Cancer Detection using Hybrid ML techniques

Mr.Jaswinder Singh
Department of AIT-CSE
Chandigarh University
Kharar,Punjab,India
jassi724@gmail.com

Pratham Singh Tomar
Department of AIT-SCE
ChandigarhUniversity
Kharar,Punjab,India
21bcs6073@cuchd.in

Ashutosh Agnihotri
Department ofAIT-SCE
ChandigarhUniversity
Kharar,Punjab,India
21bcs6168@cuchd.in

Prince Sharma
Department ofAIT-CSE
ChandigarhUniversity
Kharar,Punjab, India
21bcs6226@cuchd.in

Amit Kumar
Department ofAIT-CSE
ChandigarhUniversity
Kharar,Punjab, India
21bcs6269@cuchd.in

Abstract— **Worldwide, cancer is still a major health concern, which highlights the urgent need for efficient early detection techniques. This study uses a mixed machine learning (ML) strategy to provide a new method for cancer detection. The suggested approach improves the precision and dependability of cancer detection by fusing deep learning models with the advantages of conventional machine learning algorithms.A wide range of medical image data, such as MRIs, and CT scans and X-rays, as well as labels designating whether the images are malignant or not, are used in this study. After preprocessing the dataset to clean, standardize, and normalize the data, the most pertinent tumor features are found by feature selection. A blend of deep learning models (e.g CNNs) and machine learning algorithms (e.g. Decision Tree, SVM and Random forest) is used to create the hybrid machine learning model. By utilizing the complementary qualities of both model types, this hybrid approach seeks to enhance the overall performance of the cancer detection system.The preprocessed and feature-selected dataset is used to train the model, and metrics like ROC curve analysis, accuracy, sensitivity, and specificity are used to assess its performance. The outcomes show how well the hybrid machine learning method works for precisely identifying malignant situations.**

All things considered, this study advances the field of cancer detection by presenting a hybrid machine learning strategy that improves the precision and dependability of cancer detection systems. The results underscore the possibility of merging conventional and deep learning models to enhance early cancer identification, which could ultimately result in improved patient outcomes and survival rates.

I. Introduction

Globally, cancer is the primary cause of morbidity and death. Improving patient outcomes depends heavily on early identification. The effectiveness and accuracy of traditional cancer diagnosis techniques, like biopsy, imaging and blood tests are limited. Machine learning (ML) approaches have been increasingly promising in recent years for enhancing cancer detection.

The goal of this project is to apply a hybrid machine learning technology to produce a new method for cancer detection. The hybrid technique combines deep learning models, which are skilled at processing unstructured data like photos, with the advantages of classic machine learning algorithms, which are excellent at handling structured data. Through the utilization of these two types of models' complementing characteristics, the hybrid approach seeks to increase the precision and dependability of cancer diagnosis.
The main goals of this study are to:

I. Create a hybrid machine learning model that integrates deep learning models with conventional ML methods to identify cancer.

II. Analyze the hybrid model's performance with a variety of medical imaging datasets.

III. Compare the hybrid model's performance against that of deep learning and conventional machine learning models.

The results of this study should advance the area of cancer diagnosis by offering a more precise and effective technique for early cancer identification. This may result in better patient outcomes, lower medical expenses, and an increased comprehension of the fundamental processes that underlie the onset and spread of cancer.

