International journal of innovation in Engineering

journal homepage: www.ijie.ir

Research Paper

Machine Learning Research On Breast And Lung Cancer Detection

D Nageswara Rao[a1]

[a] *Department of Computer Science and Engineering, Chitkara University Institute of engineering and Technology, Chitkara University, Punjab, India*

ARTICLE INFO

Received: 20 December 2022

Reviewed: 12 January 2023

Revised: 23 February 2023

Accepted: 08 March 2023

Keywords:

Mammogram, Breast, Lung, Cancer, Diagnosis.

ABSTRACT

As the diagnosis of these cancer cells at late stages causes greater pain and raises the likelihood of death, the initial-state cancer finding is crucial to giving the patient the proper care and reducing the risk of dying from cancer. The publication offers a chance to research breast and lung cancer detection techniques as well as various algorithms for cancer early detection. With the aid of various image kinds and test results data sets, hybrid approaches are utilized to identify lung and breast cancer based on the size and form of the cells. The basic concept of breast and lung cancer block diagram is also explained in this study, with an emphasis on the difficulties and potential future applications of cancer detection and diagnosis techniques.

[1] **Corresponding Author**
dronamraju8@gmail.com

1. Introduction

Cancer is a complicated illness that results from the unchecked division of aberrant cells in some parts of the body, making it responsible for a number of illnesses (Krishna et al, 2019; patan et al, 2020). Therefore, early detection of cancer is challenging due to the small size of cancer cells (Ghantasala et al, 2020a). Numerous studies on the identification and classification of cancer cells have been published (Bhowmik et al, 2021). For instance, the Random Forest algorithm, Bayesian methods, Support Vector Machines, some unsupervised and semi-supervised techniques favored by Logic-Based classifiers, comparative analysis of different image PDE (Partial Differential Equations) based segmentation, and other techniques to analyses breast images (Sreehari & Ghantasala, 2019; Ghantasala et al, 2021a; Ghantasala & Kumari, 2021a; Ghantasala & Kumari, 2021b; Ghantasala et al, 2021b; Gadde et al, 2022; Pradeep Ghantasala et al, 2022a). Finding a novel, reliable approach to diagnose cancer at an earlier stage is crucial (Ghantasala et al, 2021c; Kongala et al, 2022). Researchers have a fantastic opportunity to segregate and categorize the cancer cells utilising deep learning-based convolution neural network approaches because big imaging datasets are readily available (Chandana et al, 2020; Mandal et al, 2020; Malleswari et al, 2022). Our main focus will be on creating a convolution neural network-based method for classifying and segmenting breast and lung cancer cells.

2. Literature Review

The left-censored Cox proportional hazards model was used to examine the effect of mammography on breast cancer mortality (Narod et al, 2014). K-medoid clustering is utilized to obtain breast cancer data in a syndrome cluster in order to assess an analysis (Ping et al, 2016). Neural networks and association rule mining for breast cancer were key components of study (Boutorh & Guessoum, 2015). A new hybrid Grey Wolf Optimizer technique, integrating decision trees and other applied techniques like C4. S PSOC4. S, was given (Vosooghifard & Ebrahimpour, 2015). A classifier based on non-convex LP-norm minimization and SoC was presented (Wang et al, 2015). (Sum of Coefficient). Machine learning is a useful method by looking into somatic mutations to suggest cancer sites (Chen et al, 2015). Elouedi et al (2014), applied the hybrid analysis method to investigate the problem related to breast cancer using decision trees and clustering. A learning method known as multiple clustered instances learning was devised to gather the units in histopathological images (Xu et al, 2012). In paper (Sim et al, 2011), interpretate MRI breast pictures, accuracy of cancer symptom detection, and CAD system are highly helpful concepts to consider. Edge detection and PDE (partial differential equation) based segmentation algorithms are used for the identification of lung cancer (Tripathi et al, 2019). The segmentation and classification algorithms were assessed, which combined patch-level classification with a deep learning idea. If lung cancer is suspected, the person will undergo a number of screening tests before being examined for the existence of malignant tumour cells. Physicians recommend using medical imaging tools to identify lung cancer. Imaging tools aid in spotting suspicious areas, teaching about cancer stages, approving therapies, and watching for any indications that cancer cells may be coming back. Cell segmentation can be seen as creating an accurate zone of cells nearby in an image when used in the programmed diagnosis of cancer, where identifying cells is the first but most important step. Automated cell segmentation is seen as a significant barrier in the pipeline of automated histopathological image analysis and is still a major area of research. Detection of ancient cells paired with cell segmentation (Xu & Juang, 2015; Wang et al, 2016). A pre-processing stage analysis revealed that CT is one of the less expensive ways to find lung cancer. Medical CT pictures are appropriate for adaptive and filter median (Senthil Kumar et al, 2019).

