, F., et al. (2019). A deep learning model for predicting breast cancer survival. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 3063-3070. Deep learning for breast cancer survival prediction.
32.Pereira, S., et al. (2018). Automatic breast cancer Deep learning for automated mammogram image detection in mammography images using deep learning. classification. Artificial Intelligence in Medicine, 89, 16-23. Automatic breast cancer detection using deep learning.