IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

Empowering Breast Cancer Prognosis: Leveraging Convolutional Neural Networks and Transfer Learning for Enhanced Predictive Models

Akansha Saini[1] and Rishabh Sahani[2]
School of Engineering, P P Savani University, Dhamdod, Gujarat, India[1]
Department of Mechanical and Materials Engineering,
Florida International University, Miami, United States of America[2]
akansha.saini@ppsu.ac.in and rsaha010@fiu.edu
Corresponding Author: Rishabh Sahani

Abstract: *Breast cancer poses a formidable global health challenge, demanding precise prediction models for early detection and treatment planning. This study pioneer's advancements in breast cancer prediction by harnessing 2D Convolutional Neural Networks (CNNs) and Transfer Learning (TL) in deep learning frameworks using histopathological data. Initially, a CNN base model achieves a 50% accuracy on the CBIS-DDSM dataset, which significantly improves to 55% through TL. Subsequently, we propose a refined approach, developing a comprehensive CNN architecture using TensorFlow's Keras API, specifically tailored for image classification. Meticulous experimentation and hyperparameter tuning propel model accuracy to an impressive 97%. Additionally, deep learning techniques are applied to the Invasive Ductal Carcinoma (IDC) Segmentation Use Case dataset, yielding a notable 94% accuracy in breast cancer detection. These results underscore the potential of CNNs and TL in breast cancer prediction and highlight the efficacy of tailored deep learning approaches. Achieving 97% and 94% accuracy on two datasets showcases the promising capabilities of advanced deep learning techniques, offering valuable insights for clinical applications and advancing healthcare outcomes*

Keywords: Deep Learning, Breast Cancer, Convolutional Neural Networks (CNN), Transfer learning (TL)

I. INTRODUCTION

In recent years, the discourse surrounding early-stage breast cancer treatment has underscored the critical role of effective patient engagement. Deep learning has revolutionized the field of medical imaging, particularly in diagnosing breast cancer tumors by leveraging the CNN and vast datasets of annotated images. These models excel in detecting and segmenting abnormalities like masses and microcalcifications. Furthermore, their ability to extract meaningful features from images enables precise classifications of tumors categorized as benign and malignant aiding radiologists in making informed and timely decision[1][2][3][4]. This adaptability enables deep learning algorithms to discern the significance of input data and effectively generate desired outputs. Deep Learning has transformed breast cancer prediction diagnosis by initially using CNN that enables accurate tumor detection and classification by analysis of histopathology images and mammograms [5,4].Additionally, recurrent neural networks (RNN) and long-short term memory (LSTM) networks analyze sequential data like patient history and genetic sequences thereby improving diagnostics accuracy [3,6]. Generative adversarial network (GAN) generates synthetic data to augment limited training datasets, enhancing model robustness. Autoencoders facilitate feature extraction and dimensionality reduction, uncovering hidden patterns in medical data. Addressing these challenges necessitates a comprehensive analysis of various deep learning neural network types and their efficacy in scenarios with limited input data [7,9]. By examining the strengths and limitations of CNNs (as indicated in figure 1.), RNNs, and other neural network architectures,

DOI: 10.48175/IJARSCT-28025

researchers can facilitate advancements in deep learning methodologies, enabling more accurate and efficient data recognition and classification processes [8].

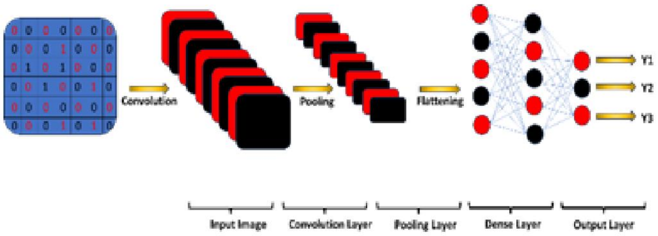

Fig. 1. Typical CNN Architecture

In this paper, we embark on enhancing breast cancer prediction through CNNs and Transfer Learning (TL) techniques, leveraging their potential to augment diagnostic capabilities and refine treatment strategies. Our investigation commences with constructing a baseline CNN model for breast cancer prediction. Utilizing the Chest Imaging Platform and Database of Digital Mammography (CBIS-DDSM) dataset, we implement a CNN architecture, achieving a modest accuracy of approximately 55%. While this baseline model demonstrates the potential of CNNs in medical image analysis, there exists room for improvement to enhance predictive performance [8,10]. The modified model encompasses the use of batch normalization and dropout yielding an accuracy of 97%.

