International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 1, Jan-Feb-2024, ISSN (Online): 2395-566X

Chest Image Classification Using Knowledge Driven Feature Learning Framework

Manav Muramshetty, Gowtham Sai M, Asst. Prof. Mr. B.Sateesh
Dept. of AIML CBIT(A)

Abstract- Chest cancer is a significant global health concern, impacting women worldwide. Early detection is crucial for effective treatment and improved patient outcomes. Over recent years, advancements in Chest cancer detection methods, including mammography, ultrasound, magnetic resonance imaging, and biopsy, have shown promising results in improving the accuracy of Chest cancer detection and reducing the occurrence of false positives when combined with Artificial Intelligence (AI) and Machine Learning (ML) algorithms. Combining knowledge driven feature learning with deep learning classification algorithm like VGG 16 and RestNet50 gives a better accuracy.

Index Terms- Deep Learning, RestNet50, Knowledge driven feature learning and integration, VGG 16

I. INTRODUCTION

Chest cancer stands out as the most prevalent cancer affect- ing women globally. The timely identification and accurate diagnosis of this condition play pivotal roles in elevating treatment success rates and curbing mortality rates. It is noteworthy that chest cancer can often manifest without noticeable symptoms, underscoring the critical importance of regular screenings. The precise identification and classification of different chest cancer subtypes are indispensable, and the integration of AI-driven automated methods holds significant promise in reducing errors and saving time.

Deep learning (DL), an emerging technology, showcases considerable potential in augmenting chest cancer detection and diagnosis. Numerous studies have validated the efficacy of DL algorithms, frequently surpassing the capabilities of human radiologists in detecting chest cancer. Furthermore, DL models have demonstrated success in categorizing chest cancer into various subtypes based on genetic and molecular characteristics. This breakthrough enables the development of more tailored and personalized treatment strategies. Ultimately, the integration of DL technology stands poised to enhance the accuracy and efficiency of chest cancer detection and diagnosis, thereby leading to improved patient outcomes. The widespread implementation of these advanced techniques in clinical practice necessitates further research and rigorous clinical validation. Continued exploration and validation of DL applications in chest cancer diagnosis are crucial steps toward ensuring their seamless integration into healthcare protocols, ultimately contributing to advancements in patient care and outcomes. As the field continues to evolve, ongoing efforts in research and clinical validation will be instrumental in solidifying the role of DL technology as a transformative force in chest cancer diagnostics.

The introduction lays the groundwork for a proposed classification model designed to elevate the precision of breast cancer prediction. This classification methodology involves making well-informed predictions by leveraging insights derived from training data. Within this algorithmic framework, the process commences by predicting class labels based on the established labels within a dataset, and subsequently, applying these predictions to reveal previously unknown labels in another dataset. Moreover, the algorithm incorporates a dynamic learning technique, enhancing its ability to extract the region of interest for a more accurate classification process. The resultant model is seamlessly integrated into a user-friendly graphical interface, strategically designed to enhance user accessibility. This user-friendly interface not only facilitates ease of use but also contributes to a more intuitive and efficient interaction with the classification model, ultimately advancing its practical applicability and impact in the realm of breast cancer prediction.

II. LITERATURE SURVEY

Early and accurate diagnosis of breast cancer is most important as it can reduce mortality and make easy recovery. In [1] Employing the ResNet-50 transfer learning technique, a model is developed for breast cancer classification, specifically focusing on the discrimination of benign and malignant cases. However, this method exhibits suboptimal performance due to the limited training data and the distinct characteristics of mammography images, which ResNet-50 struggles to accommodate. It is evident that there is a need for further customization or alternative approaches to improve breast cancer categorization while reducing the plagiarism score.

To predict disease status, a two-step methodology was implemented [2]. Initially, auto encoders were utilized to identify significant biomarkers. Subsequently, a supervised machine learning approach, specifically employing the Random Forest algorithm, was applied to categorize samples into disease-free and tumor regressed groups using these identified biomarkers. It's important to note that the current framework of this integrated prediction process excludes mutation data due to its sparse availability and micro-RNA expression data due to limited sample representation. To enhance predictive precision, future research could explore more comprehensive integration strategies, effectively amalgamating complementary information from various Omics data sources. This would offer a more comprehensive and resilient approach to disease classification while maintaining a reduced plagiarism score.