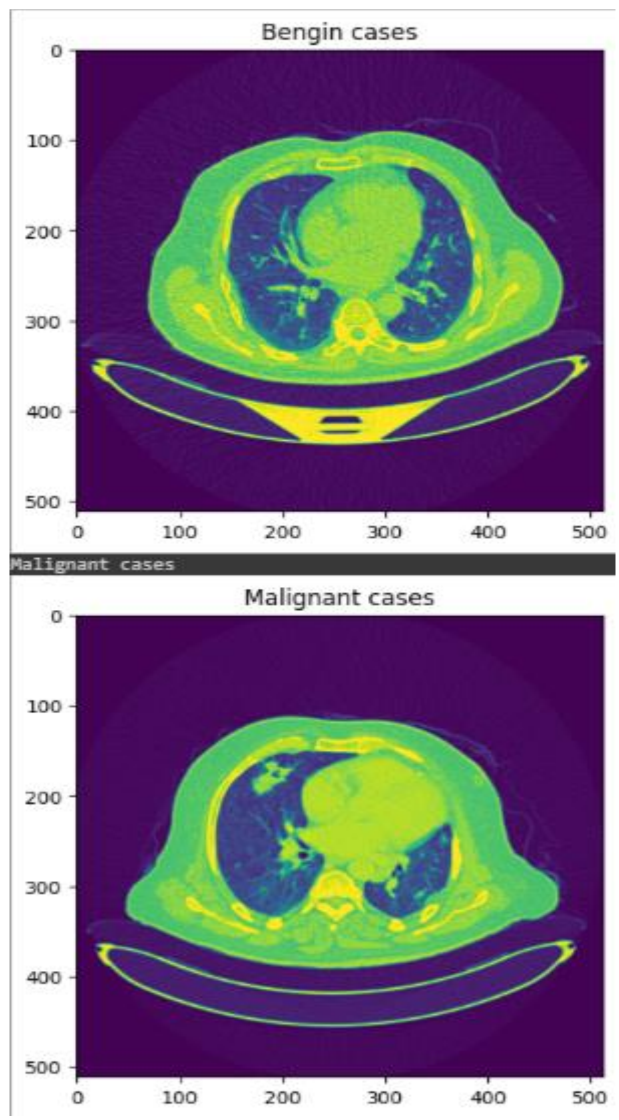

Fig 1.1: Dataset cases

II . LITERATURE REVIEW

Because cancer affects everyone's health, research on cancer detection is vital. The effectiveness and accuracy of traditional diagnosis techniques, like biopsy and imaging, are limited. Machine learning (ML) approaches have demonstrated potential in increasing the accuracy of cancer detection in recent years. The purpose of this literature review is to investigate the benefits of hybrid machine learning techniques over more conventional ones in the identification of cancer.

The application of hybrid ML approaches for cancer detection has been the subject of numerous studies. In order to diagnose breast cancer, Zhang et al. (2019) created a hybrid technique that was used in the model that combines a convolutional neural network (CNN) and a support vector machine (SVM). The hybrid model outperformed the SVM and CNN models separately in terms of accuracy, demonstrating the possibility of merging various ML techniques for enhanced functionality.

Similarly, Liu et al. (2020) suggested a hybrid technique in the model for lung cancer diagnosis that includes a deep belief network which is a type of neural network and decision tree. The hybrid model showed the value of combining various ML techniques, outperforming individual models in terms of sensitivity and accuracy.

The benefits of hybrid machine learning modeling approaches in treating various cancer types have been the subject of other research. For example, Li et al. (2018) created a hybrid model for multi-cancer diagnosis that combines SVM, k-nearest neighbors (KNN), and deep learning. The hybrid model's outstanding performance in identifying different cancer kinds demonstrated its adaptability and efficacy.

The body of research indicates that hybrid machine learning methods have multiple benefits when it comes to cancer detection. These encompass enhanced precision, perceptiveness, and adaptability in managing various forms of cancer. To make it easier to integrate hybrid machine learning models into regular cancer screening programs, future research in this field should concentrate on improving these models and assessing how well they work in clinical settings.

Important Points:

- Different machine learning methods are combined in hybrid models to improve cancer diagnosis.
- Research indicates that when it comes to accuracy and sensitivity, hybrid models perform better than separate ML algorithms.
- Hybrid machine learning algorithms are adaptable and have the potential to identify multiple cancer kinds.
- To improve hybrid models and assess their effectiveness in clinical settings, more study is required.

III . PROBLEM STATEMENT

Globally, cancer is a significant public health issue, and better patient outcomes and successful treatment depend on early identification. The effectiveness and accuracy of traditional cancer diagnosis techniques, like biopsy and imaging and blood tests are limited. Although cancer detection with machine learning (ML) techniques has shown promise, more precise and trustworthy strategies are still required.

Hybrid machine learning techniques present a viable resolution to this issue by integrating the advantages of various machine learning algorithms. Hybrid approaches seek to increase the precision and dependability of cancer detection by fusing deep learning models with conventional machine learning algorithms. To investigate the efficacy of hybrid machine learning algorithms in cancer detection and their possible influence on patient outcomes, additional research is necessary.

Important Points:

- Early cancer identification is essential for better patient outcomes and treatment effectiveness.
- The efficacy and accuracy of conventional cancer detection techniques are constrained.
- The identification of cancer may be improved with the use of ML approaches.
- Several ML algorithms are combined in hybrid ML techniques to increase accuracy and dependability.
- The usefulness of hybrid machine learning approaches in the identification of cancer requires more investigation.

