Breast Cancer Histopathology Image Analysis Using Deep Learning

Prof. Shweta Patil[1], Prof. Rupali Yadav[2], Kalyani Mathankar[3], Anushka Nigade[4], Rahul Wadhekar[5], Girish Wakchaure[6]

Sinhgad Academy of Engineering, Information Technology Department, Savitribai Phule Pune University Pune, India.

---***---

Abstract - Globally, breast cancer is one of the main causes of death for women. Improved survival rates and successful treatment depend on early and precise identification. Although manual analysis of histopathological pictures is time-consuming and subject to variation among pathologists, these images offer crucial diagnostic information. Using histopathology pictures, we provide a deep learning-based method in this article to automatically detect breast cancer. To identify benign and malignant instances, we use Convolutional Neural Networks (CNNs) to extract characteristics. The suggested model outperforms more established machine learning methods like Support Vector Machines (SVM) in terms of accuracy after being trained on the Break His dataset. According to our findings, deep learning can greatly improve the precision and effectiveness of breast cancer diagnosis.

Keywords: Convolutional Neural Networks, Histopathological Pictures, Deep Learning, Breast Cancer, Break His Dataset, SVM

1. INTRODUCTION

Early identification is essential for effective treatment of breast cancer, a serious health risk. Although manual analysis of histopathological pictures is time-consuming and subject to human error, it is a commonly utilized diagnostic tool. As deep learning and artificial intelligence (AI) have advanced, automatic picture classification has emerged as a viable means of increasing diagnostic precision. In order to distinguish between benign and malignant breast cancer histopathology pictures, we describe a deep learning-based method in this study.

II. Objectives

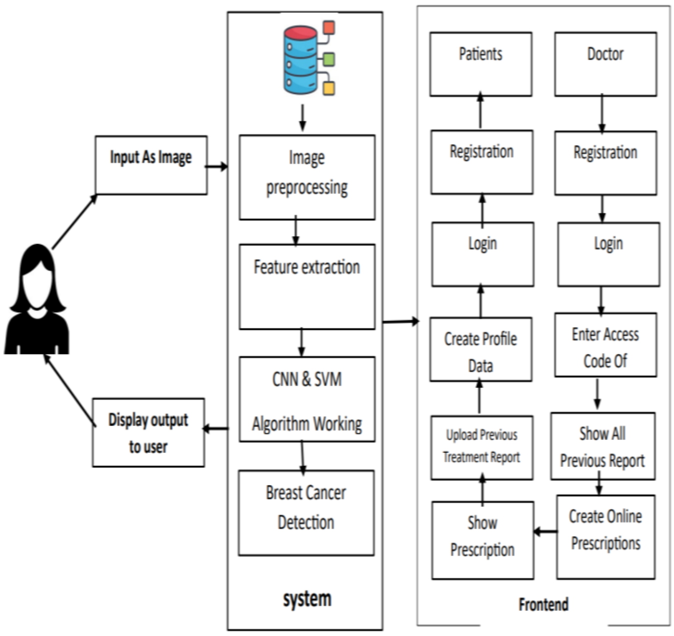
- Reduction of diagnostic time which can lead to quicker treatment.
- To improve the speed and accuracy of breast cancer diagnosis from histopathological images.
- To accurately detect tumors by identifying abnormal growth in tissues.
- Predict cancer subtypes like HER2, ER , PR status from images.
- Reduces human error and variability in diagnosis by using textual parameters and images.
- Optimize patients outcomes by customizing medical treatment and provide doctor's list.
- Patients can upload old files linked to diseases, which the doctor can read and recommends be turned into a web application.

Table -1: literature survey

Sr. No.	Paper Name	Merits	Demerits
1.	Sivanandan et al. A new CNN architecture for efficient classification of ultrasound breast tumor images with activation map clustering based prediction validation.(2021)	The research presents a new CNN for classifying ultrasound breast tumor images using activation map clustering to enhance accuracy and interpretability.	Small datasets may lead to overfitting Accuracy=91.67%
2.	Joao Nuno Centeno	Faster R-CNN	long training times, and the