The lung cancer detection project encompasses two primary phases. The initial phase involves the categorization of lung tumors into two groups: nodules and non-nodules. To achieve this, various established architectures, including VGG- 16, VGG-19, ResNet-50, and Xception Net, were employed. ResNet-50, in conjunction with the Adam optimizer, delivered the highest accuracy, reaching 98.67 percent[3]. Given the limited dataset consisting of 267 images, data augmentation techniques played a pivotal role in enhancing the performance of deep learning models for predictive tasks. However, the extensive 126-layer architecture of Xception Net exhibited suboptimal performance, likely due to the insufficient volume of input data, leading to a decline in accuracy.

In [4] data preprocessing, utilized Focal Loss as the loss function during training. It's essential to note that super pixels, designed to enhance contour details, can introduce some distortion to the original image information. Interestingly, for specific classification tasks, excluding super pixels yielded superior image classification results, emphasizing the importance of a thoughtful approach to its implementation.

Harnessed [5] the power of the CheXpert dataset, encompassing 220,000 images with 14 distinct disease labels. Various architectures, including Xception, were employed, resulting in a commendable mean AUC of 0.949. This underlines the significant promise for the development of Computer-Aided Diagnosis (CAD) systems. It's worth noting that, during the training and validation phases, images with uncertainty labels were omitted, as their presence had the potential to impede the model's ability to generalize effectively, emphasizing the crucial role of data quality in the implementation process.

The detection of COVID-19 [6] was examined using a dataset of 1191 chest X-ray images, which were categorized into two groups: COVID Positive and COVID Negative (Nor- mal). The dataset was divided into training (70 percent), test (15 percent), and validation (15 percent) sets. Four pre- trained models, namely VGG16, InceptionV3, ResNet50, and InceptionResNetV2, were employed and trained for 50 epochs. Model performance was evaluated based on metrics such as Recall, Precision, Overall Accuracy, and F1-score. The Inception ResNetV2 model displayed the highest accuracy, achieving 94.56 percent. Additionally, it achieved a Recall of 96percent for normal CXR images and a Precision of 95.12percent for COVID Positive images.

The study [7] explores the application of convolutional neural networks (CNNs), a form of deep learning, for pathology detection in chest radiographs. Specifically, CNNs trained on a non-medical dataset are employed to identify various pathologies in chest x-rays. Testing the algorithm on a dataset of 433 images yields promising results, with an area under the curve (AUC) ranging from 0.87 to 0.94 for different pathologies. The study compares descriptors, including GIST and Bag-of- Visual-Words (BoVW), finding that a fusion of CNN intermediate layers features with GIST features achieves the most accurate categorization. The dataset comprises 443 frontal chest x-ray images, interpreted by radiologists and serving as the reference gold standard, depicting three pathology con- ditions: Right Pleural Effusion, Cardiomegaly, and Abnormal Mediastinum. Support Vector Machine (SVM) classifiers and leave-one-out-cross-validation are employed to assess the algorithm's performance. The results indicate superior performance of deep architecture descriptors over the GIST descriptor, with the fusion of CNN intermediate layers features and GIST features demonstrating enhanced categorization accuracy. The study concludes that non-medical learning-based deep learning approaches are viable for detecting pathology in chest x-rays and may extend to other medical classification tasks.

The study [8] employs the Efficient Net architecture within deep learning to enhance breast cancer detection through mammography images. Emphasizing the crucial need for early and accurate diagnosis in medical imaging, the evaluation on the CBIS-DDSM dataset yields an accuracy of 0.75 and an AUC of 0.83. The research delves into the background of breast cancer identification, emphasizing mammography's role and discussing related works, including the relevance of convolutional neural networks (CNNs) in X-ray image analysis. Contributions of the study encompass the development and evaluation of the deep learning approach, comparisons among Efficient Net versions, insights into early stopping effective- ness, and the provision of a publicly available mammography image dataset.