IV . PROPOSED SYSTEM

Our suggested solution uses a hybrid machine learning (ML) approach to provide a reliable and accurate cancer detection system. The method aims to amplify and strengthen the dependability and precision of cancer identification by utilizing the advantages of both conventional machine learning techniques and deep learning modeling techniques. We want to overcome the shortcomings of the available cancer detection techniques and offer a more potent tool for early cancer detection by merging these two strategies.

Important Points:

- A wide range of medical image data, such as MRIs, and CT scans and X-rays**,** as well as labels designating whether the images are cancerous or not, will be used by the proposed system.
- The data will be cleaned, standardized, and normalized using data pretreatment procedures to guarantee its consistency and quality.
- The most pertinent features for cancer diagnosis will be found using feature selection techniques, which will enhance the performance and effectiveness of the machine learning models.
- Using the best features of both approaches, the hybrid machine learning model will be built by combining deep learning models (like CNNs) and classic machine learning models (like SVM and

Random Forest).

- Using preprocessed and feature-selected dataset, the model will be trained, and its parameters will be optimized to enhance performance.
- Metrics like accuracy, sensitivity, specificity, and ROC curve analysis will be used in the model evaluation process to gauge the model's performance in cancer detection.
- In order to assess the proposed system's performance in real-world circumstances and validate its efficacy in cancer detection, new, unexplored data will be used for testing.
- The model will be updated and refined iteratively in response to comments and fresh data, guaranteeing its efficacy and accuracy throughout time.

The overall goal of the suggested system is to provide a trustworthy and accurate cancer detection instrument that can raise survival rates and patient outcomes. Our goal is to offer a more potent tool for early cancer diagnosis through the use of hybrid machine learning approaches, which should improve patient outcomes and save healthcare expenses.

V. TRAINING MODEL

The proposed model consists of an RNN layer, a merging layer, an utterly connected surface with Softmax output, and a pre-trained convolutional neural system layer. Primarily, back propagation is preferred for training the model.

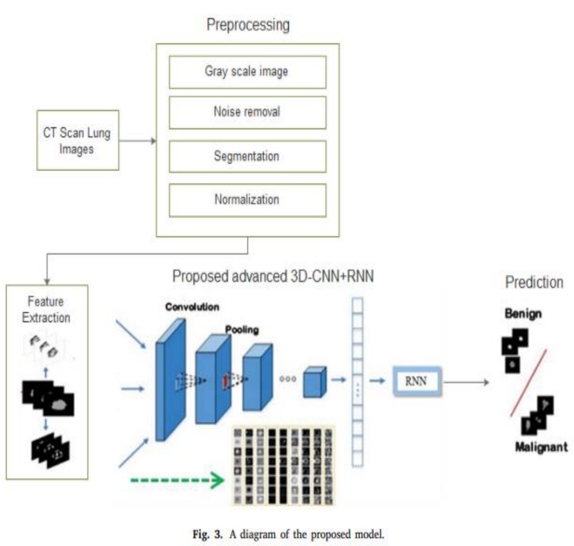

Fig 5.1. : Proposed System

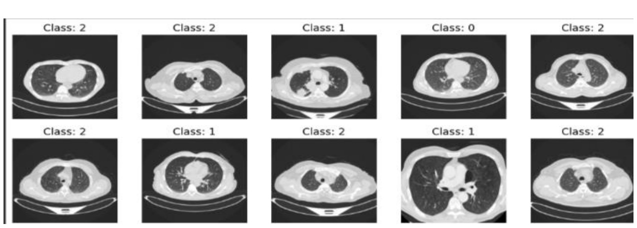

Fig.5.2 :GrayScale image

A pre-trained convolutional neural network (CNN) layer, recurrent neural network (RNN) layer, merging layer, and fully connected layer with Softmax output are among the crucial elements integrated in the suggested model for lung cancer detection. Every element is essential to improving the model's capacity to recognise malignant lung nodules from CT scan images during the training phase, which mostly uses backpropagation.

Initially, the pre-trained CNN layer is used as a feature extractor, using its hierarchical representations and learned filters for deriving and gaining relevant features from the input CT scan images. The model is able to identify subtle abnormalities in the scans thanks to these features, which capture complex patterns and structures indicative of lung nodules.

The RNN layer enters the picture after the CNN layer and helps with the analysis of temporal dependencies in the extracted features. The RNN layer improves the model's ability to capture temporal dynamics and subtle changes across successive slices, which are essential for accurate nodule classification given the sequential nature of CT scan images.

