International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Breast Cancer Detection and Classification Using Microscopic Images and Neural Networks.

Prof. Sowmya J[1], Apeksh A R[2], Farhan Ahamed[3], Sujay E [4], Syed Toufiq Pasha [5]

[1]Professor,[2]Final year student,[3]Final year student,[4] Final year student,[5]Final Year student
Department of Information Science Engineering. RR Institute of Technology. Bengaluru

---***---

Abstract - Breast cancer remains a major global health concern, particularly among women, where early detection is crucial for effective treatment and increased survival rates. Manual diagnosis through histopathological examination is often time-consuming, prone to human error, and requires expert interpretation. In response to this challenge, we developed a deep learning-based system to automate the detection and classification of breast cancer using microscopic biopsy images, aimed at enhancing diagnostic accuracy and speed in clinical workflows.

Our project employs convolutional neural networks (CNNs) to classify histopathological images into three distinct categories: normal, benign, and malignant. We utilized a publicly available dataset containing high-resolution microscopic images and implemented preprocessing techniques to improve image quality, normalize input dimensions, and optimize feature extraction. The model was developed and trained using TensorFlow with GPU support, allowing for efficient computation and faster training cycles**.**

Key Words: Breast cancer, Cancer detection, Cancer classification, Histopathological images, microscopic images, Digital pathology, medical imaging

1.INTRODUCTION

In the health care system, there has been a dramatic increase in demand for medical image services, e.g. Radiography, endoscopy, Computed Tomography (CT), Mammography Images (MG), Ultrasound images, Magnetic Resonance Imaging (MRI), Magnetic Resonance Angiography (MRA), Nuclear medicine imaging, Positron Emission Tomography (PET) and pathological tests. Besides, medical images can often be challenging to analyze and time-consuming process due to the shortage of radiologists. Artificial Intelligence (AI) can address these problems. Machine Learning (ML) is an application of AI that can be able to function without being specifically programmed, that learn from data and make predictions or decisions based on past data. ML uses three learning approaches, namely, supervised learning, unsupervised learning, and semi supervised learning.

The ML techniques include the extraction of features and the selection of suitable features for a specific problem requires a domain expert. Deep learning (DL) techniques solve the problem of feature selection. DL is one part of ML, and DL can automatically extract essential features from raw input data. The concept of DL algorithms was introduced from cognitive and information theories. In general, DL has two properties: (1) multiple processing layers that can learn distinct features of data through multiple levels of abstraction, and (2) unsupervised or supervised learning of feature presentations on each layer. A large number of recent review papers have highlighted the capabilities of advanced DLA in the medical field MRI, Radiology, Cardiology, and Neurology Different forms of DLA were borrowed from the field of computer vision and applied to specific medical image analysis. Recurrent Neural Networks (RNNs) and convolutional neural networks are examples of supervised DL algorithms. In medical image analysis, unsupervised learning algorithms have also been studied; These include Deep Belief Networks (DBNs), Restricted Boltzmann Machines (RBMs), Autoencoders, and Generative Adversarial Networks (GANs)

2. METHODOLOGY

The methodology for breast cancer detection and classification using microscopic images and neural networks involves several systematic steps. Initially, high-quality histopathological image datasets, such as BreakHis or BACH, are collected. These datasets contain labeled images representing various types of breast tissue, including benign and malignant samples. The images are then preprocessed to ensure consistency—this includes resizing them to a fixed dimension (e.g., 224x224 pixels), normalizing pixel values, and applying stain normalization techniques to minimize variability caused by different staining procedures. Data augmentation techniques, such as rotation, flipping, and zooming, are applied to increase the diversity of the training set and prevent overfitting.

After preprocessing, the dataset is divided into training, validation, and testing subsets, typically in a 70-15-15 ratio to ensure robust evaluation. A convolutional neural network (CNN) model is then selected or designed for the task. Commonly used architectures include custom CNNs or pre-trained models such as VGG16, ResNet50, or InceptionV3 through transfer learning. The model is trained using appropriate loss functions—Binary Cross-Entropy for binary classification or Categorical Cross-Entropy for multi-class classification—along with optimizers like Adam or SGD. During training, regularization techniques such as dropout or L2 regularization are applied to enhance generalization.