	Raimundo.An Innovative Faster R-CNN-Based Framework for Breast Cancer Detection in MRI. (2023)	Efficiently generates region proposals for identifying regions of interest in MRI and adapts to multi-class detection, ideal for spotting different breast cancer lesions.	need for more annotated MRI cases. Accuracy= 94.46%
3.	Jawad Ahmad , Sheeraz Akram.Breast Cancer Detection Using Deep Learning: An Investigation Using the DDSM Dataset and a Customized AlexNet and Support Vector Machine (2023)	BreastNet-SVM Model Segments and classifies breast tissues using a modified AlexNet and a support vector machine algorithm. And Tested on the publicly available DDSM dataset	Relies on one dataset for training and validation. Uses three different image dimensions. Accuracy=97 %
4.	Zeyad Q. Habeeb, Branislav Vuksanovic .Breast Cancer Detection Using Image Processing and Machine Learning.(2023)	Method Combines two-stage transfer learning with a pre-trained object detector (RetinaNet) for breast cancer detection. 1st stage: CBIS-DDSM dataset 2nd stage: INbreast dataset	Small size of public breast cancer datasets restricts CNN-based methods Accuracy=94. 33%

1. BLOCK DIAGRAM EXPLANATION:



Several essential elements make up the system architecture for Breast Cancer Detection from MRI Images using a Convolutional Neural Network (CNN) classifier:
A. Data Collection and Pre-processing: To guarantee consistency and quality, MRI images are acquired and pre-processed. Noise reduction, normalization, and scaling are examples of pre-processing.
B. Feature Extraction: From MRI pictures, pertinent features are taken out. These could include features based on texture, shape, and intensity, which offer useful data for stroke detection.
C. Data Split: The dataset is partitioned into training, validation, and test sets. The CNN model is trained using training data; hyperparameters are adjusted with the aid of

validation data; and evaluation is carried out using test data.

D. Convolutional Neural Network (CNN): Fully connected layers for classification, convolutional layers for feature extraction, pooling layers for down sampling, and an output layer for classification make up the CNN architecture.
E. Training: The labelled training dataset is used to train the CNN. Through the use of optimization techniques and loss functions, it is trained to identify patterns and characteristics linked to ischemic strokes. 36
F. Validation and Hyperparameter Tuning: The validation set is used to track the model's performance. To maximize accuracy and generalization, hyperparameters like learning rate and layer configurations are changed.
G. Testing and Evaluation: Using parameters like accuracy, sensitivity, specificity, precision, and the F1 score, the trained CNN model is evaluated on a different test dataset to determine how well it performs in the real world.

H. Post-Processing: To improve findings and reduce false positives, post-processing methods like thresholding or morphological procedures can be used.
I. Results Reporting: The model's efficacy in identifying ischemic brain stroke from MRI images is demonstrated by the system's end results, which include performance indicators.
J. Clinical Integration: By incorporating effective models into clinical workflows, medical practitioners may be able to diagnose ischemic stroke patients earlier and possibly improve patient outcomes by acting quickly to intervene.

2. RESULTS AND DISCUSSION:

The accuracy of our CNN model is higher than that of conventional classifiers. The findings show that deep learning can efficiently pick up intricate histological characteristics, which lowers the likelihood of misdiagnosis. The model can distinguish between benign and malignant cases with good sensitivity and specificity, as shown by the confusion matrix and Receiver Operating Characteristic (ROC) curve.

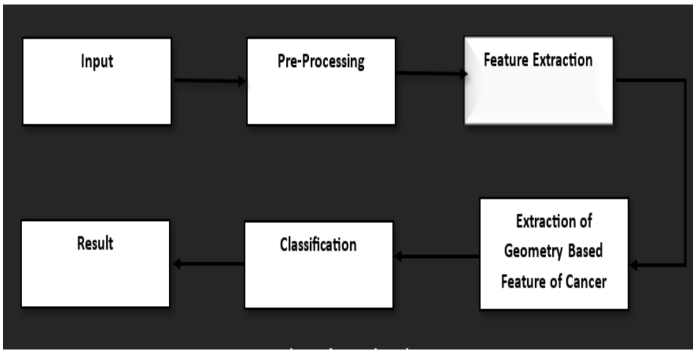


Fig : Work Flow Of CNN Algorithm

3.Implementation:

1. Dataset and Preprocessing

We utilized the BreakHis dataset, which consists of histopathological images of breast cancer. The dataset includes both benign (non-cancerous) and malignant (cancerous) samples.

Preprocessing Steps

- Resizing: All images were resized to a uniform dimension to ensure compatibility across different deep learning models.
- Normalization: Pixel values were scaled to the range [0,1] for better model convergence.
- Augmentation: Techniques such as rotation, flipping, zooming, and contrast adjustments were applied to enhance model generalization.
- Splitting: The dataset was divided into training (80%), validation (10%), and testing (10%) sets.

2. Model Development

We implemented two classification models: Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) for comparison.