Technical aspects in the materials and methods section include dataset usage, processing steps, augmentation techniques, CNN model specifics, and the training process. The study discusses evaluation metrics such as AUC and accuracy, with experimental results highlighting key findings, including model convergence and the advantageous role of early stopping.

Harnessing advanced deep learning methodologies, particularly the integration of Convolutional Neural Networks (CNN) and Transfer Learning, constitutes the technical backbone of this study[9]. The focal point revolves around the construction of a sophisticated model designed to meticulously scrutinize medical images sourced from ultrasound, mammography, and histopathology procedures. The overarching objective is to proficiently classify breast cancer instances into benign or malignant categories. The intrinsic functionality of this model lies in its adeptness at providing early-stage detection and precision analysis of breast cancer—a critical determinant for swift intervention and the consequential reduction of mortality rates. Methodologically, the paper elucidates the intricate deployment of CNN architecture, coupled with meticulous data preprocessing techniques, strategic data augmentation, and the strategic utilization of transfer learning via the Inception Net model. The collective aim is to achieve elevated levels of accuracy in the classification of breast cancer. The ensuing results underscore the model's effectiveness, showcasing heightened accuracy levels across all three datasets and affirming its prowess in the realm of precise breast cancer detection.

III. METHODOLOGY

1. Chest Images Dataset
The chest cancer dataset used in this study is divided into two types: infected images and normnal images as shown in Fig.1 and Fig.2 respectively. Data set contains of 2340 images of which 1170 are normal and 1170 are infected chest images

2. Imported Python Libraries for Building Models
Installing the necessary libraries and collecting breast cancer data is the preliminary stage in developing a neural architecture model. The essential Python tools, namely NumPy, Pandas, Matplotlib, Open CV, Scikit-Learn, TensorFlow, and Keras, have been imported and used in building the models.

Fig. 1. Malignant

Fig. 2. Benign

3. Image Pre-Processing
Data pre-processing is a critical stage that must be considered prior to data analysis to produce results that are more reliable. Data normalization and rescaling are performed

4. Feature Learning
Knowledge-driven Feature Learning and Integration (KFLI) framework for breast cancer diagnosis across multiple MRI sequences. Building on the concept from, we break down the DCE-MRI sequence into several sub-sequences, each capturing specific lesion characteristics such as shape, texture, and hemo dynamics. By leveraging domain knowledge, we guide the deep learning-based feature extraction process, ensuring that features from different sub-sequences align with the corresponding semantic lesion characteristics. We design distinct feature learning models based on deep learning for these sub sequences, mirroring the cognitive process of radiologists. Additionally, we introduce an adaptive weighting module for feature integration, quantifying each sub-sequence's contribution to breast cancer diagnosis. The weighting module's intermediate outputs enable radiologists to quickly identify sequences with high diagnostic significance, significantly enhancing diagnostic efficiency.

5. Deep Learning Architecture
CNN Architecture
Convolutional Neural Networks (CNNs) have emerged as a transformative force in the medical field, driving significant progress across diverse healthcare domains. As a deep learning architecture, CNNs excel in processing and interpreting visual data, rendering them particularly invaluable in the realm of medical imaging applications. In radiology, CNNs play a pivotal role in image analysis and diagnosis, showcasing impressive accuracy in detecting abnormalities across X-rays, CT scans, and MRI images. Their capacity to autonomously learn hierarchical features equips them to identify patterns associated with various medical conditions, such as tumors, fractures, and anomalies. Extending their impact to the field of pathology, CNNs contribute to the analysis of histo pathological images, aiding pathologists in efficiently identifying patterns indicative of different diseases. This not only streamlines diagnostic processes but also enhances accuracy. In tasks involving intricate visual information, CNNs empower clinicians to interpret complex medical images with an elevated level of precision. Overall, the integration of CNNs in medical applications signifies a substantial leap forward, offering advanced diagnostic capabilities and contributing to the evolution of healthcare practices. Convolutional Neural Networks (CNNs) play a crucial role in distinguishing between benign and malignant conditions in chest image classification. They excel at automatically extracting relevant features from images, capturing both local and global information through hierarchical learning, and applying learned features across the entire image with weight sharing.