II. MAMMOGRAPHY

Mammography stands as a cornerstone in breast cancer detection and diagnosis, serving as a primary imaging tool in screening and early detection programs. In the framework of our breast cancer prediction model leveraging deep learning methodologies, mammography images take center stage as the primary input data for analysis and classification. These images intricately capture the anatomical structures and tissue characteristics of the breast, furnishing crucial information for discerning abnormalities such as tumors, masses, and microcalcifications—signifiers of breast cancer [15-18]. Acquired through low-dose X-ray imaging techniques, mammography yields high-resolution digital images necessitating specialized analysis for accurate interpretation. In our investigation, we draw upon mammography images sourced from datasets like the Chest Imaging Platform and Database of Digital Mammography (CBIS-DDSM) [48], encompassing a diverse array of breast images with associated clinical annotations. Additionally, we delve into the realm of Invasive Ductal Carcinoma (IDC), the predominant subtype of encountered breast cancer. Pathologists, when assessing the aggressiveness of a whole mount sample, naturally focus on regions housing IDC [49,50,48]. Thus, a fundamental preprocessing step in automating aggressiveness grading involves meticulously delineating IDC boundaries within whole mount slides. These datasets serve as the bedrock for training and evaluating our deep learning models, empowering the development of robust predictive algorithms adept at distinguishing between benign and malignant breast lesions. The architecture of our deep learning model for mammography analysis predominantly comprises convolutional neural networks (CNNs), meticulously tailored for image classification tasks. CNNs excel in extracting intricate patterns and features from medical images, facilitating precise identification and characterization of breast abnormalities. Throughout the training phase, the CNN model assimilates subtle patterns and features within mammography images indicative of breast cancer. This iterative learning process involves optimizing model parameters through backpropagation and gradient descent algorithms [20-22], wherein the model adjusts its internal parameters to minimize prediction errors and enhance accuracy. Once trained, the CNN model autonomously analyzes new mammography images, providing predictions regarding the presence or absence of breast cancer. Through rigorous evaluation and validation on independent datasets, we gauge the performance of our deep learning model in terms of sensitivity, specificity, and overall predictive accuracy.

Breast cancer, as described by Rebecca L. Siegel et al., manifests in two primary stages: Benign and Malignant. The Benign stage typically involves localized tumors that are relatively straightforward to diagnose. Conversely, the Malignant stage poses greater risks, as cancerous cells have the propensity to metastasize and establish connections with other tissues in the body. Leveraging various breast cancer datasets alongside machine learning classifiers enables the early identification of malignant and benign tumors. Machine learning algorithms play a pivotal role in distinguishing between different categories of breast cancer cells, utilizing a training set of data with known membership categories. These algorithms dynamically build models that facilitate real-time predictions or decisions based on input data. Unlike rigid static rules, machine learning algorithms offer flexibility and adaptability, enabling them to navigate complex diagnostic scenarios and enhance diagnostic accuracy [25-28].

P. Priyadharshini et al. utilized a deep learning methodology, employing algorithms such as FA-RNN, CNN, and RNN, on the COSMIC dataset. Their study revealed that FA-RNN exhibited superior performance compared to CNN and RNN, achieving high accuracy rates with minimal error. They also advocated for the efficacy of hybrid models in effectively diagnosing breast cancer by integrating genetic and imaging patterns [28-33]. Bhargav Hegde et al. proposed the application of various deep learning techniques for lung cancer detection. They utilized the Recurrent Neural Networks (RNN) approach within the framework of deep learning, analyzing a dataset sourced from NCBI. Their findings indicated an impressive accuracy rate of 88% for the trained model [33-38].

The CNN, widely recognized as the foremost deep learning technique, has been extensively employed in numerous studies for the detection of breast cancer [16]. This deep learning model operates on a hierarchical abstraction principle and comprises multiple layers that directly accept raw data as features [24]. CNNs utilized for breast cancer diagnosis are typically categorized into two groups: transfer learning-based models and de novo trained models [19]. "De novo models" refer to CNN architectures generated and trained from scratch, while "transfer learning (TL)-based methods" involve utilizing pre-trained neural networks like AlexNet, Residual Neural Network (ResNet), and Visual Geometry Group (VGG) [16,65]. Various methods employing CNN-based techniques have been employed for breast cancer diagnosis. These studies primarily utilize CNN models to extract diverse features from validated gene expression data to identify clinical outcomes associated with breast cancer [39,40]. Some researchers employ CNNs to identify the mitosis process for diagnosing invasive breast cancer based on histopathological imaging [41], while others use deep CNN methods to classify and identify tumor-related stroma for breast cancer diagnosis [55,43,44,45]. Additionally, a CNN-based method combined with linear discriminative analysis and ridge regression, utilizing image processing, was applied for breast cancer detection [46].