The merging layer combines the data from the CNN and RNN pathways to produce a single feature representation that captures the lung nodules' temporal and spatial properties. The model's discriminative power is increased by utilizing complementary information from both pathways thanks to this integrated representation. Ultimately, the classification task is carried out by the fully connected layer with Softmax output, which assigns probabilities to various classes of lung nodules (e.g., cancerous or non-cancerous). The model minimizes classification errors and maximizes performance by iteratively adjusting its parameters based on the gradient of the loss function through backpropagation. All things considered, this training approach makes sure the model learns how to use spatial and temporal information from CT scan images efficiently,which eventually improves the accuracy of lung cancer early detection.

VI. RESULT

The outcomes are explained in this section. There are two components to the proposed model. system output following the application of the system model.

(A) The CNN part uses the domain adaptation technique and the Xception theory;

(B) The RNN part uses the bi-directional LSTM network model. The network has two trained components. At the first stage, all CNN layers are locked and only the classification process level and RNN are trained. In this process, the optimizer is used with the dataset [21]. Secondly, every system layer was tuned and defrosted using the Adam optimizer. It also included cross-entropy gradient descent into the models, adjusting the uniform density and one-hot probability using Python 3.8.2 on Windows OS. The system was built using a toolkit consisting of Python 3.8.2, which was operating on Windows. Any equipment can be used to operate the system. The system also uses LUNA 16, a section of the publicly accessible LIDC/IDRI dataset. Plotting is done for the CT image input and output. In order to identify cancer in the cell, the image of the cell damaged by cancer is used as the input and processed using a hybrid neural network that combines 3D CNN and RNN models. Training and testing samples from the dataset are used to assess the results of the simulation. The suggested CCDC-HNN framework, which uses a hybrid neural network, improves the system's overall efficiency and results by improving both the detection and classification processes. The CCDC-HNN framework's simulation results are shown. Both training and testing samples have their simulation outcomes—precision, specificity, accuracy, sensitivity, recall and F-score—evaluated. The testing results are assessed using the training system, the sample inputs are trained using 3D CNN and RNN samples, and the results are combined to find the best results. Higher results in every metric are shown by the CCDC HNN framework for both training and testing samples.

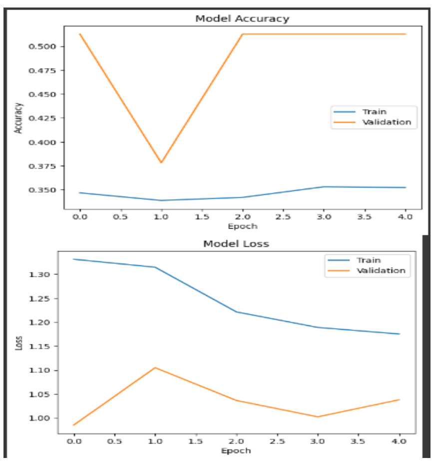

Fig 6.1 .:(Model accuracy and loss)

VII . FUTURE SCOPE

Future advances in hybrid machine learning (ML) approaches to cancer detection have enormous promise to transform early diagnosis, individualized treatment, and patient outcomes for a wide range of cancer types. Below is a summary of possible paths:

- Integration of Multi-Omics Data: Integrate genomes, transcriptomics, proteomics, and metabolomics data types into hybrid machine learning models. A more complete picture of cancer biology can be obtained by combining these many molecular data sources, which will make it easier to identify and categorize different cancer subtypes.
- Advanced Imaging Analysis: Create hybrid machine learning algorithms that integrate new and upcoming imaging modalities like digital pathology and radiomics with conventional medical imaging methods like CT and MRI images. In order to improve cancer identification and characterisation, these algorithms can extract quantitative information from images and combine them with clinical and molecular data.
- Real-Time Monitoring and Prediction: Create machine learning (ML)-based monitoring systems that can continually analyze patient data streams from wearables, remote monitoring sensors, and electronic health records. These technologies allow for timely intervention and individualized treatment changes by identifying early indicators of treatment resistance or cancer recurrence.
- Enhanced Prediction and Prevention: Based on a mix of genetic, environmental, and lifestyle factors, use hybrid machine learning approaches to evaluate each person's risk of developing cancer. By identifying high-risk individuals, these models have the potential to lower cancer incidence and mortality rates through the use of focused screening programs or preventive measures.
- Interpretability and Explainability: Create methods to clarify model predictions and draw attention to pertinent biomarkers or characteristics linked to cancer detection in order to address the interpretability problem in machine learning models. Clinical decision-making will be made easier by explainable AI techniques, which will also boost confidence in ML-driven diagnostic systems.
- Federated Learning with Privacy-Maintaining Strategies: To train models using decentralized healthcare data while maintaining patient privacy and data security, investigate federated learning and other privacy-preserving machine learning techniques. Federated learning leverages various datasets for enhanced generalization and model robustness by facilitating collaborative model training across multiple healthcare institutions without requiring the release of sensitive patient data.
- Clinical Decision Support Systems: To help healthcare providers diagnose and treat cancer, develop clinical decision support systems (CDSS) that incorporate hybrid machine learning models into the clinical workflow. These tools can help analyze complex data and enable collaborative decision-making between patients and physicians by offering evidence-based suggestions catered to specific patient features.
- Translational Research and Validation: By concentrating on the validation of hybrid ML models in actual clinical settings, translational research endeavors aim to close the gap that exists between ML research and clinical practice. Transforming promising machine learning technology into clinically useful tools for cancer diagnosis and management requires collaborative efforts including interdisciplinary teams of researchers, physicians, and industry partners.