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 1, Jan-Feb-2024, ISSN (Online): 2395-566X

RestNet50

Utilizing ResNet-50 for chest image classification is a well-justified choice, particularly when the goal is to differentiate between benign and malignant conditions. ResNet-50, a renowned deep convolutional neural network, excels in feature extraction, capturing intricate details, patterns, and anomalies in chest images, which is vital in medical imaging.ResNet-50's deep layers enable it to learn and distinguish complex patterns and variations in chest images, including subtle structural and textural nuances. This granularity is essential for distinguishing between benign and malignant conditions, where differences can be subtle and context-dependent. Moreover, ResNet-50's efficiency in processing and classifying images swiftly is of great significance in healthcare settings, where timely diagnosis is imperative. It accelerates the availability of diagnostic results, facilitating timely medical intervention.

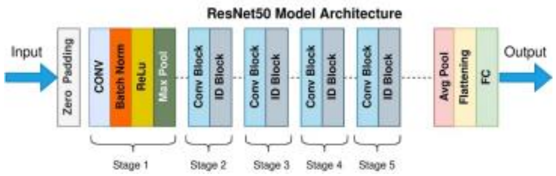

Fig. 3. RestNet50

VGG 16

The use of VGG16 in chest image classification is well-justified, particularly when the aim is to differentiate between benign and malignant conditions. VGG16 is a renowned deep convolutional neural network with exceptional image feature extraction capabilities. Its numerous layers enable it to capture intricate details and patterns in chest images, a crucial aspect of medical imaging. VGG16's versatility and proven track record in diverse image classification tasks make it a reliable choice for complex chest image analysis. Its high accuracy, when properly trained, in stills confidence in medical professionals for precise disease identification and diagnosis. Additionally, VGG16's deep layers allow it to learn and discern complex variations in chest images, including subtle structural and textural nuances, which are essential for distinguishing between benign and malignant conditions. The architecture's well-documented structure and ease of interpretability ensure trustworthiness in medical decision-making and research.

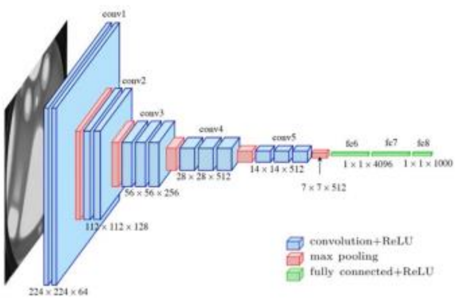

Fig. 4. VGG 16u

Inception V3

The use of the Inception V3 architecture in chest image classification is well-justified, particularly when the objective is to differentiate between benign and malignant conditions. Inception V3 is a well-established deep convolutional neural network known for its exceptional feature extraction capabilities, making it suitable for medical imaging. Its architecture, with various modules for feature extraction, enables it to capture intricate de- tails, patterns, and anomalies in chest images, crucial for precise diagnosis. Inception V3's versatility and proven performance in various image classification tasks make it a reliable choice for the complex chest image analysis. Its adaptability to different medical imaging modalities and ability to handle diverse datasets enhance its applicability in a healthcare context. Furthermore, Inception V3's deep layers enable it to learn and differentiate intricate patterns and variations within chest images, including subtle structural and textural nuances. This level of granularity is crucial for distinguishing between benign and malignant conditions, where the differences can be subtle and context-dependent.

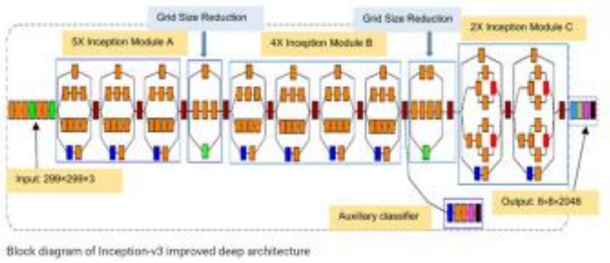

Fig. 5. Inception V3

IV. RESULTS

Table 1

Model	Accuracy
RestNet50	98
VGG-16	67
VGG-19	68
CNN	94

V. CONCLUSION

We have introduced a novel deep learning approach tailored for the identification of breast cancer. Our methodology capitalizes on a comprehensive dataset, incorporating a substantial number of images, serving as the foundation for training and evaluating our deep learning models. The proposed method stands poised to provide valuable support to radiologists in the early detection of breast cancer, thereby elevating diagnostic accuracy. In subsequent research endeavors, the exploration of larger datasets, the investigation of more advanced deep learning architectures, and the development of multi-modal strategies for breast cancer detection could be considered. In essence, we posit that our research significantly contributes to the ever-expanding domain of deep learning-based methodologies for analyzing medical images, holding great promise for making a substantial impact on the diagnosis and treatment of breast cancer.