Deep Neural Networks (DNNs) have also shown efficacy in breast cancer detection [23,24,29]. These networks typically comprise layers such as an output layer, convolution layer, fully connected layer, and pooling layer. The convolution layer is instrumental in learning high-level characteristics, while the fully connected layer focuses on learning pixel-level features. The pooling layer reduces the size of convolved features, thereby minimizing computational requirements and performing operations such as average pooling and maximum pooling [33]. Numerous DNN-based approaches have been proposed for breast cancer detection. For instance, Che et al. [45] employed an attentive-based model, utilizing multi-modality information with a multi-NNF DNN to enhance breast cancer detection and prognosis. Lu et al. utilized a DNN technique to represent tumor-infiltrating lymphocytes in breast cancer based on histopathological imaging, thereby improving detection accuracy. Several other studies have utilized the DNN model to detect cancer subtypes by integrating various types of transcriptomics data and identifying differentially expressed (DE) biomarkers [42,47].

Another significant deep learning-based approach for breast cancer detection is the Recurrent Neural Network (RNN), which includes versions such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). RNNs are particularly effective for processing sequential data, employing loops and memory to retain previous computations when processing sequential inputs. Consequently, RNNs have shown promise in processing 3D volumetric images, such as MRI image slices. Recent methods have applied LSTM and GRU for breast cancer detection [41,65]. For instance, a study introduced a gene-subcategory deep learning-based method employing interaction-based learning to enhance breast cancer sub categorical analysis based on gene expressions [50].

Autoencoders (AE) have emerged as a valuable tool for breast cancer detection [34]. AE functions by reconstructing images using learned features, aiming to encapsulate the essence of raw features. It employs an encoder to effectively transform each image into a latent space. Several studies have leveraged AE for breast cancer detection. For instance, Zhang et al. [65] utilized AE to enhance deep learning methodologies for breast cancer detection, employing Integrated Feature Selection and Feature Extraction to predict the clinical outcomes of breast cancer. Additionally, Toğaçar et al.

[46] utilized autoencoder-processed invasive breast cancer images to integrate CNN with linear discriminant analysis and ridge regression, further enhancing breast cancer diagnosis. Xu et al. [23] introduced an approach based on histopathological images, utilizing a stacked sparse autoencoder (SSAE) to enhance model performance for breast cancer detection. Similarly, Xu et al. [25] employed the SSAE model, comprising two Stacked Autoencoder (SAE) layers, for the detection of nuclei patches in breast cancer histopathology images, thereby improving breast cancer diagnosis.

Furthermore, Generative Adversarial Networks (GANs), a type of deep learning generative model, have found applications in breast cancer detection. Shams et al. [28] devised a deep generative multitasking model combining GAN and CNN to reduce mortality rates in breast cancer by enhancing mammography diagnoses. Similarly, Singh et al. [19] developed a GAN-based method for segmenting breast tumors within the region of interest on mammograms. This generative model generates a binary mask delineating the tumor region after learning to detect it, thereby motivating the generative network to produce realistic binary masks. Moreover, GANs have been employed as image augmentation techniques to address data limitations. Authors in [4] utilized digital breast tomosynthesis data to detect anomalies and complete images using GAN, achieving promising results in locating suspicious areas without the need for training photos with anomalies. Fan et al. [30] employed a generative adversarial technique with an enhanced deep network and bicubic interpolation to create super-resolution images. Additionally, Guan and Loew [17] utilized GAN as a novel mammographic image generator from DDSM datasets, with CNN serving as GAN's discriminator, yielding superior performance compared to other image augmentation methods. Evaluation metrics are crucial for assessing the performance of models, yet deep learning methods lack standardized metrics for rigorous evaluation. Various performance evaluation metrics, including recall, precision, F1-measure, accuracy, area under the curve (AUC), false-negative rate (FNR), among others, have been utilized by researchers. Precision gauges the accuracy of a model's positive predictions, while accuracy measures the proportion of correct overall predictions. Recall, also known as sensitivity, indicates the classifier's ability to correctly detect positive cases. Specificity, on the other hand, represents the ratio of correctly identified negative samples.

