AI-Driven Breast and Lung Cancer Detection

Mrs. Harshitha B K[1], M Ronith[2], Karthik N Hebbar[2], Rakshith Srinivasan[2], Shravan Karthik[2]

[1] Assistant Professor, Department of Information Science and Engineering, Dayananda Sagar Academy ofTechnology and Management, Bengaluru, India.
[2] B.E. Students, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, India.

Abstract

Cancer, notably breast and lung cancer, continuesto pose a global health threat, emphasizing the need for early detection strategies to improve survival rates. This literature survey focuses on AI-driven early detection of breast and lung cancer through deep learning and Convolutional Neural Network(CNN) models. A novel automated system for breast cancer detection is proposed, employing techniques like convolutional neural networks, sparse autoencoder, and stacked sparse autoen- coder. Comparative analyses underscore the superiority of the stacked sparse autoencoder, presenting a robust and accurate non-invasive detection solution. Shifting to lung cancer, early diagnosis is crucial, prompting a critical assessment of existing computer-aided diagnosis techniques. A new model for lung cancer detection is proposed to enhance accuracy. With lungcancer being a leading global cause of mortality, the review explores advanced technologies in computed tomography (CT) screening. The analysis aids radiologists in precise assessments, addressing challenges posed by vast CT image information.

Index Terms—*Breast cancer, Lung cancer Machine learning, Deep learning, Convolutional neural network, EfficientNetB0.*

I. INTRODUCTION

Breast cancer and lung cancer stand as formidable challenges in the realm of healthcare, with far-reaching impacts on global mortality rates. According to the Centresfor Disease Control and Prevention (CDC), breast cancer ranks as the most prevalent form of cancer among women. Thesurvival outcomes for individuals afflicted with breast cancer vary significantly, hinging on factors such as cancer type and stage at diagnosis. Breast cancer originates in the cells of the breast, commonly in the lobules or ducts, potentially invading adjacent healthy tissues and lymph nodes. Abnormal cell growth is the catalyst for this malignancy, emphasizing the critical need for early detection to mitigate the progression of unwanted cells and ensure timely intervention.

The complexity of breast cancer lies in its diverse manifestations, demanding a precise diagnosis distinguishing between benign and malignant tumours. Early detection isparamount, as the treatment and prevention strategies differ markedly based on the nature of the tumour. Regrettably,the absence of a dedicated diagnostic machine for early- phase cancer detection poses a significant challenge. Efforts to develop a sophisticated diagnostic system leveraging Machine Learning methodologies have gained prominence. This approach employs algorithms for tumour classification, offering a more accurate and time-efficient means of identifying cancerous cells compared to conventional testslike mammograms, ultrasound, and biopsies.

Similarly, lung cancer, a leading cause of cancer-related deaths, presents unique challenges due to its tendency to manifest symptoms in advanced stages. Despite the reliability of CT imaging as a potent diagnostic tool, challenges persistin accurately identifying cancerous cells, including intensity variations in CT scan images and potential misjudgements of anatomical structures by medical professionals. Recognizing the significance of early detection and the limitations in current diagnostic capabilities, computer-aided diagnosis (CAD) has emerged as a promising supplement for radiologists and doctors.

A multitude of systems and research endeavours in lung cancer detection using image processing and machine learning techniques have been explored. However, existing systemsexhibit variances in detection accuracy, prompting the ongoingpursuit of higher precision, ideally approaching 100%. This literature survey delves into recent advancements, scrutinizing systems developed for breast cancer and lung cancer detectionbased on CT scan images. Through a comprehensive analysis, we aim to identify the most promising systems, theirlimitations, and contribute to the discourse by proposing anew model that strives to enhance accuracy and efficacy in lung cancer detection.

II. LITERATURE SURVEY

A. Recent advancement in cancer detection using machinelearning: Systematic survey of decades, comparisons and challenges.

Recent advancements in cancer detection using machinelearning have garnered significant attention in the medicalfield. This systematic survey aims to explore the evolution of machine learning techniques for detecting various types of cancer, including breast, brain, lung, liver, skin cancer, and leukemia.

Machine Learning Techniques in Cancer Detection: This section delves into the application of machine learning methodologies, including supervised, unsupervised, and deep learning, in aiding the diagnosis, treatment, and prognosis of different cancer types. Emphasis is placed on the analysis of medical imaging modalities such as MRI, PET, CT, and mammography, highlighting their role in enhancing cancerdetection accuracy.

Cancer Types and Associated Challenges: The document provides comprehensive insights into six major types of humanbody cancer, elucidating their respective characteristics,etiology, symptoms, and risk factors. Furthermore, it outlines the challenges and constraints inherent in employing machine learning techniques for the detection and classification of these cancers.