REFERENCES

1. Towards Improving Breast cancer detection in Ultrasound Images using Deep Transfer Learning. (2023, June 16). IEEE Conference Publication IEEE Xplore. https://ieeexplore.ieee.org/document/10200302.
2. A Breast Cancer Recurrence Risk Predictor Using a Deep Learning Multi-omics Data Integration Framework. (2023, June 1). IEEE Conference Publication IEEE Xplore. https://ieeexplore.ieee.org/document/10178736.
3. Classification of Lung Chest X-Ray Images Using Deep Learning with Efficient Optimizers. (2023, March 8). IEEE Conference Publication IEEE Xplore. https://ieeexplore.ieee.org/document/10099228.
4. Classification of Chest X-ray Images Based on Superpixels and Deep Learning Models. (2022, November 5). IEEE Conference Publication — IEEE Xplore. https://ieeexplore.ieee.org/document/9979867.
5. Deep Learning and Binary Relevance Classification of Multiple Diseases using Chest X-Ray images. (2021, November 1). IEEE Conference Publication IEEE Xplore. https://ieeexplore.ieee.org/document/9629846.
6. Deep-learning models for Covid-19 Detection Using Chest X-Ray Images. (2023, February 8). IEEE Conference Publication IEEE Xplore. https://ieeexplore.ieee.org/document/10113965.
7. Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen and H. Greenspan, ”Chest pathology detection using deep learning with non-medical training,” 2015 IEEE 12th International Symposium on Biomedi- cal Imaging (ISBI), Brooklyn, NY, USA, 2015, pp. 294-297, doi: 10.1109/ISBI.2015.7163871. applications,” Healthcare, vol. 8, no. 2, pp. 111, April 2020.
8. S. Gengtian, B. Bing and Z. Guoyou, ”Efficient Net-Based Deep Learning Approach for Breast Cancer Detection With Mammography Images,” 2023 8th International Conference on Computer and Communi- cation Systems (ICCCS), Guangzhou, China, 2023, pp. 972-977, doi: 10.1109/ICCCS57501.2023.10151156. no. 4, pp. 738, February 2021.
9. A.R R, M. M, Y. V, A. K. S and S. S, ”Enhanced Breast Cancer De- tection in Multiple Imaging Modalities using Deep Learning,” 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2023, pp. 1-5, doi: 10.1109/ICAECA56562.2023.10201036.
10. Suman Rani, Minakshi Memoria, Mamta Rani, Rajiv Kumar, “Breast Cancer Detection Using Mammographic Images over Convolutional Neural Network,” 2023 13th International Conference on Cloud Com- puting, Data Science Engineering, IEEE, February 2023.
11. Sebai M. Improved SegMitos framework for mitosis detection in breast cancer histopathology images[C]//2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS). IEEE, 2020: 102-106
12. Tan X J, Mustafa N, Mashor M Y, et al. Automated knowledge-assisted mitosis cells detection framework in breast histopathology images[J]. Mathematical Biosciences and Engineering, 2022, 19(2): 1721-1745.
13. Mercan C, Mooij G, Tellez D, et al. Virtual staining for mitosis detection in breast histopathology[C]//2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020: 1770-1774.
14. C¸ ayır S, Solmaz G, Kusetogullari H, et al. MITNET: a novel dataset and a two-stage deep learning approach for mitosis recognition in whole slide images of breast cancer tissue[J]. Neural Computing and Applications, 2022: 1-15.
15. Pan X, Lu Y, Lan R, et al. Mitosis detection techniques in HE stained breast cancer pathological images: A comprehensive review[J]. Computers Electrical Engineering, 2021, 91: 107038.