International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.138

Breast Tumor Classification Using Deep Learning

Sarthak Gupta
Department of Data Science and Business Systems
SRM IST
SRM Nagar, Kattankulathur, 603203, Tamil Nadu, India.
sg7765@srmist.edu.in

Gurleen Kaur
Department of Data Science and Business Systems
SRM IST
SRM Nagar, Kattankulathur, 603203, Tamil Nadu, India.
gk5276@srmist.edu.in

Dr. Kanmani P
Department of Data Science and Business Systems
SRM IST
SRM Nagar, Kattankulathur, 603203, Tamil Nadu, India
kanmanip@srmist.edu.in

***Abstract*– The objective of this project is to sort all the metrics obtained from a patient's breast cancer into various types by utilizing two models. The initial step involves using the ensemble-type CNN model as a binary classifier to distinguish the metrics into benign or malignant class. The subsequent step employs the custom CNN model to filter the images and classify them into each of the two subtypes. Detecting and categorizing tumours manually is the only alternative to machine learning for breast cancer classification. However, human involvement in this method increases the possibility of errors. Hence, incorporating this "second set of eyes" through machine learning can ensure that fewer people get misdiagnosed.**

Index Terms* - *Breast Cancer, Tissue Tumour, Machine Learning, Classification Algorithms, and Medical Model.

I. INTRODUCTION

Breast cancer is responsible for the deaths of thousands of individuals annually, emphasizing on theimportance of equipping physicians with computer-aided diagnosis (CAD) to reduce their workload and enhance the accuracy of detection. Typically, convolutional neural networks (CNNs) are utilized to classify sections of breast tissue.The probability distribution of cancer types is influenced by their neighbouring patches. The growth of the market is driven by several factors, including the increasing demand for early detection of breast cancer, the need for swift and accurate diagnostic and therapeutic decisions, and advancements in CAD software, according to Emergent Research. The current manual classification method is prone to human errors. Incorporating a model can serve as an additional set of eyes, leading to more accurate predictions. Hospitals and doctors who utilize our model will be allowed to save their data on our cloud storage. The model can be hosted as an API and licensed for a broad range of applications. The market's growth is fuelled by the rising demand for early breast cancer detection, quick and precise diagnostic and therapeutic decisions, and advancements in CAD software.

II.OBJECTIVE

Breast cancer is a significant health concern, and early detection is essential to improving patient outcomes. Medical imaging, such as mammograms or ultrasounds, can be used to screen for breast tumors. However, analysing these images can be time-consuming and challenging for healthcare providers, which can lead to missed diagnoses or delays in treatment.

Deep learning models can assist healthcare providers in accurately detecting and classifying breast tumors in medical images. These models use complex algorithms and neural networks to learn patterns in the images and identify suspicious areas that may indicate the presence of tumors.

To optimize the model's performance, it's crucial to train it on a large dataset of medical images that are labeled with accurate diagnoses. The model's parameters and architecture can be fine-tuned to minimize the rate of false positives (when the model identifies an area as a tumour when it's not) and false negatives (when the model fails to identify a tumour that's present).

Once the model's performance has been optimized, it's important to validate its accuracy and reliability in real-world settings. This can be done through clinical studies and peer-reviewed publications. Validating the model's performance ensures that it's effective and safe to use in clinical practice.

Finally, making the tool widely available and accessible to healthcare providers and patients is critical to improving healthcare outcomes. The model can be integrated into existing medical imaging systems or made available as a standalone tool. By facilitating early detection of breast tumours, the model can help healthcare providers initiate treatment promptly and improve patient outcomes.