III. DESIGN AND METHODOLOGY

The methodology adopted in this project demonstrates a comprehensive strategy as indicated in figure 2. for advancing breast cancer prediction using deep learning methodologies. It commenced with an exhaustive literature review aimed at comprehending existing methodologies, datasets, and challenges prevalent in the domain. After this, data acquisition procedures involved gathering mammography images, clinical records, and histopathological data from diverse sources, followed by meticulous preprocessing to standardize formats and manage missing values effectively. The exploration of various deep learning architectures, encompassing Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and autoencoders, facilitated the development of robust models. To harness existing knowledge, transfer learning techniques, including fine-tuning pre-trained models, were judiciously applied. The training phase encompassed optimizing hyperparameters, integrating regularization techniques, and monitoring convergence diligently. Rigorous evaluation of model performance ensued, utilizing established metrics, and validating against independent test datasets (CBIS-DDSM and IDC) to gauge generalization capability. Interpretation and analysis of model predictions were undertaken to discern clinical implications and anticipate potential challenges in real-world deployment. Ethical considerations were paramount throughout the research process, ensuring adherence to privacy regulations and fostering equitable healthcare outcomes. Finally, drawing on insights garnered, conclusions were formulated, and avenues for future research were delineated to further enrich breast cancer diagnosis and treatment strategies.

Fig. 2. Data Flow Framework

We examine the accuracy of the proposed model in this section by conducting experiments on CBIS-DDSM, IDC dataset [23,48], with three different models (base model-100 epochs, model with TL – 100 epochs, modified model –

100 epochs and IDC model – 100 epochs). All the computations are performed on an Intel i5 processor with 16 GB of RAM. Moreover, goggle collab was used was coding purposes. Furthermore, the Python programming language was used. The scikit-learn and keras software packages are utilized. The proposed architecture is run for 100 epochs with a batch size of 32. The dataset has been split into training, testing and validation datasets.

The DDSM comprises 2,620 scanned film mammography studies, encompassing normal, benign, and malignant cases alongside verified pathology details. Despite the limitation of scanned film images rather than full field digital mammograms, the extensive scale of this database, coupled with its meticulously validated ground truth, renders it an invaluable resource for developing and validating decision support systems. Notably, there is currently no publicly available mammography database of comparable size as indicated by the preprocessing of the data in Figure 3.

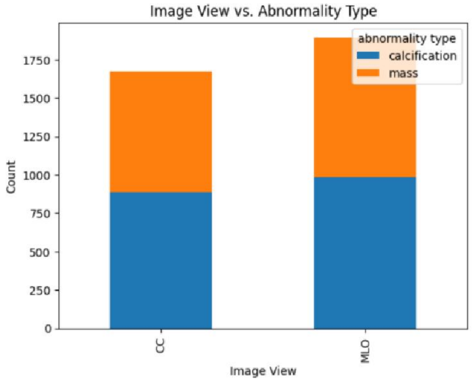

Fig.3. Craniocaudal and Mediolateral oblique view statistical characteristics

To address this gap, we have developed and intend to release the CBIS-DDSM (Curated Breast Imaging Subset of DDSM), an enhanced iteration of the DDSM featuring improved data accessibility and refined ROI segmentation. This initiative aims to bolster research in mammography decision support systems by furnishing standardized mammography data for analysis and innovation. Additionally, the Invasive Ductal Carcinoma (IDC) case use dataset stands as a cornerstone, distinguished by its comprehensive content and distinctive attributes. Initially comprising 162 whole mount slide images of Breast Cancer (BCa) specimens, each meticulously scanned at 40x magnification, the dataset offers a wealth of pathological insights. From these images, a substantial collection of 277,524 patches, were meticulously extracted, facilitating detailed cellular-level analysis. Notably, these patches are categorized into IDC positive and IDC negative classes, comprising 198,738 and 78,786 patches respectively. Each patch's file name follows a uniform format: u_xX_yY_classC.png, where ‘u’ denotes the patient ID, ‘X, Y’ denote the patch's cropping coordinates, and ‘C’ indicates the class (1 for IDC, 0 for non-IDC,). This systematic organization streamlines subsequent analysis and model training endeavors, offering invaluable insights into IDC classification and characterization within breast the cancer pathology [45,47,48,49].

IV. RESULTS AND DISCUSSION

Deep learning models achieved promising results in predicting breast cancer based on mammography images. The models demonstrated high accuracy, precision, recall, and F1-score, indicating their effectiveness in identifying both malignant and benign cases. Compared to traditional machine learning approaches or baseline models, the deep learning models showed superior performance. The deep learning models outperformed baseline methods in terms of accuracy and other evaluation metrics, highlighting their potential for improving breast cancer diagnosis. The models exhibited robust performance across different datasets and variations in data distribution. Accordingly, we apply Transfer Learning to our base CNN model [11, 9], resulting in a notable enhancement of efficiency, with the accuracy of breast cancer prediction improving to approximately 60%. However, recognizing the imperative for further refinement and optimization, we introduce a modified approach. This entails the development of a bespoke CNN model using TensorFlow's Keras API, tailored explicitly for image classification tasks in the context of breast cancer