3. Models and Methods

We are aware that if cancer is discovered at an early stage, the likelihood of survival dramatically rises for the majority of patient groups. Many lives have been saved by current screening programmers for colorectal, breast, and cervical cancer, but they are not sensitive enough. Research is necessary to improve our biological comprehension of early malignancies, to improve our capacity to recognize them, and to investigate the technologies employed (Rupa et al, 2022). Several places are actively conducting high-caliber research with an emphasis on cancer early detection. Lung cancer is a fatal illness. There are numerous different sizes and shapes of cancer cells (Kishore et al, 2021). Early detection of it is challenging.

Bayesian methods, Logic-Based classifiers, Support Vector Machines, for instance the Random Forest algorithm, and a few other unsupervised, semi-supervised methods used for breast image classification are used to detect cancer cells in a microscopic image (Reddy et al, 2021; CADe, 2020). Tumor Marker Test, Complete Blood Count, and image into multiple regions are also important to detect cancer cells in a microscopic image. Marker Controlled, Watershed Segmentation, Edge Identification, and PDE based segmentation techniques are compared to other image segmentation methods for the detection of lung cancer (Pradeep Ghantasala et al, 2022b). Establishing a novel, reliable approach to assess cancer, the tissues contained in the input mammography images of breast cancer, and the input CT image of the lungs is crucial.

3.1. Pre-processing of images

The median filter will be applied to lung cancer CT scan images at the pre-processing stage for adaptive contrast enhancement. The AHE has the ability to enhance local contrast and enhance the details in the image (Ghantasala et al, 2021d). In order to increase contrast in mammography pictures of breast cancer, contrast-restricted adaptive histogram equalization, a kind of AHE, will be applied.

3.2. Designing of model to identify the presence of lung cancer:

Using a convolution neural network and chest CT images, the suggested method can identify lung cancer. The CT image's lung areas are first removed, and each slice in those regions is segmented to look for malignancies. The CNN architecture is trained using the segmented tumour regions. The patient photos are then tested using CNN. Figure (1) depicts the suggested model. The trained system will be able to identify the presence of cancer in the CT scan of the lung.

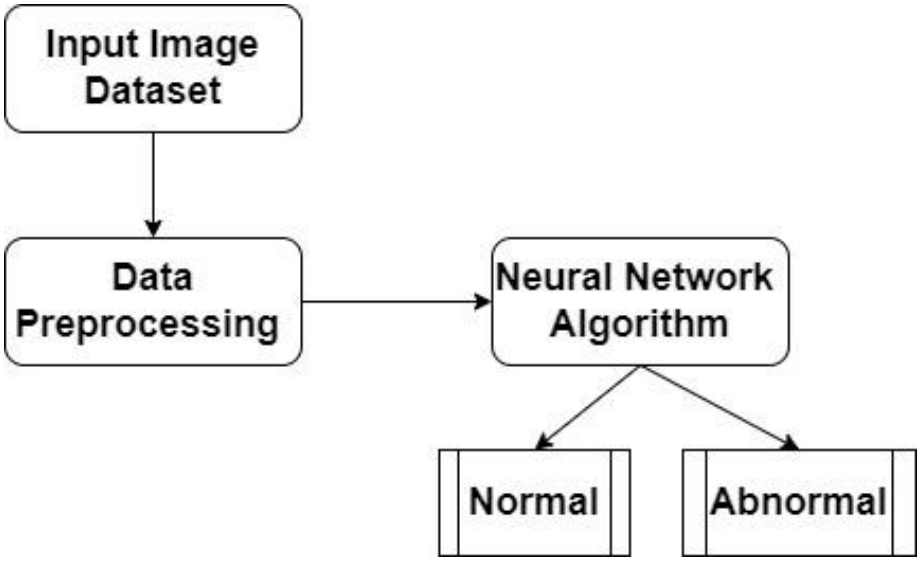

Fig. 1. Block schematic of the suggested lung cancer model

3.3. Design of Model to Identify the Presence of Breast Cancer

The suggested method for using Deep Convolution Neural Network to detect breast cancer based on mammography images is shown in Figure (2). After image improvement, a picture will be divided into sections with like characteristics using image segmentation. In the first method, the Region of interest pooling is started using circular contours. The second strategy, which combines the threshold and region-based approaches, will be used to resolve the region of interest pooling. Deep CNN will be used to classify the Region of interest pooling (Ghantasala et al, 2020b).