Comparative Analysis and Performance Metrics: A comparative analysis of various machine learning approaches is presented, evaluating their performance on benchmark datasets using metrics such as accuracy, sensitivity, specificity,and F-measure. This critical assessment sheds light on the effectiveness and limitations of different methodologies incancer detection.

Conclusion and Future Directions: Concluding remarks underscore the promising prospects of machine learningin cancer detection and classification. Additionally, this section outlines future research avenues aimed at addressing challenges related to data quality, standardization, integration, validation, interpretability, and ethical considerations within the domain.

B. Cancer Diagnosis Using Deep Learning: A Bibliographic Review.

The paper presents an in-depth exploration of cancer diagnosis fundamentals, evaluation criteria, medical methodologies, and cutting-edge deep learning techniques across various cancer types.

Fundamentals of Cancer Diagnosis: This section elucidates the core steps involved in cancer diagnosis, including pre-processing, image segmentation, post-processing, and feature extraction. Additionally, it delineates conventional classification methods utilized by medical professionals, such as the ABCD-rule, seven-point checklist method, Menzies method, and pattern analysis.

Deep Learning Techniques in Cancer Diagnosis: An overview of artificial neural networks and their diverse variants is provided, encompassing convolutional neuralnetworks, generative adversarial networks, deep autoencoders,restricted Boltzmann machines, recurrent neural networks, long short-term memory networks, multi-scale convolutional neural networks, and multi-instance learning convolutional neural networks. Links to online resources for each technique are also supplied.

Applications Across Cancer Types: This segment summarizesthe utilization of deep learning methodologies in diagnosing breast cancer, lung cancer, brain cancer, and skin cancer. It references relevant studies employing deep learning for cancer diagnosis, citing performance metrics and associated datasets.

Discussion on Challenges and Limitations: A critical discussion is presented on the challenges and constraints encountered in leveraging deep learning for cancer diagnosis, including issues pertaining to data availability, quality,distribution, model complexity, and generalization. Potential solutions and future research directions are proposed to address these challenges.

Summary and Conclusions: This section offers a concise recapitulation of the document's key findings and contributions, emphasizing the advantages and potential of deep learning in cancer diagnosis. Furthermore, it advocates for continued research and development efforts in this burgeoning field.

C. Breast Cancer Detection using Machine Learning Tech-niques.

This literature review examines a study focusing on automated breast cancer diagnosis using machine learning techniques, specifically employing a convolutional neural network (CNN) and recursive feature elimination (RFE). Additionally, it investigates the performance comparison between CNN and various other machine learning algorithms such as SVM, KNN, logistic regression, naive Bayes, andrandom forest.

Dataset and Preprocessing: The study utilizes the BreaKHis 400X dataset, comprising 7,858 images of benign andmalignant breast tumors. Preprocessing techniques include RFE-based feature selection to reduce data dimensionality and complexity, as well as segmentation, normalization, and partitioning of the dataset into training and testing subsets.

Classifiers: An exploration of different machine learning classifiers is provided, elucidating their operational principles and advantages. This includes SVM, KNN, random forest, logistic regression, and naive Bayes, detailing how these classifiers employ distinct methodologies to learn from data and predict tumor class labels.

Proposed Methodology: The authors propose a CNN model for classifying breast images into benign or malignant categories. The architecture and layers of the CNN model, such as convolutional, ReLU, pooling, dropout, flatten, and dense layers, are explained. Additionally, the mechanism by which the CNN model extracts and integrates features from images to generate final predictions is illustrated.

Conclusion: The study concludes that the CNN model surpasses existing methods in terms of accuracy, precision, and dataset size. It asserts the CNN model's superiority in terms of efficiency, flexibility, and robustness for breast cancer detection. Moreover, future research directions for enhancing the model and extending its application to other medical domains are suggested.

D. Deep Learning Techniques for Breast Cancer Detection Using Medical Image Analysis.

This document investigates the application of deep learning models, including Convolutional Neural Network (CNN), Sparse Autoencoder (SAE), and Stacked Sparse Autoencoder (SSAE), for breast cancer detection using mammogram images. Additionally, it compares the performance of these models with existing methods and discusses the advantages and drawbacks of various breast imaging modalities.

Preprocessing and ROI Segmentation: A comprehensive overview of preprocessing techniques for mammogram images is provided, encompassing noise removal, artifact suppression, background separation, and pectoral muscle removal. The document also details the segmentation of the region of interest (ROI) using the seeded region growing (SRG) technique, which relies on the intensity values of neighbouring pixels.

