iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.56115

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

Detection of Breast Cancer from Mammogram using Convolution Neural Network

Shubham Parulekar[1], Mihir Gadkar[2], Milind Paraye[3]
Department of Electronics and Telecommunications, Sardar Patel Institute of Technology, Bhavans Campus, Old D N Nagar, Munshi Nagar, Andheri West, Mumbai

Abstract: ***Breast cancer is the leading cause of cancer among women. In the last few years, the number of cases of breast cancer has skyrocketed among younger women. Detection of breast cancer during the initial stages can greatly reduce the risk of fatality. Mammograms which are X-ray images of breast tissues are used extensively by doctors to determine the early onset of breast cancer. However, due to human error, a lack of resources and knowledge can result in inaccurate predictions which can prove fatal. The power of Artificial Intelligence to process and predict images has greatly increased in the past few years. Many modern medical devices utilize the power of sophisticated AI algorithms to aid in detecting and predicting the early onset of breast cancer. In this paper, we demonstrate how we can utilize a convolution neural network to predict the early onset of breast cancer and help solve the issue of inaccurate detection of cancer cells from mammogram images.***
Keywords: Neural Network, CNN, Mammogram, Breast Cancer Detection

I.INTRODUCTION

Breast cancer is one of the most commonly diagnosed types of malignant cancer among women of all ages with nearly 1.8 million new cases of breast cancer being detected every year worldwide and delayed detection can often result in a patient's death [2]. According to data obtained from renowned medical journals, worldwide breast cancer is detected in women every 2 minutes and 1 woman dies from the effects of breast cancer every 13 minutes [3]

Fortunately, the probability of having malignant breast cancer cells present in a large percentage of the population is quite low. Over the years various kinds of therapies have seen some success in clinical trials and real-world applications and have been very effective in reducing the incidence and progression of breast cancer tumors [2]. The early detection of breast cancer immensely improves the chances of survival. The American Cancer Society claims in a study that if breast cancer is detected early when it's still in its localization state, the 5-year relative survival rate is up to 99% [4]. It has been proven by clinical trials that mammogram screening is the best way to find breast cancer early when it is easier to treat and before it is big enough to feel or cause symptoms

Mammogram screening has a lot of advantages but being excessively reliable leads to a high risk of false positive and false negative outputs. Scientists have tried to solve this problem by utilizing a computer-aided diagnostic system(CAD) since the 1990s, The use of CAD has led to a rapid increase in detection accuracy but there has not been any significant development in the field of computer-aided diagnosis for many years and accuracy rates have plateaued. Deep learning algorithms have shown remarkable promise in the field of object detection and various experiments that leverage deep learning to detect cancer in the brain and lungs have seen moderate success, Deep Learning can be utilized to detect malignant cancer cells in mammogram images, but the challenge with this approach is the positioning and density of cancer cells in the mammogram image

Keeping these challenges in mind, we aim to help radiologists by developing an algorithm that can help detect cancerous tissue quickly and accurately within minutes without any expensive machinery. Once deployed, the model will outline the areas of interest in the patient's mammograms, detecting the presence of cancer cells and then predicting if the cells are benign or malignant and providing a report that can be further examined by radiologists with ease.

II.PROPOSED SYSTEM

Our model for breast cancer detection is developed by training it on a large and diverse dataset of about 5,000 high-quality mammogram images to ensure that the model provides results with extremely high accuracy. the model is essentially divided into various modules like the pre-processing module, a feature selection module, an association rules mining module, and a classification module.

The pre-processing module processes a set of images removing redundant data and replaces some attributes that are missing from the dataset with relevant data to ensure data continuity. The pre-processing module makes sure that the data that's fed to the module for training is consistent, and does not contain any empty or blank attributes that can negatively affect the accuracy of our trained model.

The patch selection module selects features from the images that are most prominent and reflects a pattern across all the labeled images. The associate rules mining module identifies relationships between unrelated data points helping in finding common features across extracted features. The classification module receives these images and using a weighted deep learning neural network (WDLNN) classifies them into benign and malignant.

III.CONSTRAINT

It is not possible to identify all different types of breast cancer cells utilizing only mammogram images. Sometimes, while detecting cancer cells in mammograms the patch of cancer cells is distributed in a wide area causing the model to give inaccurate results and give false positive and false negative results. A major reason for this is the density of cancer cells in breast tissues is not constant and shows extreme variation [6].