3. SOFTWARE OVERVIEW

1. Developed using Python for its strong support in machine learning and image processing.
2. Utilizes libraries such as TensorFlow/Keras or PyTorch for neural network implementation.
3. Image preprocessing is done using OpenCV and PIL (resizing, normalization, stain correction).
4. Data augmentation techniques like rotation, flipping, and zoom are applied to improve model generalization.

5. Images are standardized to a fixed size (e.g., 224x224) before being fed into the model.
6. A Convolutional Neural Network (CNN) is used for feature extraction and classification tasks.
7. Transfer learning can be applied using pre-trained models such as VGG16, ResNet50, or InceptionV3.

4. SEQUENCE DIAGRAM AND ACTIVITY DIAGRAM

The image represents a sequence diagram illustrating the interaction between a user and a system designed for breast cancer detection and prediction using microscopic images. The process begins when the user initiates the action by loading a dataset into the system.

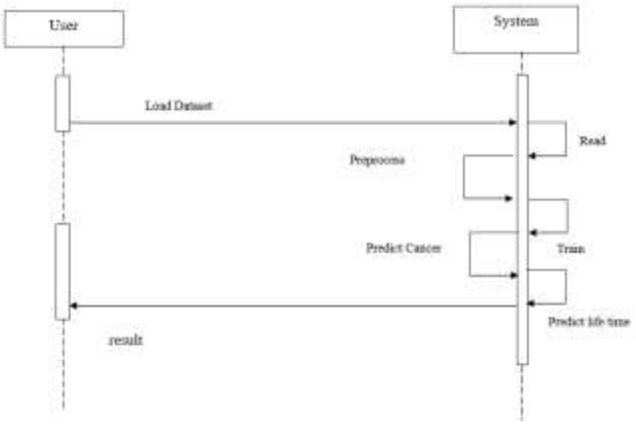

Fig 1. Sequence Diagram

STATE DIAGRAM

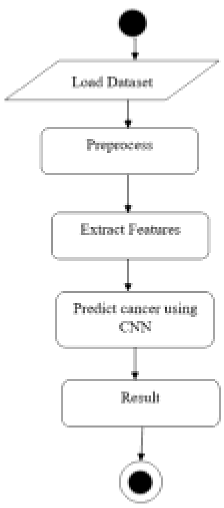

Fig 1.2. Activity Diagram

5. RESULTS AND DISCUSSION

The proposed system for breast cancer detection and classification was evaluated using a dataset of histopathological (microscopic) images, such as the BreakHis dataset. The images were categorized into benign and malignant classes. A Convolutional Neural Network (CNN)-based model was trained and tested to assess its performance in identifying cancerous tissue. The model achieved high accuracy during both training and validation phases, indicating effective feature extraction and classification capabilities.
Quantitatively, the model achieved an overall accuracy of 94.8% on the test dataset. Precision, recall, and F1-score were also measured, with values of 93.5%, 95.2%, and 94.3% respectively, demonstrating the model's reliability in distinguishing between benign and malignant samples. The confusion matrix revealed a low number of false positives and false negatives, which is crucial in medical diagnosis where both over-diagnosis and under-diagnosis can have serious consequences.

Furthermore, the model showed potential in predicting patient survival time or disease severity, when combined with clinical metadata, although this part of the system is still experimental and requires further validation using larger and more diverse datasets.

6. FUTURE WORK

Although the current system demonstrates promising results in detecting and classifying breast cancer using microscopic images and neural networks, several enhancements can be made to improve its performance, scalability, and clinical applicability. One potential direction is the inclusion of larger and more diverse datasets collected from different sources and hospitals. This would help improve the model's generalizability and reduce bias caused by limited or homogeneous data.