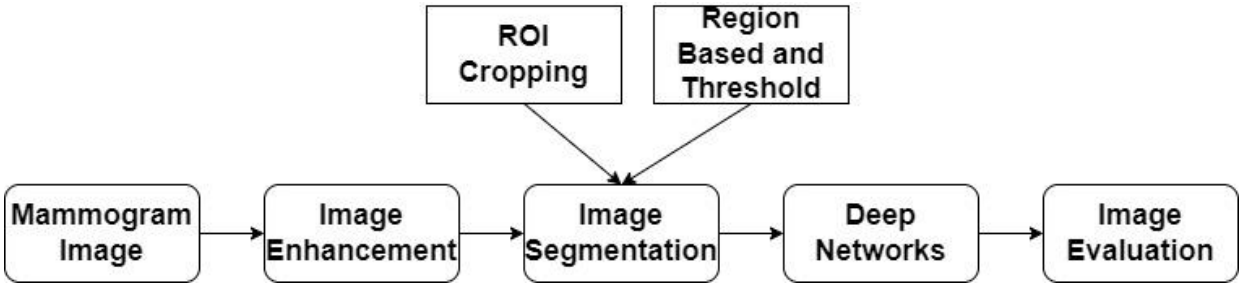

Fig. 2. Block schematic of the suggested breast cancer model

A classifier's performance is assessed using evaluation tools such the F1 score, the confusion matrix, the accuracy, and the receiver-operating curve's area under the curve (Kapula et al, 2022; Veerraju et al, 2022). It employed a confusion matrix to precisely display the classifier's performance on a table. Additionally, a detected result for an image region is implicit and can be either correct or erroneous. The same five classifiers that were employed in this study's voting procedures. Each model has a coefficient (weight) that is typically inversely related to how accurate it is in classifying data. Due to the fact that they can still influence the ultimate decision, this only partially resolves the issue of subpar models.

4. Conclusion

In the research, we provide a thorough assessment of the available machine learning-based diagnostic techniques for early identification of breast and lung cancer, which can help patients, receive better care. A variety of models based on voting have been widely utilized in machine learning to predict cancer. The deep-learning multimodal technique is effective for detecting cancer. Additionally, we must create cutting-edge instruments with great sensitivity and specificity for the detection of cancer.

References

- Bhowmik, C., Pradeep Ghantasala, G. S., & AnuRadha, R. (2021). A Comparison of Various Data Mining Algorithms to Distinguish Mammogram Calcification Using Computer-Aided Testing Tools. In Proceedings of the Second International Conference on Information Management and Machine Intelligence: ICIMMI 2020 (pp. 537-546). Springer Singapore. https://doi.org/10.1007/978-981-15-9689-6_58
- Boutorh, A., & Guessoum, A. (2015, April). Classication of SNPs for breast cancer diagnosis using neural-network-based association rules. In 2015 12th International Symposium on Programming and Systems (ISPS) (pp. 1-9). IEEE. DOI: 10.1109/ISPS.2015.7244998
- CADe, M. (2020). CADx for Identifying Microcalcification Using Support Vector Machine. Journal of Communication Engineering & Systems, 10(2), 9-16p.
- Chandana, P., Ghantasala, G. P., Jeny, J. R. V., Sekaran, K., Deepika, N., Nam, Y., & Kadry, S. (2020). An effective identification of crop diseases using faster region based convolutional neural network and expert systems. International Journal of Electrical and Computer Engineering (IJECE), 10(6), 6531-6540.