Several studies have been conducted that show that in a large percentage of women between the ages of 40 and 49, simple mammogram screening is unable to detect the presence of cancer [5]. Over 94% of women withdrew from further testing when mammogram screening detected no malignant cancer cells [5]. On average up to 10 percent of women who undertook BC Cancer Breast Cancer Screening require more extensive testing at very specific areas of the breast tissue for a more accurate and reliable result [7]

IV.IMPLEMENTATION

The implementation of the model is divided into 4 major steps:

1) Image Input
2) Image Pre-processing.
3) Model Building and Training
4) Output and Classification Result

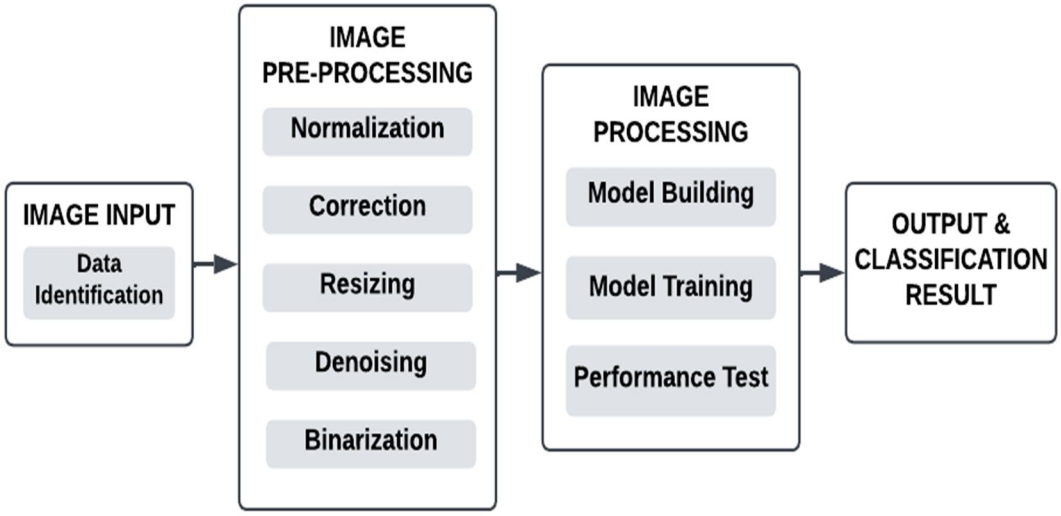


Fig. 1 Implementation Flowchart

The image input will be in JPG and PNG format and will be ingested one by one. Image clarity is an important factor in obtaining better precision and avoiding data that do not match. To ensure data clarity and suitability, we perform image pre-processing in the initial phase before it is sent to the training model. Pre-processing consists of cropping, enlarging, removing noise, etc.

Uploaded mammogram images are scanned pixel by pixel for patch detection to locate the mass of the group of cells that the model could only. Non-maximal blanking is used to detect the region of interest and ignore other irrelevant parts of the image. After detecting the patch area, the next important step in developing the classification model is sorting the image.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

After the patch training the classification model using thousands of images over hundreds of iterations, the patch classifier takes an input image and classifies it into 5 categories: background, malignant mass, benign mass, malignant calcification, and benign calcification.

Classified image patches undergo feature extraction in which the different characteristics of four types of cell mass are measured and extracted namely: cell size, cell color, cell shape, cell mass number per unit, and the total number of cell groups. These functions are extracted and then compared with other functions present in the dataset followed by applying the classification algorithm where the model recognizes these patches in image

A. Image Input

For Training our Image classification and prediction model, we have used the Wisconsin Breast Cancer Dataset (Diagnostics) which is a binary dataset with images labeled as benign or malignant to train our classification model and determine the type of cancer (benign and malignant).

1) Identifying Data

The dataset we used has 5000 high-quality mammogram images. The dataset is divided into training and validation sections with the training category having 2000 images each for benign and malignant classes and the validation folder having 500 images in each category. To build a model the dataset is split into training and testing data with 80% of data used to train and 20% percentage of data for testing purposes

B. Image Pre-processing

The next step in this process is image data pre-processing, which includes multiple sub-steps whose main goal is to transform the image data using transformation methods into an acceptable state that can be further analyzed by the model. While working on medical and healthcare-related data, extremely high accuracy is required.

One way to ensure the model's outputs are highly accurate is by training the model on a large dataset. Many times we do not possess a large and comprehensive image dataset and adequate processing power to train the model, for situations like these image preprocessing has been scientifically proven to improve the quality of image datasets by enhancing the required important features and removing unwanted distortions in the images.

Listed below are the steps that we used for image Preprocessing:

1) Image Stack Acquisition

An image stack is a series of individual images placed on top of each other. Obtaining image stacks instead of individual images adds a z-dimension image and is essential for certain probes. It helps to have a bigger stack of images We avoid over-fitting and increase the final precision as the model has more data trained in it.

2) Image Selection

This method examines the features used for the selection of images that are more valuable for analysis. The selection is made based on the criteria of having a good image quality, features required content, and being unique compared to other selected images.