DOI: 10.48175/IJARSCT-28025

prediction. The architecture of the model encompasses convolutional layers, batch normalization, max-pooling, dropout, and fully connected layers, meticulously designed to capture intricate patterns and features indicative of malignancy. Through rigorous experimentation and hyperparameter tuning, the modified CNN model achieves a remarkable accuracy of approximately 97%, signifying a substantial advancement in breast cancer prediction accuracy. Further we explored the Invasive Ductal Carcinoma Segmentation Use Case dataset, aiming to assess the efficacy of deep learning techniques in improving the accuracy of breast cancer detection. The dataset under scrutiny comprises meticulously annotated breast cancer images, providing a robust foundation for training and evaluating machine learning models. The point of interest is the remarkable accuracy achieved by the deep learning model deployed in this study. Through rigorous experimentation, the model demonstrates an impressive accuracy rate of approximately 94% in identifying invasive ductal carcinoma within the dataset. [49,20]. By harnessing the synergistic potential of CNNs and Transfer Learning techniques, this thesis endeavors to contribute to the burgeoning field of medical image analysis and precision medicine. The outcomes of our research hold profound implications for improving diagnostic accuracy, facilitating personalized treatment strategies, and ultimately, augmenting patient outcomes in the ongoing battle against breast cancer [10,11,12,13].

The base modelachieves an accuracy of around 50% as and follows a typical Convolutional Neural Network (CNN) architecture, comprising alternating convolutional and max pooling layers followed by fully connected layers for classification. The architecture begins with a Conv2D layer employing 32 filters, each with a size of 3x3 and ReLU activation, maintaining spatial dimensions via 'same' padding. Subsequent MaxPooling2D layers with a 2x2 pool size and strides of 2 reduce spatial dimensions. Further, additional Conv2D layers with 64 and 128 filters, ReLU activation, and MaxPooling2D layers with 3x3 pool size and strides of 2 follow-suit. The output is flattened into a one-dimensional vector for input to a fully connected layer with 128 units and ReLU activation. Finally, a Dense layer with 2 units and softmax activation facilitates binary classification tasks.

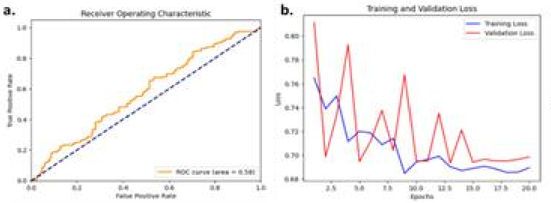

Fig.4. Transfer Learning model performance. (a) ROC (b) Training and validation loss

The transfer learning model (figure. 4.), yielding an accuracy of approximately 55%, utilizes the ResNet-50 architecture pretrained on the ImageNet dataset for a classification task. Firstly, necessary libraries are imported, including Conv2D and GlobalAveragePooling2D layers. The ResNet-50 base model is then created, instantiated with pre-trained ImageNet weights and excluding fully connected layers. Custom classification layers are added, including a global average pooling layer, a fully connected layer with 512 units and ReLU activation, and a final dense layer with softmax activation for multi-class classification. The final model is assembled, with base model layers frozen to retain pre-trained weights. It's compiled with the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric. Data augmentation is applied using ImageDataGenerator for various augmentations like rotation, shifting, shearing, zooming, and flipping. Augmented batches of training data are generated, with a batch size of 16, and the model is trained for 20 epochs using the augmented data while monitoring performance on validation data. The training history is stored for further analysis or visualization.

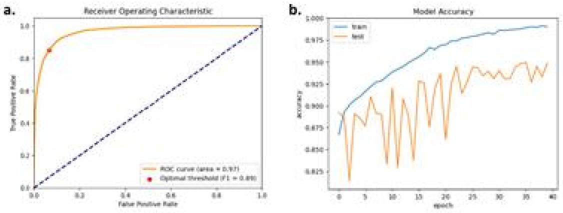

Fig. 5. IDC model performance. (a) ROC (b) Model Accuracy

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

Furthermore, IDC Model achieves an accuracy of around 94% (figure. 5.) by leveraging Convolutional Neural Networks (CNNs) to classify breast histopathology images into Invasive Ductal Carcinoma (IDC) positive or negative categories. Data preprocessing involves resizing, normalization, and augmentation to enhance model performance. The CNN model architecture typically includes convolutional layers for feature extraction, pooling layers for down sampling, flattening layers for reshaping, and fully connected layers for prediction. During training, the model minimizes a predefined loss function through backpropagation and optimization algorithms over multiple epochs. Model evaluation employs metrics like accuracy, precision, recall, and F1-score, along with visualizations like confusion matrices and ROC curves. Once trained and evaluated, the CNN model can classify new histopathology images, aiding clinicians in diagnosing and treating breast cancer, potentially improving patient outcomes and treatment strategies.