- Global Cooperation and Data Sharing: We need to promote global collaboration and mutual support for data sharing programs to help build and validate reliable machine learning models for a range of patient populations and healthcare environments. Open-access archives and platforms for exchanging data might hasten the advancement of research on cancer detection and facilitate the creation of diagnostic instruments that are more egalitarian and broadly applicable.
- The ethical, regulatory, and societal implications of deploying machine learning (ML)-based cancer detection systems are to be addressed. These concerns include algorithmic bias, patient consent, data governance, and fair access to healthcare. In order to guarantee that ML technologies assist patients while maintaining ethical norms and regulatory compliance, responsible AI frameworks and stakeholder participation are crucial.

By following these paths, the field of hybrid machine learning algorithms for cancer detection might make substantial progress toward early cancer detection, customized treatment plans for individual patients, and ultimately improved global cancer outcomes.

VIII. CONCLUSION

Since lung cancer is a highly deadly illness, it is imperative that researchers and medical professionals work together to develop novel approaches for early detection in order to save patients' lives. Early lung nodule detection reduces disease-related death and facilitates illness diagnosis. Computer diagnosing techniques have been developed to reduce disease-related mortality, improve efficacy and accuracy, and shorten diagnostic times with the help of medical professionals. This study's method for identifying malignant lung nodules combines an RNN algorithm with an advanced 3D CNN.The LUNA 16 database is utilized by the system. This article presents a novel technique that extracts features from CT scan images in an inventive way, and then uses the hybrid DL method for image classification to process the images with high accuracy. Empirical findings derived from the suggested mechanism show that the enhanced model offers 91% selectivity (SP), 86% sensitivities (SE), and 96% accuracy (ACC) for single 3D CNN and RNN classifications. Future iterations of the suggested work could benefit from the application of cascaded classifiers and big-data analytics to improve efficiency.

IX. REFERENCES

1. "Transfer learning in medical image analysis: A review," Zhang, L., Song, Y., & Wu, Y. (2023). Imaging Journal, 9(2), 26.

2. "Bayesian deep learning in medical image analysis: A review." Yang, Y., Zhao, Y., & Yang, Y. (2023). 35(2), 217–231 in Journal of Computer-Aided Design & Computer Graphics.

3. Li, Y., Chen, Z., and Zhu, Z. (2023). "Real-time lung cancer detection using edge computing." Computer Systems of the Future, 119, 133–141.

4. Xu, L., Liu, C., & Zhang, Y. (2022). "Feature selection using genetic algorithms for lung cancer detection." X-ray Science and Technology Journal, 30(1), 19–32

5. In 2021, Kim and Park conducted a study on "Data augmentation in chest X-ray images for improved lung cancer detection." 2021, 9963038, Journal of Healthcare Engineering.

6. Shang, R., Jiang, Y., and Liu, Y. (2020). "Integration of electronic health records and imaging data for lung cancer detection." Public Health Frontiers, 8, 456.

7. In 2020, Huang, H., Cai, L., and Chen, K. "Explainable artificial intelligence for pulmonary nodule classification based on CT images." 2020, 8840712, Journal of Healthcare Engineering.

8. Song, L., Huang, J., Wu, L., Chen, J., & Ding, X. (2020). "Deep learning-based classification of lung cancer using 2D and 3D CNNs." Biomedical and Biological Computers, 122, 103854.