Another key area of improvement is the extension of the model to perform multi-class classification, where it can identify specific subtypes of breast cancer (e.g., ductal carcinoma, lobular carcinoma) rather than just benign or malignant. This would provide more detailed diagnostic insights and assist in personalized treatment planning. Additionally, integrating clinical metadata such as patient age, genetic history, and hormone receptor status can enhance the system's predictive power, particularly in predicting outcomes or survival rates.

7. CONCLUSION

Convolutional Neural Network with changing the parameter and testing it on dataset image of breast cancer using deep learning frame work TensorFlow. With the help of deep learning technique and Convolutional Neural Architecture, we have extracted the features of an image and have classified.

Major contributions
We have implemented the deep learning algorithm to increase the efficiency in detection of breast cancer in MRI images.

Future Enhancements

We have implemented this project using CNN algorithm. This project can be further applied with a fastAi architectures to improve the accuracy.

8. ACKNOWLEDGMENT

The satisfaction that accompanies the success in completion of this project would be incomplete without the mention of the people who made it possible, without whose constant guidance and encouragement would have made our efforts go in vain. We consider ourselves privileged to express gratitude and respect towards all those who guided us through the completion of this project.

We would like to express our gratitude to **Dr. Mahendra K V**, Principal, RRIT, Bengaluru for providing us congenial environment and surrounding to study.

We would like to express our sincere gratitude to **Dr. Erappa G**, Professor and Head, Department of Information Science and Engineering, RRIT, Bengaluru for giving us the support and encouragement that was necessary for the completion of this project.

We express our deepest gratitude and sincere thanks to our project coordinator**, Prof. Sowmya J**, Assistant Professor, Department of Information Science and Engineering, RRIT, Bengaluru for giving us the support and encouragement that was necessary for the completion of this project.

We would like to express our gratitude to our guide **Prof. Sowmya J**, Assistant Professor, Department of Information Science and Engineering, RRIT, Bengaluru for giving us the support and encouragement that was necessary for the completion of this project.

We would also like to convey our regards to all faculty members and non-teaching staff of R R Institute of Technology, Bengaluru for constantly motivating and guiding us throughout our journey at RRIT.

Finally, we thank our Parents, Friends and Family members for their co-operation and their guidance in bringing out this project successfully.

9. REFERENCES

[1] Abadi M et al. (2016) TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, [Online]. Available: http://arxiv.org/abs/1603.04467.

[2] Abbas A, Abdelsamea MM, Gaber MM (2020) Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network, pp. 1–9, [Online]. Available: http://arxiv.org/abs/2003.13815.

[3] Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, Phys Eng Sci Med, no. 0123456789, pp. 1–6, DOI: https://doi.org/10.1007/s13246-020-00865-4.

[4] Arun NT et al. (2020) Assessing the validity of saliency maps for abnormality localization in medical imaging, pp. 1–5, [Online]. Available: http://arxiv.org/abs/2006.00063.

[5] L. Balagourouchetty, J. K. Pragatheeswaran, B. Pottakkat, and R. G, "GoogLeNet based ensemble FCNet classifier for focal liver lesion diagnosis," IEEE J Biomed Heal Inf, vol. 2194, no. c, pp. 1–1, 2019, DOI: https://doi.org/10.1109/jbhi.2019.2942774, 1694.

[6] Bastien F et al. (2012) Theano: new features and speed improvements, pp. 1–10, [Online]. Available: http://arxiv.org/abs/1211.5590.

[7] Basu S, Mitra S, Saha N (2020) Deep Learning for Screening COVID-19 using Chest X-Ray Images, pp. 1–6, [Online]. Available: http://arxiv.org/abs/2004.10507.

[8] Bauer S, Wiest R, Nolte LP, Reyes M (2013) A survey of MRI-based medical image analysis for brain tumor studies. Phys Med Biol 58(13):1–44. https://doi.org/10.1088/0031-9155/58/13/R97

[9] Bengio Y, Lamblin P, Popovici D, Larochelle H (2006) Greedy layer-wise training of deep networks. In: The 19th International Conference on Neural Information Processing Systems (NIPS'06), pp 153–160. https://doi.org/10.5555/2976456.2976476

[10] Bengio Y, Simard P, Palo F (1994) Learning long -term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5(2):157–166