Cancer Detection by Deep Learning Technique: This section delves into the specifics of three deep learning techniques employed for cancer detection: CNN, SAE, and SSAE. It elucidates how these techniques learn features from image patches, either in a supervised or unsupervised manner, and their training and testing processes using a softmax classifier. Furthermore, parameters and settings such as learning rate, number of epochs, cost function, decoderfunction, and training algorithm are discussed.

Performance Measures and Density Scoring: An evaluation of the deep learning techniques' performance is conducted using various metrics including accuracy, misclassification rate, sensitivity, specificity, precision, and negative predictive value. Additionally, the document elaborates on scoring the density of mammogram patches, indicative of breast cancer risk, utilizing multiple instance learning (MIL) techniques. Experimental results using the mini-MIAS database are presented and compared with previous techniques.

References: The document includes a comprehensive list of references spanning breast cancer detection, deep learning techniques, medical image processing, and mammographicdatabases, citing a total of 20 papers.

E. Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model

The paper titled ”Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model” provides an overview of breast cancer, including its prevalence, diagnosis, and treatment. It outlines the primary objectives, which are to review literature on infrared digital imaging and propose a novel model for breast cancer detection utilizing deep learning and image processing techniques.

Previous Techniques and Comments: This section evaluates various techniques for breast cancer detection, such as mammography, near-infrared fluorescence, thermography, and artificial neural networks. It discusses their advantages and disadvantages while presenting findings from related studies on these techniques.

Proposed Model: The architecture and workflow of the proposed model are detailed in this section. Comprising four components—pre-processing of breast thermal images, image classification framework, support vector machine model, and results—the model utilizes a pre-trained Inception V3 model for feature extraction and a linear support vector machine for classification. Tested on a dataset of 67 subjects with 480 images, the model demonstrates high accuracy and sensitivity.

Conclusions: This section summarizes the main contributions and limitations of the paper while suggesting future research directions. The paper asserts that the proposed model is a valuable addition to breast cancer detection, highlighting the potential of infrared imaging coupled with an agent for highly accurate tumor detection. Acknowledging the need for more data, sensitivity, and precise diagnosis, the paper underscores avenues for further exploration in the field.

F. Deep learning for lung Cancer detection and classification

The document delves into the utilization of deep learning techniques for the detection and classification of lung cancer nodules from CT scan images. It outlines a novel approach integrating deep learning methods with various feature extraction techniques to enhance accuracy and efficiency in lung cancer diagnosis.

Pre-processing and Segmentation: This section details the preprocessing steps involved in enhancing, denoising, and extracting the lung region from input images. Techniques such as histogram equalization, adaptive bilateral filtering, and the artificial bee colony algorithm are employed for optimal image preparation.

Feature Extraction and Selection: The document elucidates the extraction of four types of features—texture, intensity, volumetric, and geometric—from segmented lung images. Furthermore, it introduces a fuzzy particle swarm optimization algorithm for selecting the most relevant features among them, enhancing the efficiency of the classification process.

Classification and Evaluation: A novel convolutional neural network method, integrated with fuzzy particle swarm optimization, is proposed for lung cancer nodule classification. This method aims to reduce computational complexity while improving classification accuracy. The performance of the proposed method is evaluated using two datasets—the Lung Image Database Consortium (LIDC) dataset and a real-time dataset from Arthi Scan Hospital. Comparative analysis with existing methods demonstrates the superiority of the proposed method in achieving high accuracy, sensitivity, specificity, and low error rates.

G. Computer-Assisted Decision Support System in PulmonaryCancer detection and stage classification on CT images.

The document titled ”Computer-Assisted Decision Support System in Pulmonary Cancer Detection and Stage Classification on CT Images” outlines the primary focus and purpose of the research, emphasizing the importance of early detection and diagnosis of pulmonary cancer using computer-assisted methods.

Background and Motivation: This section provides context on the significance of pulmonary cancer as a leading cause of global mortality and underscores the urgency for early detection and diagnosis. It reviews existing literature on the use of artificial neural networks (ANNs) and convolutional neural networks (CNNs) for lung nodule detection and classification, highlighting challenges associated with heterogeneous datasets.

Proposed Approach: The proposed method is presented, consisting of data recording, dataset collection, model training, and diagnosis decision-making steps. The DFCNet, a deep fully convolutional neural network, is introduced for classifying pulmonary nodules into four lung cancer stages. The performance is evaluated on diverse datasets with varying scan conditions and compared with existing CNN techniques.