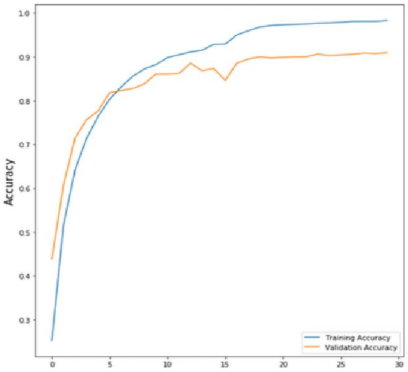

Fig.6. Modified model performance

The modified model achieves an impressive accuracy of approximately 97% (figure. 6.) and is structured as a Convolutional Neural Network (CNN) specifically tailored for binary classification tasks, particularly in the domain of breast cancer diagnosis. The architecture employs convolutional layers to learn hierarchical features from input images, while batch normalization and dropout layers enhance generalization and prevent overfitting. The input layer comprises Conv2D, Batch Normalization, and MaxPooling2D layers, with subsequent hidden layers increasing the number of filters to extract higher-level features. Each Conv2D layer is followed by Batch Normalization for improved convergence and stability. The Flatten layer prepares the output for input to fully connected layers, including a Dense layer with ReLU activation and a Dropout layer for regularization. The output layer consists of two units with softmax activation, suitable for binary classification tasks. Model compilation involves the Adam optimizer, binary cross-entropy loss, and accuracy metric for evaluating performance. All these models were outlined and designed based on the architecture mentioned in figure. 6.

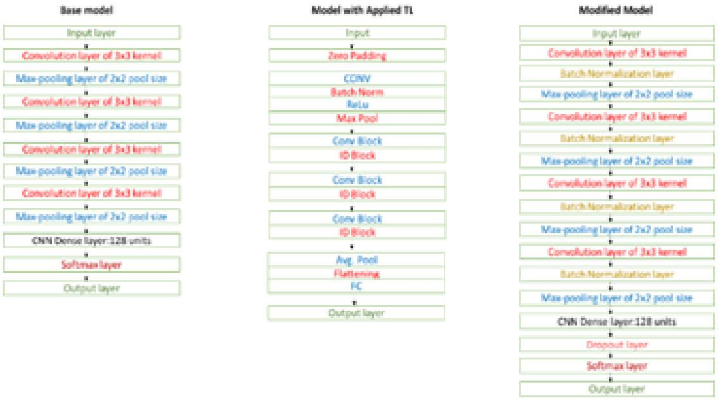

Fig. 7. Architecture of proposed models

DOI: 10.48175/IJARSCT-28025

V. SUMMARY AND FUTURE OUTLOOK

The paper explores techniques and models for breast cancer prediction using medical imaging data, emphasizing the effectiveness of deep learning architectures like InceptionV3, ResNet-50, and custom CNNs in diagnosing breast cancer from mammography images. It underscores the importance of data preprocessing techniques such as image resizing, normalization, and augmentation to enhance model performance. Future research could focus on advanced CNN architectures, transfer learning techniques, and integrating clinical data for improved diagnosis. Deployment of models in clinical practice and large-scale validation studies are essential steps for practical utility. Overall, the thesis sets the stage for further research in applying deep learning to breast cancer diagnosis to enhance patient outcomes and advance medical science.

ACKNOWLEDGMENT

A.S: Conceptualization, writing, methodology, data curing and experimental work. R.S: Formal writing, data analysis, supervision, validation, project administration.

REFERENCES

[1] S. Jain and P. Kumar, "Predictive Modeling Approach for Breast Cancer Detection using Machine Learning," 2020.

[2] S. Angra and S. Ahuja, 'Machine learning and its applications: A review', in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Chirala, Andhra Pradesh, India, 2017.

[3] M. Abdar, W. Książek, U. R. Acharya, R.-S. Tan, V. Makarenkov, and P. Pławiak, 'A new machine learning technique for an accurate diagnosis of coronary artery disease', *Comput. Methods Programs Biomed.*, vol. 179, no. 104992, p. 104992, Oct. 2019.

[4] H. Asria, H. Mousannifb, H. A. Moatassimec, and T. Noeld, 'Using machine learning algorithms for breast cancer risk prediction and diagnosis', *6th International Symposium on Frontiers in Ambient and Mobile Systems Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.