- Chen, Y., Sun, J., Huang, L. C., Xu, H., & Zhao, Z. (2015). Classification of cancer primary sites using machine learning and somatic mutations. BioMed research international, 2015. https://doi.org/10.1155/2015/491502
- Elouedi, H., Meliani, W., Elouedi, Z., & Amor, N. B. (2014, August). A hybrid approach based on decision trees and clustering for breast cancer classification. In 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR) (pp. 226-231). IEEE. DOI: 10.1109/SOCPAR.2014.7008010
- Gadde, S. S., Anand, D., Sasidhar Babu, N., Pujitha, B. V., Sai Reethi, M., & Pradeep Ghantasala, G. S. (2022). Performance Prediction of Students Using Machine Learning Algorithms. In Applications of Computational Methods in Manufacturing and Product Design: Select Proceedings of IPDIMS 2020 (pp. 405-411). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0296-3_36
- Ghantasala, G. P., & Kumari, N. V. (2021a). Identification of Normal and Abnormal Mammographic Images Using Deep Neural Network. Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146, 7(1), 71-74.
- Ghantasala, G. P., & Kumari, N. V. (2021b). Breast Cancer Treatment Using Automated Robot Support Technology for Mri Breast Biopsy. International Journal of Education, Social Sciences and Linguistics, 1(2), 235-242.
- Ghantasala, G. P., Kallam, S., Kumari, N. V., & Patan, R. (2020a, March). Texture recognization and image smoothing for microcalcification and mass detection in abnormal region. In 2020 international conference on computer science, engineering and applications (ICCSEA) (pp. 1-6). IEEE. DOI: 10.1109/ICCSEA49143.2020.9132858
- Ghantasala, G. P., Kumari, N. V., & Patan, R. (2021c). Cancer prediction and diagnosis hinged on HCML in IOMT environment. In Machine Learning and the Internet of Medical Things in Healthcare (pp. 179-207). Academic Press. https://doi.org/10.1016/B978-0-12-821229-5.00004-5
- Ghantasala, G. P., Reddy, A. R., & Arvindhan, M. (2021d). Prediction of Coronavirus (COVID-19) Disease Health Monitoring with Clinical Support System and Its Objectives. In Machine Learning and Analytics in Healthcare Systems (pp. 237-260). CRC Press.
- Ghantasala, G. P., Tanuja, B., Teja, G. S., & Abhilash, A. S. (2020b). Feature Extraction and Evaluation of Colon Cancer using PCA, LDA and Gene Expression. Forest, 10(98), 99.
- Kapula, P. R., Jeslin, J. G., Hosamani, G., Vats, P., Shelke, C. J., & Shukla, S. K. (2022, April). The Block Chain Technology to Protect Data Access Using Intelligent Contracts Mechanism Security Framework for 5g Networks. In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 202-206). IEEE. DOI: 10.1109/ICACITE53722.2022.9823471
- Kishore, D. R., Syeda, N., Suneetha, D., Kumari, C. S., & Ghantasala, G. P. (2021). Multi scale image fusion through Laplacian Pyramid and deep learning on thermal images. Annals of the Romanian Society for Cell Biology, 3728-3734.
- Kongala, L., Shilpa, T., Reddy Madhavi, K., Ghantasala, P., & Kallam, S. (2022). Applying Machine Learning to Enhance COVID-19 Prediction and Diagnosis of COVID-19 Treatment Using Convalescent Plasma. In Intelligent Computing and Applications: Proceedings of ICDIC 2020 (pp. 479-489). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-4162-7_45
- Krishna, N. M., Sekaran, K., Vamsi, A. V. N., Ghantasala, G. P., Chandana, P., Kadry, S., ... & Damaševičius, R. (2019). An efficient mixture model approach in brain-machine interface systems for extracting the psychological status of mentally impaired persons using EEG signals. Ieee Access, 7, 77905-77914. DOI: 10.1109/ACCESS.2019.2922047
- Malleswari, D. N., Rao, D. N., Vidyullatha, P., Ghantasala, G. P., & Sathiyaraj, R. (2022, December). Enhanced SS-FIM Algorithm For High Utility Uncertain Itemsets. In 2022 IEEE 2nd International Conference on Mobile

Networks and Wireless Communications (ICMNWC) (pp. 1-5). IEEE. DOI: 10.1109/ICMNWC56175.2022.10031871