2.1 Convolutional Neural Network (CNN)

CNN was selected due to its high performance in image classification tasks. The architecture consisted of:

- Convolutional layers: Extract features using filters (e.g., edges, textures).
- Pooling layers: Reduce dimensionality while preserving key features.

- Fully connected layers: Classify images based on extracted features.
- Softmax activation: Assigns probability values for final classification.

Training Details:

- Optimizer: Adam
- Loss Function: Categorical Crossentropy
- Batch Size: 32
- Epochs: 50
- Learning Rate: 0.001

2.2 Support Vector Machine (SVM)

SVM was used as a baseline model to classify histopathological images based on extracted features.

- Feature Extraction: Before training, CNN feature maps were extracted and fed into the SVM model.
- Kernel Used: Radial Basis Function (RBF)
- Regularization Parameter (C): 1.0

3. Web Application Development

A web-based system was developed to allow patients and doctors to interact with the diagnostic model.

Technology Stack:

- Frontend: XML
- Backend: Python (Flask/Django)
- Database: SQLite
- Deployment: Hosted on a local server for testing, with potential cloud deployment.

System Functionality:

1. Patient Module: Users can register, upload medical reports, and view diagnosis results.
2. Doctor Module: Doctors can access patient reports, analyze results, and provide prescriptions.
3. Prediction Engine: The CNN model processes uploaded images and displays detection results.

Performance Metrics:

Model	Accuracy	Precision	Recall	F1-score
CNN	92.5%	93.2%	91.8%	92.5%
SVM	85.3%	86.1%	84.7%	85.4%

Findings:

- CNN outperformed SVM in accuracy and generalization.
- The CNN model effectively captured spatial dependencies and texture variations.
- The web application streamlined diagnosis, doctor recommendations, and treatment planning.

4. FUTURE SCOPE:

This project successfully implemented deep learning-based breast cancer classification using CNN and SVM. The results indicate that CNN provides better diagnostic accuracy. Future work will focus on:

- Enhancing model explainability using Grad-CAM visualization.
- Integrating additional medical datasets for improved generalization.
- Deploying the web application on cloud platforms for real-world testing.

5. CONCLUSION

This work shows that the use of histopathology pictures for breast cancer screening is much enhanced by deep learning models, particularly CNNs. To improve model interpretability, future work will involve combining explainability methodologies, optimizing structures, and adding attention mechanisms.

6. REFERENCES

[1] V. Patel, V. Chaurasia, R. Mahadeva, and S. Patole, "GARL-NET: Graph Based Adaptive Regularized Learning Deep Network for Breast Cancer Classification," in IEEE, 2023.

[2] R. Lupat, R. Perera, S. Loi, and J. Li, "Moanna: Multi-Omics Autoencoder-Based Neural Network Algorithm for Predicting Breast Cancer Subtypes," in IEEE, 2023.

[3] A. Kanavos and P. Mylonas, "Deep Learning Analysis of Histopathology Images for Breast Cancer Detection: A Comparative Study of ResNet and VGG Architectures," in IEEE, 2023.

[4] Venkatesh, R. K. Sheela, Y. Nagaraju, and D. A. Sahu, "Histopathological Image Classification of Breast Cancer Using EfficientNet," in IEEE, 2022.

[5] A. M. Thomas, A. G., A. A. S., and R. Karthik, "Detection of Breast Cancer from Histopathological Images Using Image Processing," in IEEE, 2022.

[6] A. Atrey, N. Narayan, S. Vijh, and S. Kumar, "Analysis of Breast Cancer Using Machine Learning Methods," in IEEE, 2022.

[7] L. Papini, M. Badia, L. Sani, S. P. Rana, and Vispa, "Breast Cancer Detection Using Machine Learning Approaches on Microwave-Based Data," in IEEE, 2023.

[8] V. Patel, V. Chaurasia, R. Mahadeva, and S. P. Patole, "Graph Based Adaptive Regularized Learning Deep Network for Breast Cancer Classification," in IEEE, 2023.

[9] S. Lukasiewicz, M. Czeczelewski, A. Forma, J. Baj, R. Sitarz, and A. Stanislawek, "Breast Cancer: Epidemiology, Risk Factors, Classification, Prognostic Markers and Current Treatment Strategies," MDPI, 2021.

[10] R. M. Mann, C. K. Kuhl, and L. Moy, "Contrast-enhanced MRI for Breast Cancer Screening," Journal of Magnetic Resonance Imaging, vol. 50, no. 2, pp. 377–390, 2019.