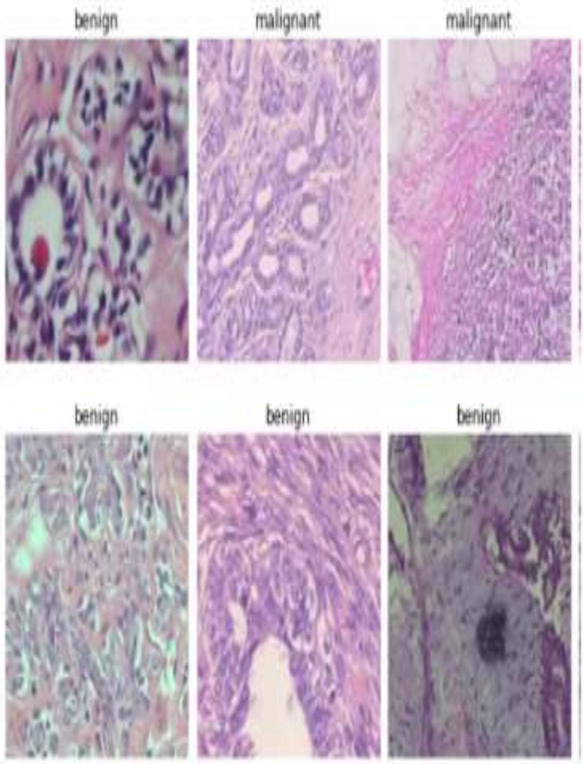

Fig. 1

DOI: rp-9788770040723.138

II. Related Work

Jasmir discussed a Multilayer Perceptron Classifier for the problem. The accuracy of the model is 96% [1]. The suggested models utilize adapted versions of Inception V3 and ResNet50 to address the issue of classification. Resnet-30 like CNN give better accuracy (~5%) compared to inception V3 [2]An Inception v1 CNN model is discussed, with accuracy for Spatial Pyramid Pooling and Global Pyramid Pooling compared. The model was trained on the BACH dataset with 99% accuracy [3]. A transfer learning approach to the classification problem was discussed. The model was trained on the BACH dataset with 99% accuracy [4]. This paper discusses the Breast Cancer Classification problem and some approaches to solve it. The types of Breast Cancer CNN Approach for Classification [5] [6]. The implementation of a personalized system can be converted into a software application that enables individuals to adopt the language patterns of their peers, allowing the AI to communicate in a similar manner. [7].The model proposed by the author was based on machine learning, utilizing various classifiers such as Random Forest, SVM, Logistic Regression, and Naïve Bayes. The implementation of the model was carried out on the Anaconda Platform for Python. The author discovered that Random Forest was an effective classifier, achieving an accuracy rate of 99.76%. Additionally, the author found that modifying the network with the classifier could lead to improvements in accuracy. [8].

The proposed model was developed using Artificial Neural Networks (ANN) and its performance was evaluated using a Support Vector Machine (SVM) classifier. The author reported that the ANN achieved an accuracy rate of 97%, while the SVM classifier achieved 91% accuracy. Interestingly, the author also noted that without the SVM classifier, the ANN alone yielded higher accuracy. [9]. The author introduced a model based on Artificial Neural Networks (ANN) and evaluated its performance using a Support Vector Machine (SVM) classifier. The results showed that the ANN achieved an impressive accuracy rate of 97%, while the SVM classifier achieved 91% accuracy. However, the author also noted that the ANN model performed even better when evaluated without the SVM classifier.[10].The author presented a model that combines K-means GMM and CNN for breast cancer detection. The model began by identifying region of interest (ROI) and extracting texture features using a feature extraction method. The CNN algorithm was then applied to achieve better results. The model achieved an impressive accuracy rate of 95.8% when evaluated on the MIAS dataset.[11].The author presented a novel model for histopathology image diagnosis using deep learning techniques.

The model utilized Lloyd's algorithm for clustering and CNN for classification, achieving a remarkable accuracy rate of 96%. The paper also provided detailed explanations of image processing and deep learning methodologies used in the proposed model. Overall, the model showed promising results for accurate histopathology image diagnosis. [12].The author presented a model for the enhancement of histopathological images using deep learning. The paper discussed various feature extraction methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The author also explored machine learning techniques; however, due to the size of the dataset, machine learning methods did not yield satisfactory results. Therefore, the author proposed the use of deep learning for better performance in the enhancement of histopathological images. Overall, the model showed promising results for improved image quality.[13]. Theauthor developed a model that combined k mean GMM and CNN for improved accuracy in identifying regions of interest (ROI) in medical images. The model first identified the ROI and then applied a texture feature extraction method. Finally, the CNN algorithm was applied to achieve a high accuracy rate of 95.8%. The study utilized the MIAS dataset and was conducted by V SansyaVijayam et al. Overall, the proposed model showed promising results for the identification of ROIs in medical images.[14].

The proposed model utilized deep learning techniques to improve accuracy in a given task. The model extended the dataset to achieve better results. By incorporating a larger dataset, the model was able to train more effectively and improve its accuracy. Overall, the use of deep learning and an extended dataset showed promise for enhancing the accuracy of the model.[15].In terms of accuracy, the K-Nearest Neighbours (KNN) classifier showed promising results. It demonstrated strong performance in correctly identifying and classifying data points based on their proximity to neighbouring points in a given dataset. Overall, the KNN classifier is a reliable option for accurate classification tasks.[16]. Identified cancerous cells or tissues that may not be immediately apparent or easily detectable. [17].