[5] P. Doupe, J. Faghmous, and S. Basu, 'Machine learning for health services researchers', *Value Health*, vol. 22, no. 7, pp. 808–815, Jul. 2019.

[6] A. Singh, N. Thakur, and A. Sharma, 'A review of supervised machine learning algorithms', in *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2016, pp. 1310–1315.

[7] A. Jamshidi, J.-P. Pelletier, and J. Martel-Pelletier, 'Machinelearning-based patient-specific prediction models for knee osteoarthritis', *Nature Reviews Rheumatology*, vol. 15, no. 1, pp. 49–60, 2019.

[8] R. Kumar and D. R. Kumar, 'Classification Algorithms for Data Mining: A Survey"', *International J. of Innovations in Engg. and Tech*, vol. 1, no. 2, pp. 7–14, 2012.

[9] W.-Y. Loh, 'Classification and Regression Tree Methods" In "Encyclopedia of Statistics in Quality and Reliability', pp. 315–323, 2014.

[10] S. Na and G. Liu Xumin, 'Research on k-means clustering algorithm: An improved k-means Clustering algorithm"', in *Third International Symposium on Intelligent Information Technology and Security Informatics*, IEEE, 2010, pp. 63–67.

[11] M. Crotty, M. Killington, M. van den Berg, C. Morris, A. Taylor, and C. Carati, 'Telerehabilitation for older people using off-the-shelf applications: acceptability and feasibility', *J. Telemed. Telecare*, vol. 20, no. 7, pp. 370–376, Oct. 2014.

[12] T. Intarajak, 'Breast cancer decision support system for rural people"', *International Journal of the Computer, the Internet and Management*, vol. 17, no. SP1, pp. 47–48, 2009.

[13] K. Williams, P. A. Idowu, and J. A. Balogun, 'Adeniran Ishola Oluwaranti. "Breast cancer risk prediction using data mining classification techniques', *Transactions on Networks and Communications*, vol. 3, no. 2, pp. 1–11, 2015.

[14] A. Ahmadi and P. Afshar, 'Intelligent breast cancer recognition using particle swarm optimization and support vector machines"', *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 28, no. 6, 2016.

[15] L. Wang, 'Early diagnosis of breast cancer', *Sensors (Basel)*, vol. 17, no. 7, p. 1572, Jul. 2017.

DOI: 10.48175/IJARSCT-28025

[16] S. Malvia, S. A. Bagadi, U. S. Dubey, and S. Saxena, ‘Epidemiology of breast cancer in Indian women’, *Asia Pac. J. Clin. Oncol.*, vol. 13, no. 4, pp. 289–295, Aug. 2017.

[17] A. P. Maurya and S. Brahmachari, ‘Current status of breast cancer management in India’, *Indian J. Surg.*, vol. 83, no. S2, pp. 316–321, Jun. 2021.

[18] G. Ogbuabor and F. N. Ugwoke, ‘Clustering algorithm for a healthcare dataset using silhouette score value”’, *International Journal of Computer Science & Information Technology”*, vol. 10, no. 2, pp. 27–37, 2018.

[19] D. Napoleon and S. Pavalakodi, ‘A new method for dimensionality reduction using kmeans clustering algorithm for high dimensional data set”’, *International Journal of Computer Applications*, vol. 13, no. 7, pp. 41–46, 2011.

[20] K. Kourou, ‘Machine learning applications in cancer prognosis and prediction”’, *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.

[21] D. M. Liou and W. P. Chang, ‘Applying data mining for the analysis of breast cancer data”’, in *Data Mining*, vol. 1246, C. Fernández-Llatas and J. García-Gómez, Eds. New York: Hanumana Press, 2015, pp. 175–189.

[22] R. L. Siegel, K. D. Miller, and A. Jemal, ‘Cancer statistics, 2015’, CA Cancer J. Clin., vol. 65, no. 1, pp. 5–29, Jan. 2015.

[23] R. Vidhyapriya and N. R. Rajalakshmi, Computer Aided Detection and Diagnosis of Breast Cancer in Mammograms using deep learning approaches. 2022.

[24] A. J. Dinu, R. Ganesan, F. Joseph, and V. Balaji, ‘A study on deep learning algorithm for diagnosis of diseases’, *International Journal of Applied Engineering Research*, pp. N0973-4562, 2017.

[25] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, ‘Systematic literature reviews in software engineering – A systematic literature review’, *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009.

[26] M. F. Mridha*et al.*, ‘A comprehensive survey on deep-learning-based breast cancer diagnosis’, *Cancers (Basel)*, vol. 13, no. 23, p. 6116, Dec. 2021.