- Mandal, K., Ghantasala, G. P., Khan, F., Sathiyaraj, R., & Balamurugan, B. (2020). Futurity of Translation Algorithms for Neural Machine Translation (NMT) and Its Vision. In Natural Language Processing in Artificial Intelligence (pp. 53-95). Apple Academic Press.
- Narod, S. A., Sun, P., Wall, C., Baines, C., & Miller, A. B. (2014). Impact of screening mammography on mortality from breast cancer before age 60 in women 40 to 49 years of age. Current Oncology, 21(5), 217-221. https://doi.org/10.3747/co.21.2067
- Patan, R., Ghantasala, G. P., Sekaran, R., Gupta, D., & Ramachandran, M. (2020). Smart healthcare and quality of service in IoT using grey filter convolutional based cyber physical system. Sustainable Cities and Society, 59, 102141. https://doi.org/10.1016/j.scs.2020.102141
- Ping, Q., Yang, C. C., Marshall, S. A., Avis, N. E., & Ip, E. H. (2016). Breast cancer symptom clusters derived from social media and research study data using improved $ k $-medoid clustering. IEEE transactions on computational social systems, 3(2), 63-74. DOI: 10.1109/TCSS.2016.2615850
- Pradeep Ghantasala, G. S., Nageswara Rao, D., & Patan, R. (2022a). Recognition of Dubious Tissue by Using Supervised Machine Learning Strategy. In Applications of Computational Methods in Manufacturing and Product Design: Select Proceedings of IPDIMS 2020 (pp. 395-404). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0296-3_35
- Pradeep Ghantasala, G. S., Reddy, A. R., & Mohan Krishna Ayyappa, R. (2022a). Protecting Patient Data with 2F-Authentication. Cognitive Intelligence and Big Data in Healthcare, 169-195. https://doi.org/10.1002/9781119771982.ch7
- Reddy, A. R., Ghantasala, G. P., Patan, R., Manikandan, R., & Kallam, S. (2021). Smart Assistance of Elderly Individuals in Emergency Situations at Home. Internet of Medical Things: Remote Healthcare Systems and Applications, 95-115. https://doi.org/10.1007/978-3-030-63937-2_6
- Rupa, C., MidhunChakkarvarthy, D., Patan, R., Prakash, A. B., & Pradeep, G. G. (2022). Knowledge engineering–based DApp using blockchain technology for protract medical certificates privacy. IET Communications, 16(15), 1853-1864. https://doi.org/10.1049/cmu2.12439
- Senthil Kumar, K., Venkatalakshmi, K., & Karthikeyan, K. (2019). Lung cancer detection using image segmentation by means of various evolutionary algorithms. Computational and mathematical methods in medicine, 2019. https://doi.org/10.1155/2019/4909846
- Sim, K. S., Chia, F. K., Chong, S. S., Tso, C. P., & Abbas, S. F. (2011). Real time based computer-aided design MRI breast cancer detection and data management system. In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV) (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- Sreehari, E., & Ghantasala, P. G. (2019). Climate changes prediction using simple linear regression. Journal of Computational and Theoretical Nanoscience, 16(2), 655-658. https://doi.org/10.1166/jctn.2019.7785
- Tripathi, P., Tyagi, S., & Nath, M. (2019). A comparative analysis of segmentation techniques for lung cancer detection. Pattern Recognition and Image Analysis, 29, 167-173. https://doi.org/10.1134/S105466181901019X
- Veerraju, S., Ghantasala, G. P., Verma, E., Kumar, A., Akram, S. V., & Pant, K. (2022, December). Cognitive Radio Technique for Blockchain-based Vehicles with IoT. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 385-392). IEEE. DOI: 10.1109/IC3I56241.2022.10073245
- Vosooghifard, M., & Ebrahimpour, H. (2015, October). Applying Grey Wolf Optimizer-based decision tree classifer for cancer classification on gene expression data. In 2015 5th international conference on computer and knowledge engineering (ICCKE) (pp. 147-151). IEEE. DOI: 10.1109/ICCKE.2015.7365818

- Wang, S., Chen, F., Gu, J., & Fang, J. (2015, March). Cancer classification using collaborative representation classifier based on non-convex lp-norm and novel decision rule. In 2015 Seventh International Conference on Advanced Computational Intelligence (ICACI) (pp. 189-194). IEEE. DOI: 10.1109/ICACI.2015.7184775
- Wang, S., Yao, J., Xu, Z., & Huang, J. (2016). Subtype cell detection with an accelerated deep convolution neural network. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19 (pp. 640-648). Springer International Publishing. https://doi.org/10.1007/978-3-319-46723-8_74
- Xu, Y., Zhu, J. Y., Chang, E., & Tu, Z. (2012, June). Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 964-971). IEEE. DOI: 10.1109/CVPR.2012.6247772
- Xu, Z., & Huang, J. (2015). Efficient lung cancer cell detection with deep convolution neural network. In Patch-Based Techniques in Medical Imaging: First International Workshop, Patch-MI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 9, 2015, Revised Selected Papers 1 (pp. 79-86). Springer International Publishing. https://doi.org/10.1007/978-3-319-28194-0_10