3) Image Alignment

The image is aligned in a position that puts the required content upright and center to significantly increase object detection accuracy. The various types of transforms deployed for image alignment are perspective transform, affine transform, etc.

C. Model Building and Training

Deep learning models help us in getting the best possible results to predict and detect breast cancer. In our experiment, we used a convolution Neural network to detect cancer cells in breast tissues after evaluating a number of factors including the training dataset required, accuracy, speed, the computational power required, etc.

We employ a patch-based sampling algorithm to identify patches of interest in the images. This algorithm is quicker and more efficient than other widely used image classification algorithms. The patch-based sampling algorithm is well suited for processing images that have a wide range of textures which range from regular to stochastic.

The working of the algorithm can be broken down into 3 basic steps

1) Patch Sampling

This step is the starting point in training the algorithm to classify a patch of cells. The fundamental theory is to treat the image data as a collection of independent patches, sampling from a representative set of image patches. Here we use an image descriptor vector for each of the patches independently and utilize the resulting distribution of samples as a complete detailed description of the image.

2) Patch Classification

To perform grouping or division on complicated images we leverage a system that utilizes a sliding window-shaped classifier to perceive patches in a picture and create a grid-like network of probabilistic outcomes.

3) Patch Segmentation

In this step, we encapsulate the results of the patch ratings conducted in the above steps and provide a final rating of the image segmentation results. To drastically improve the field of reception, we add an additional convolutional neural network layer on top of the patch rating outputs which has the added benefit of converting the whole patch classifier into a filter.

Hence, the top layers effectively use the patch sorter to 'scan' the whole image, looking for signs of cancerous lesions and extracting higher-level features that can eventually be used for full image classification

By researching and experimenting with various neural networks and pre-trained algorithms we found that using a pre-trained neural network with tiers like the deep belief network(DBN) with multiple hidden layers which are refashioned for image classification can improve the speed and accuracy of training models, so we choose the Densenet201 model as pre-trained weights.

This model was trained in the Imagenet competition on more than a million images and it's a 201-layer deep convolution neural network. This pre-trained convolution neural network can classify images into up to 1000 categories [9]. To ensure we continuously move towards minimizing the loss function we keep the learning rate of 0.0001.

Simplifying the working of the algorithm, it classifies the cancer cells by identifying the texture of the tissue. The convolution neural network is made up of many layers, the most prominent of which are the input layer, feature extraction layer, and classification layer.

The input layer provides the input data in the format of images, this input data specifies the width, height, and number of channels. The input dataset is partitioned into two sets - a training set and a test set with the data split into 80% and 20% respectively. Before providing the image data to our model, we convert it into an acceptable format

The feature extraction layer consists of multiple convolution and pooling layers. The Convolutional layers remodel the input data by manipulating a patch of locally connecting neurons from the layer before it. Our CNN model utilizes a lot of filters to detect changes in pixel values of images to empower the model further to detect a recurring pattern in the uploaded images.

The model selects a set of relevant characteristics like variables and predictors by employing various methods of variable selection like variable Subset selection, and attribute selection. In the beginning, we start with a low number of filters to specifically detect the low-level features. As we go deeper into the CNN layers we use multiple filters to detect high-level features. Fundamentally, Feature detection consists of scanning the input with a filter of a certain size and then implementing complex matrix calculations to obtain a detailed feature map.

As our neural network deepens and we have multiple convolution layers extracting and detecting multiple features, we add a layer to provide for spatial variation so that the model can recognize an object with high accuracy even if its appearance does not match with other similar objects detected. A pooling layer then reduces the spatial dimensions of the feature map to a format compatible with the poling size of(2,2) thereby reducing the massive computation power required and the number of parameters to learn. This flattened output from a convolution layer is then fed to the next convolution layer. We deploy a global average pooling layer and then to standardize the inputs to be provided to a layer for each minibatch we use a batch normalization of 50. As the output would be binary (i.e. benign or malignant), the dense layer has 2 neurons.

For further fine-tuning the model we use Adam's Optimization Model to accelerate the gradient descent algorithm, and softmax Activation Function to transform the output into an array of probabilities, and to evaluate the performance of our model on our featured data we use Binary Cross Entropy.

The graphs below show how the accuracy of the model increases and the loss function decreases as we train the model for multiple iterations

Fig. 2 Training v.s. Validation Data Accuracy

Fig. 3 Training v.s. Validation Data Loss

V.MODEL PERFORMANCE EVALUATION

A. Performance Metrics

Listed below are a few widely used metrics to measure the performance of machine learning and Deep learning models. Also listed below is the performance of our model wrt to these metrics a)

1) Recall F1-Score and Precision Metrics

These metrics are extremely important to understand the outcomes of our model, and see the areas in which our model lacks. Recall shows us the True Positive, False Positive, True Negative, and False Negative predictions of the model. Precision value gives us the ratio of positive values that are correctly predicted to Actual positive values.