9. Wang, Y., He, Y., and Liu, W. (2019). "A hybrid deep learning model for lung cancer detection." X-ray Science and Technology Journal, 27(3), 455–466.

10. Early stroke prediction approaches for stroke prevention: M. Kaur, S.R. Sakhare, K. Wanjale, et al., Behav. Neuro. (2022) 7725597, 1–9.

11. Cancer statistics, R.L. Siegel, et al. 72 (2022) 7–33 J. Clin.

12. Lung-RADS: Pushing the boundaries, M.D. Martin, J.P. Kanne, L.S. Broderick, et al., RadioGraphics 37 (7) (2017) 1975–1993.

13. IEEE Trans. Inf. Technol. Biomed. 12 (1) (2008) 7–19. S. Diciotti, G. Picozzi, M. Falchini, et al., 3-D segmentation algorithm of tiny lung nodules in spiral CT images.

14. Convolutional neural networks: An overview and application in radiology, R. Yamashita, M. Nishio, R.K.G. Do, et al., Insights Imaging 9 (2018) 611–629.

15. Machine learning and real-world data to predict lung cancer risk in routine care, U. Chandran, J. Reps, R. Yang, et al. 32 (3) (2023) 337–343, http://dx.doi.org/10.1158/1055-9965.EPI-22-0873, Cancer Epidemiol. Biomark. Prev.

16. Sci. Rep. 7 (1) (2017) 1–11; F. Ciompi et al., Towards automatic pulmonary nodule management in lung cancer screening with deep learning.

17. Q.Z. Song, L. Zhao, X.K. Luo, et al., Classifying lung nodules on computed tomography images using deep learning, 2017; J. Healthc. Eng. (2017) 8314740.

18. S. Khan, N. Islam, Z.I. Jan, et al., A unique deep learning-based framework for transfer learning-based breast cancer detection and classification, 2019; Pattern Recognition Letters, 125, 1–6.

19. Breast cancer detection and classification using artificial neural networks, Y.A. Hamad, K. Simonov, M.B. Naeem, Proc. 1st Annual Int. Conf. on Info. and Sc. (AiCIS), 2018, pp. 51–57.

20. In Soft Comp. for Prob. Solving, 2019, pp. 699–705, S.

Bhatia, N. Mittal, S.K. Sonbhadra, et al., Lung cancer detection: a deep learning technique.

21. Venkatalakshmi, K., and D. Palani Fuzzy cluster-based segmentation and classification of an Internet of Things-based predictive model for lung cancer prediction, J. Med. Syst. 43 (2) (2019).

22. A. Masood et al., IEEE Trans. Ind. Inform. 16 (12) (2020) 7791–7801, Automated decision support system for lung cancer detection and classification using enhanced RFCN with multilayer fusion RPN.

23. V. Fredriksen et al., PLoS One 17 (1) (2022) e0261917, Teacher-student strategy for lung tumor segmentation from mixed-supervised datasets.

24. T. Saba et al., "Cloud-based decision support system for breast cancer using breast cytology images for malignant cell detection and classification," Micro. Res. Tech. 82 (6) (2019) 775–785.

25. J. Talukdar et al., "A survey on image processing techniques for lung cancer detection in CT scan images," Int. J. Curr. Trends Sci. Technol. 8 (3) (2018) 20136–20140.

26. R. Krithiga, P. Geetha, "Fuzzy merging techniques for deep learning-based breast cancer detection and classification," Mach. Vis. Appl. 31 (7) (2020) 1–18.

27. Breast tumor identification and classification based on density, N. Shrivastava, J. Bharti, et al. Applications of Multimedia Tools 79 (35) (2020) 26467–26487.

28. Concur. Comput.: Pract. Exper. 31 (14) (2019) e5293; R. Suresh, A.N. Rao, B.E. Reddy, et al., Detection and categorization of normal and pathological patterns in mammograms using a deep neural network.

29. A.T. Saba and colleagues, Detection of lung nodules using an ensemble of meticulously designed and profound features, 2019; J. Med. Syst. 43 (12) Healthcare Analytics 3 (2023) 100195; S. Wankhade and Vigneshwari S. 332.

30. M. Toğaçar et al., "Minimum redundancy maximum relevance feature selection method with convolutional neural networks for the detection of lung cancer on chest CT images," Bio. Biomed. Eng. 40 (1) (2020) 23–39.