Pre-processing: This section details the pre-processing steps applied to CT scan images, including resizing, duplication, enhancement, segmentation, and data augmentation. These steps are crucial for preparing input data for the DFCNet and enhancing training data diversity and quality.

Deep Neural Network Architecture: The architecture and parameters of the DFCNet are explained, comprising convolutional layers with Parametric ReLU (PReLU) activation functions, max-pooling layers, batch normalization layers, dense layers, a deconvolutional layer, and a softmax layer. Training techniques such as large margin softmax loss and dropout regularization are employed.

Implementation Details: The section provides implementation specifics, including datasets used, training/testing data division, patch size, false-positive reduction methods, and stage label assignment techniques. Experimental results and performance evaluation metrics, such as accuracy, sensitivity, specificity, and Dice score, are reported and compared with existing CNN methods.

Conclusion and Future Work: Summarizing the research contributions and findings, this section discusses limitations and suggests future directions. It highlights the potential of the proposed method in aiding radiologists for improved nodule detection accuracy and treatment options. Furthermore, it proposes extending the application of DFCNet to other biomedical images and diseases, alongside leveraging IoT-enabled healthcare methods for personalized medical information access and clinical decision-making efficiency.

Lung cancer identification: a review on detection and classification

The paper focuses on the significance of detecting and classifying lung nodules from CT images for the early diagnosis and treatment of lung cancer. It reviews variousmethods and techniques aimed at achieving this goal.

CAD System Overview: This section discusses two types of computer-aided diagnosis (CAD) systems: handcraft-based and CNN-based. Handcraft-based systems utilize predefined features like texture, shape, and intensity, while CNN-based systems employ deep learning to automatically learn features from data.

CNN Architecture: A brief introduction and comparison of different CNN architectures, including LeNet, AlexNet, VGGNet, GoogLeNet, ResNet, and DenseNet, are provided. Additionally, CNN models for 3D shapes, such as multi-view CNN and volumetric 3D CNN, are described.

Nodule Detection and Classification: Recent studies utilizing CNN-based CAD systems for lung nodule detection and classification are summarized in this section. The paper compares their results and methodologies while highlighting challenges and future research directions in the field.

Conclusion: The paper concludes that CNN-based CAD systems exhibit promising performance and potential in assisting radiologists with lung cancer diagnosis. Itemphasizes the need for further research to enhance the accuracy, speed, and generalization capabilities of these systems.

H. Lung Cancer Detection using CT Scan Images

The document introduces a research paper focusing on the development of a new model for detecting and classifyinglung cancer nodules from CT scan images. It highlights the utilization of image processing and machine learning techniques in this endeavour.

Literature Review: This section reviews several existing methods for lung cancer detection and classification, pinpointing their limitations. The method proposed by Ignatius and Joseph [8], utilizing watershed segmentation and feature extraction, is identified as the current best solution, albeit lacking nodule classification into benign or malignant categories.

Proposed Model: The document proposes a novel model that enhances the current best solution by incorporating additional features such as centroid, diameter, and mean intensity. Additionally, it employs support vector machine (SVM) as a classifier to differentiate between benign and malignant nodules. Image preprocessing techniques, including median and Gaussian filters, are employed to remove noise and enhance image quality.

Implementation: This section details the implementation of the proposed model using MATLAB and real patient CT scan images obtained from the Lung Image Database Consortium (LIDC) archive. Training time, prediction time, and accuracy metrics of the proposed model are reported and comparedwith the current best solution.

Results and Evaluation: The document presents the results and evaluation of the proposed model, demonstrating its superior accuracy, sensitivity, and specificity compared to the current best solution. Scatter plots illustrating the trained model and nodule classification are provided. However, the document acknowledges certain weaknesses, such as the model's inability to classify nodules into different stages, and suggests future avenues for improvement.

Conclusion: In conclusion, the document asserts that the proposed model represents an improvement over the current best solution. It identifies areas for further enhancement and suggests future research directions to address these shortcomings.

III. PROPOSED SYSTEM

In our pursuit of enhancing cancer detection methodologies, we present a specialized approach focusing solely on deep learning (DL) techniques for both breast and lung cancer. This innovative methodology leverages cutting-edge DL models and tailored pre-processing techniques designed to maximize accuracy and efficiency in cancer diagnosis.

Dataset and Pre-processing: The foundation of our methodology involves the utilization of diverse datasets tailored for each cancer type. Image enhancement, denoising, and segmentation are conducted employing advanced algorithms such as histogram equalization, adaptive bilateral filter, artificial bee colony, and K-means clustering. Feature extraction encompasses a variety of methods, including Recursive Feature Elimination (RFE) and texture, intensity, volumetric, and geometric features extraction.