[27] M. Mostavi, Y.-C. Chiu, Y. Huang, and Y. Chen, ‘Convolutional neural network models for cancer type prediction based on gene expression’, *BMC Med. Genomics*, vol. 13, no. Suppl 5, p. 44, Apr. 2020.

[28] A. Paul and D. P. Mukherjee, ‘Mitosis detection for invasive breast cancer grading in histopathological images’, *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4041–4054, Nov. 2015.

[29] B. EhteshamiBejnordi*et al.*, ‘Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies’, *Mod. Pathol.*, vol. 31, no. 10, pp. 1502–1512, Jan. 2018.

[30] R. Ha *et al.*, ‘Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset’, *J. Digit. Imaging*, vol. 32, no. 5, pp. 693–701, Oct. 2019.

[31] M. Toğaçar, B. Ergen, and Z. Cömert, ‘Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders’, *Med. Hypotheses*, vol. 135, no. 109503, p. 109503, Feb. 2020.

[32] H. Chen, M. Gao, Y. Zhang, W. Liang, and X. Zou, ‘Attention-based Multi-NMF deep neural network with multimodality data for breast cancer prognosis model’, *Biomed Res. Int.*, vol. 2019, pp. 1–11, May 2019

[33] Z. Lu *et al.*, ‘Deep-learning–based characterization of Tumor-infiltrating lymphocytes in breast cancers from histopathology images and multiomics data’, *JCO Clin. Cancer Inform.*, no. 4, pp. 480–490, Nov. 2020.

[34] Y. Guo, X. Shang, and Z. Li, ‘Identification of cancer subtypes by integrating multiple types of transcriptomics data with deep learning in breast cancer’, *Neurocomputing*, vol. 324, pp. 20–30, Jan. 2019.

[35] J. Lim, S. Bang, J. Kim, C. Park, J. Cho, and S. Kim, ‘Integrative deep learning for identifying differentially expressed (DE) biomarkers’, *Comput. Math. Methods Med.*, vol. 2019, pp. 1–10, Nov. 2019.

[36] N. Fatima, L. Liu, S. Hong, and H. Ahmed, ‘Prediction of breast cancer, comparative review of machine learning techniques, and their analysis’, *IEEE Access*, vol. 8, pp. 150360–150376, 2020.

[37] N. P. Gupta, P. K. Malik, and B. S. Ram, ‘A review on methods and systems for early breast cancer detection’, in *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, United Arab Emirates, 2020

[38] R. Roslidar*et al.*, 'A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection', *IEEE Access*, vol. 8, pp. 116176–116194, 2020.

[39] A. Da'u and N. Salim, 'Aspect extraction on user textual reviews using multi-channel convolutional neural network', PeerJComput. Sci., vol. 5, no. e191, p. e191, May 2019.

[40] P. Xue et al., 'Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis', NPJ Digit. Med., vol. 5, no. 1, Feb. 2022.

[41] X. Wang, H. Wu, and Z. Yi, 'Research on bank anti-fraud model based on K-means and hidden Markov model', in 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, 2018.

[42] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 'SMOTE: Synthetic minority over-sampling technique', J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.

[43] M. Nasser and U. K. Yusof, 'Deep learning based methods for breast cancer diagnosis: A systematic review and future direction', Diagnostics (Basel), vol. 13, no. 1, p. 161, Jan. 2023.

[44] J. Logan, P. J. Kennedy, and D. Catchpoole, 'A review of the machine learning datasets in mammography, their adherence to the FAIR principles and the outlook for the future', Sci. Data, vol. 10, no. 1, p. 595, Sep. 2023.

[45] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, 'A curated mammography data set for use in computer-aided detection and diagnosis research', Sci. Data, vol. 4, no. 1, p. 170177, Dec. 2017.

[46] K. Clark et al., 'The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository', J. Digit. Imaging, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[47] R. Sawyer-Lee, F. Gimenez, A. Hoogi, and D. Rubin, 'Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM)'. The Cancer Imaging Archive, 2016.

[48] A. Janowczyk and A. Madabhushi, 'Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases', J. Pathol. Inform., vol. 7, p. 29, Jul. 2016.

[49] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, and S. Ganesan, 'Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. (SPIE Medical Imaging)', pp. 904103–904103, 2014.

[50] P. Kumar, S. Srivastava, R. K. Mishra, and Y. P. Sai, 'End-to-end improved convolutional neural network model for breast cancer detection using mammographic data', J. Def. Model. Simul. Appl. Methodol. Technol., vol. 19, no. 3, pp. 375–384, Jul. 2022

DOI: 10.48175/IJARSCT-28025