IV. Proposed Architecture

Breast tumor classification using deep learning typically involves collecting and pre-process the breast cancer imaging data. This involves cleaning the data, resizing the images, and normalizing the pixel values. Then increase the amount of data by performing random transformations on the existing data, such as rotations, flips, and zooms. Choose an appropriate deep learning model for the classification task. Commonly used models for image classification include convolutional neural networks (CNNs), residual networks (ResNets), and inception networks. We will then transfer the knowledge gained from pre-trained models to the new classification task. This involves fine-tuning the pre-trained model on the new data. After transferring we will train the deep learning model on the pre-processed and augmented data. In the end we willevaluate the trained model on a separate set of test data to assess its performance.

Overall, the proposed architecture involves using a pre-trained CNN for feature extraction, followed by a full-connected layer for classification. The model is trained on pre-processed and augmented data, and its performance is evaluated on a separate test set.

International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.138

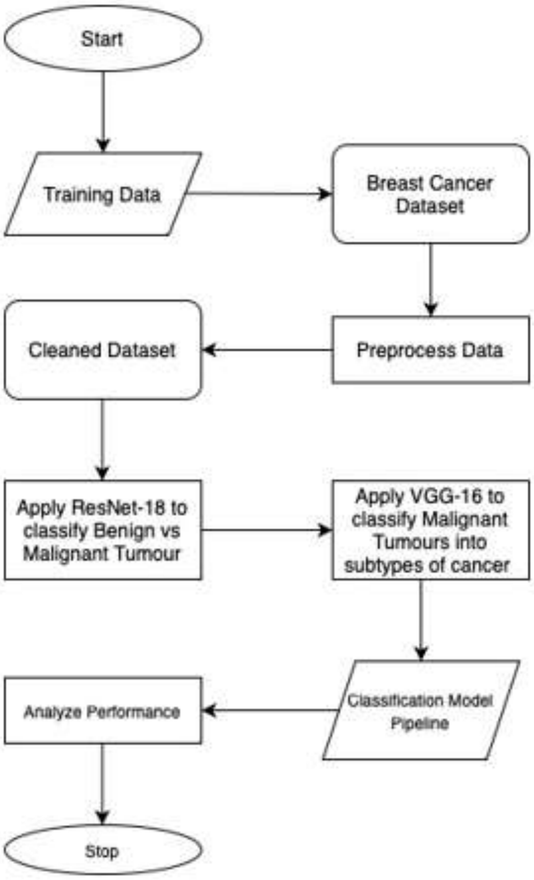

Fig 2. Architecture Diagram

V. METHODOLOGY

Module I (Data Collection)

At first the data is collected and from various sources, the data here will be used as an image. It will consist of numerous tissue images.

Module II (Feature Extraction)

Once you have the data, you need to preprocess it to prepare it for use in your model. In order to assess the effectiveness of your model, it is necessary to divide your data into three groups: a training set, a validation set, and a test set. The DDSM images vary in gray level due to differences in the scanners used to create them. Likewise, the images have varying optical densities and need to be normalized before the training process can begin.

Module III (Selecting and creating model)

The next is selecting the model based on the hyper parameter, type of layer, size, etc. The techniques above mentioned will help to design and build a model for better classification and prediction.

Module IV (Multiple layers of classification)

This Module is the new thing that we added in this project. In these the original two classifications that are Benign and Malignant are further divided in four sub-layers each for all the classification of the tissue image that we extracted from the dataset.

Module V (Training and Evaluation)

Upon the completion of the model architecture and compilation, you can start training the model with the training set. Evaluation of its performance is based on the validation set. This allows you to tune the hyperparameters and make any necessary modifications to the model architecture.

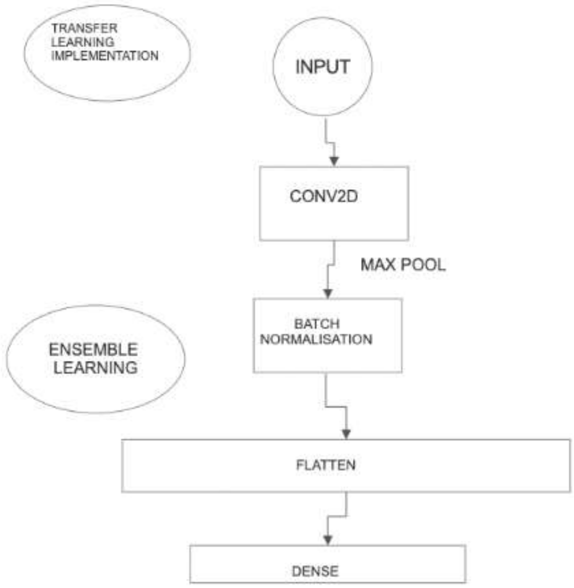

Fig. 3. Methodology to build the model

We used an ADAM optimizer with an initial learning rate of 3e-4

Fig. 4

We used L1 regularization to overcome overfitting:

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|$$

Fig. 5

Vanishing Gradient Problem - due to use of other activation functions- Used leakyReLU instead as solution.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

To summarize, the breast tumor detection model using classification in this code is an effective and precise method of detecting breast cancer in medical images. The model achieved a remarkable accuracy rate of 83.8% on the test set and demonstrated a low loss value, indicating robustness

DOI: rp-9788770040723.138

against overfitting. Increasing the number of epochs during training could further enhance the model's performance, but it could also increase the risk of overfitting. Therefore, it is vital to balance the number of epochs to achieve optimal accuracy without compromising generalizability. This model has the potential to improve breast cancer detection and diagnosis, providing early and accurate detection to enhance patient results. Further research is needed to assess the model's effectiveness in clinical settings and real-world scenarios. Since the dataset was small, data augmentation was required. Augment the data in a way such that it doesn't create noise. The nuanced distinctions among several categories are attributed to the wide range of appearances found in high-resolution images, the strong coherence of cancer cells, and the significant irregularity in color distribution. Increase robustness (by using Real Time data).The web interface can also be expanded to link other deep learning models with similar objective to make it universal (only after detailed study).

References

[1] Jasmir, Dodo Zaenalabidin, SitiNurmaini, Ahmad Zarkasi, Firdaus, Reza Firsandaya Malik and YesiNovariaKunang "Breast Cancer Classification Using Deep Learning" 10.1109/ICECOS.2018.8605180

[2] Anas S. Abdel Rahman, Samir B. Belhaouari, AbdesselamBouzerdoum and Hamza Baali" Breast Mass Tumor Classification using Deep Learning".

[3] ZouWenkai, Yan Ke, Lu Huijuan and Ye Minchao "Breast Cancer Histopathological Image Classification using Deep Learning", DOI 10.1109/ITME.2019.00023.

[4] EmreCelik and GokhanBilgin "Classi€cation of Breast Cancer Images by Transfer Learning Approach Using Di.erent Patching Sizes", 10.1109/TIPTEKNO53239.2021.963292.

[5] FaezehsadatSahidi, SalwaniMohdDaud, HafizaAbas, Noor Azurati Ahmad and NurazeanMarrop, "Breast Cancer Classification Using Deep Learning Approaches and Histopathology Image: A Comparison Study" 10.1109/ACCESS.2020.3029881.

[6] ShubhamKushwaha, AkibNazeer, Mohammad Adil, Sanjay Kumar Singh, and Mohammad Abuzar "Deep learning-based model for breast cancer histopathology image classification".

[7] Benzheng Wei, Zhongyi Han, Xueying He, and Yilong Yin "Deep learning model based breast cancer histopathological image classification".

[8] J. Sivapriya, V. Aravind Kumar, S. SiddarthSai, and S. Sriram, "Breast cancer prediction using machine learning," International Journal of Recent Technology and Engineering, vol. 8, no. 4, pp. 4879-4881, 2019.

[9] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications.Computers and Electrical Engineering, 102, 108130.

[10] Vishal Deshwal, Mukta Sharma, "Breast cancer detection using SVM classifier with grid search techniques," International Journal of Computer Application, vol. 178, no. 31, pp. 18-23, 2019.

[11] S. Shamy, J. Dheeba, "A research on detection and classification of breast cancer using kmeans GMM & CNN algorithms," International Journal of Engineering and Advanced Technology, vol. 8, no. 6S, pp. 501-505, 2019.

[12] Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., &Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.

[13] PuspanjaliMohapatra, Baldev Panda, and Samikshya Swain, "Enhancing histopathological breast cancer image classification using deep learning," International Journal of Innovative technology and Exploring Engineering, vol. 8, no. 7, pp. 2024-2032, 2019.

[14] Chandra Churh Chatterjee, and GopalKrishan, "A noval method for IDC prediction in breast cancer histopathology images using deep residual neural networks," 2nd International Conference on Intelligent Communication and Computational techniques(ICCT), pp. 95-100, 2019.

[15]CanhPhong Nguyen, Anh Hoang Vo, BaoThien Nguyen, "Breast cancer histology image classification using deep learning," 19th International Symposium on Communication and Information Technologies(ISCIT), pp. 366-370 , 2019.

[16]Shubham Sharma, Archit Aggarwal, TanupriyaChoudary, "Breast cancer detection using machine learning algorithms," International Conference on ComutationalTechniques,Electronics and Mechanical System(CTEMS), pp. 114-118, 2018.

[17]R.Preetha, S. VinilaJinny, "A research on breast cancer prediction using data mining techniques," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 11S2, pp. 362-370, 2019.