DL Model Architectures: Our methodology introduces a sophisticated DL architecture for both breast and lung cancer detection. We employ a hybrid approach that integrates EfficientNetB0, a state-of-the-art transfer learning layer, with a customized convolutional neural network (CNN) model designed specifically for each cancer type. The CNN architecture includes convolutional, ReLU, pooling, dropout, flatten, and dense layers to comprehensively analyse the intricate features within the medical images.

Dataset Split and Evaluation Metrics: The dataset is strategically divided into training and testing sets, allocating 70% for training to ensure robust model training. Our methodology employs a range of evaluation metrics, including F1 score, recall, precision, and accuracy, providinga comprehensive assessment of the DL models' performance for both breast and lung cancer detection.

Customized CNN Model: In addition to leveraging the transfer learning capabilities of EfficientNetB0, our methodology introduces a cancer-specific CNN model for each type. These models are meticulously designed to capture nuances specific to breast and lung cancer images, enhancing the adaptability and accuracy of the detection system.

Implementation and Comparative Analysis: The proposed DL methodology is implemented and rigorously tested against benchmark datasets for both breast and lung cancer. Comparative analyses are conducted, focusing exclusively on deep learning approaches, and highlighting the efficiency, accuracy, and computational complexity of our specialized models. This includes a detailed examination of confusion matrices, sensitivity, specificity, and overall classification accuracy tailored to breast and lung cancer characteristics.

IV. CONCLUSION

In conclusion, our cutting-edge methodology marks a paradigm shift in cancer detection, specifically targeting breastand lung cancers through exclusive deployment of deep learn- ing (DL) techniques. The integration of EfficientNetB0 as a transfer learning layer, coupled with cancer-specific CNN models, signifies a leap in accuracy and efficiency. Through meticulous dataset handling, advanced preprocessing methods,and a diverse set of evaluation metrics, our models ex- hibit unparalleled performance. Rigorous implementation and comparative analyses underscore the methodology's prowess, surpassing conventional approaches in terms of efficacy, pre- cision, and computational efficiency. Our DL-centric approachemerges as a potent and forward-thinking solution, promising heightened precision in the early identification of breast and lung cancers within the domain of medical image analysis.

REFERENCES

[A] Tanzila Saba, Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges, Journal of Infection and Public Health, Volume 13, Issue 9, 2020, Pages 1274-1289, ISSN 1876-0341, https://doi.org/10.1016/j.jiph.2020.06.033.

[B] Munir, K.; Elahi, H.; Ayub, A.; Frezza, F.; Rizzi, A. Cancer Diagnosis Using Deep Learning: A Bibliographic Review. Cancers 2019, 11, 1235, https://doi.org/10.3390/cancers11091235.

[C] Bhise, S., Gadekar, S., Gaur, A.S., Bepari, S. and Deepmala Kale, D.S.A., 2021. Breast cancer detection using machine learning techniques. Int. J. Eng. Res.Technol, 10(7), pp.2278-0181.

[D] Selvathi, D., Aarthy Poornila, A. (2018). Deep Learning Techniques for Breast Cancer Detection Using Medical Image Analysis. In: Hemanth, J., Balas, V. (eds) Biologically Rationalized Computing Techniques for Image Processing Applications. Lecture Notes in Computational Vision and Biomechanics, vol 25, https://doi.org/10.1007/978-3-319-61316-1 8.

[E] Mambou, S.J.; Maresova, P.; Krejcar, O.; Selamat, A.; Kuca, K. Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model. Sensors 2018, 18, 2799, https://doi.org/10.3390/s18092799.

[F] Asuntha, A., Srinivasan, A. Deep learning for lung Cancer detection and classification. Multimed Tools Appl 79, 7731–7762 (2020), https://doi.org/10.1007/s11042-019-08394-3.

[G] Anum Masood, Bin Sheng, Ping Li, Xuhong Hou, Xiaoer Wei, Jing Qin, Dagan Feng, Computer- Assisted Decision Support System in Pulmonary Cancer detection and stage classification onCT images, Journal of Biomedical Informatics,Volume 79, 2018, Pages 117-128, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2018.01.005.

[H] Thakur, S.K., Singh, D.P. & Choudhary, J. Lung cancer identification: a review on detection and classification. Cancer Metastasis Rev 39, 989–998 (2020), https://doi.org/10.1007/s10555-020-09901-x.

[I] Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A.K. Singh, A. Elchouemi, Lung Cancer Detection using CT Scan Images, Procedia Computer Science, Volume 125, 2018, Pages 107-114, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.12.016.