The F1 score is the weighted average of recall and precision, the accuracy of the model is directly proportional to the F1 value. For our model, the Precision, Recall, and F1 score for the detection of malignant breast cancer cells is 93%, 92%, and 92% respectively. The values of Precision, Recall, and F1 score metrics for detecting benign breast cancer cells are 88%, 87%, and 88% respectively:

2) Confusion Matrix

The confusion matrix is an essential metric to determine if the model had wrongly classified any data. the metric has 2 rows representing the predicted values vs. the actual class. This metric gives us a view of the false positive and false negative values.

A confusion matrix of our model shows us that our model accurately classified 230 malignant cells and 118 benign cells, at the same time it identified 30 false negatives and 20 false positives from the test data.:

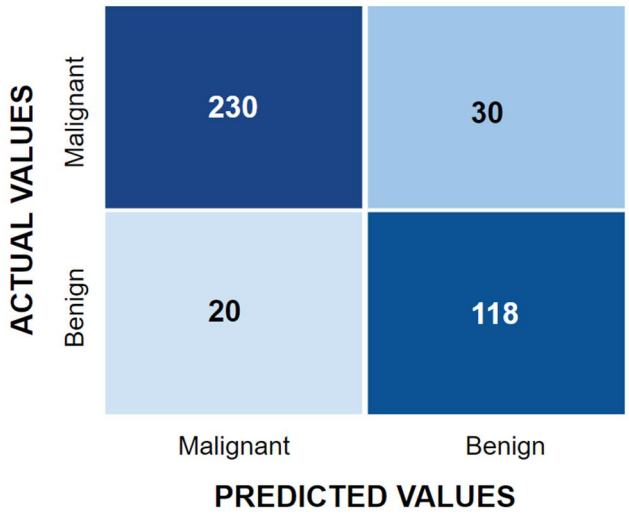


Fig. 4 Confusion Matrix

VI.OUTPUT AND RESULT

The results we get from training the model on a large dataset of the labeled image is that the model becomes incredibly accurate at detecting cancer cells based on the shape, texture, density, and profile of these cells These results clearly demonstrate that convolution neural networks utilizing an end-to-end training approach can be successfully utilized for classification of image and detecting breast cancer cells.

Shown below are a few outputs of our model on untrained images

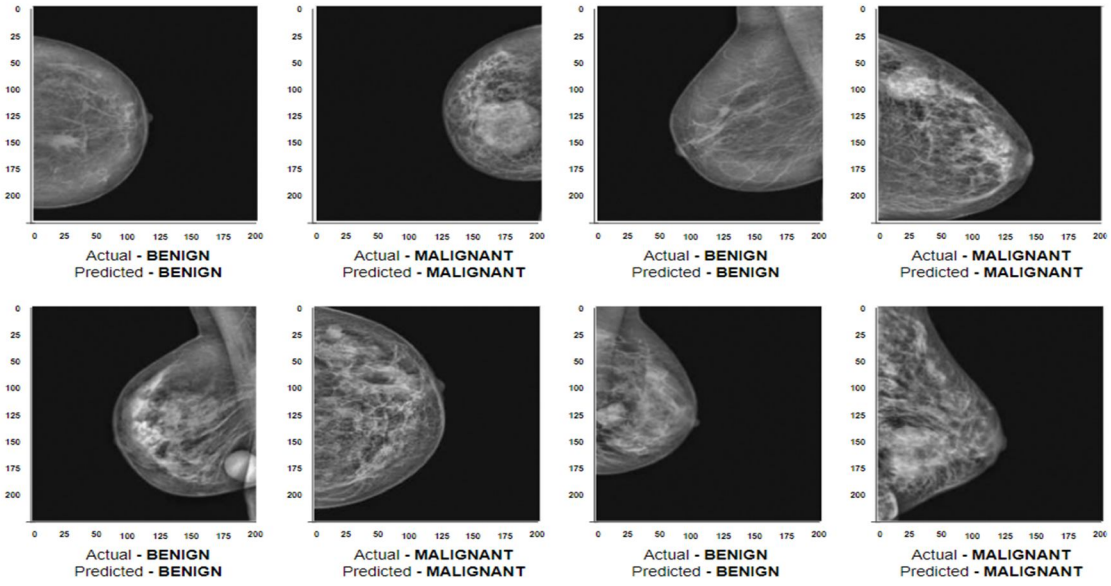


Fig. 4 Model Results

The extremely high accuracy of the model is a testament to the power of new advanced deep-learning models that help us solve problems and make the impossible, possible

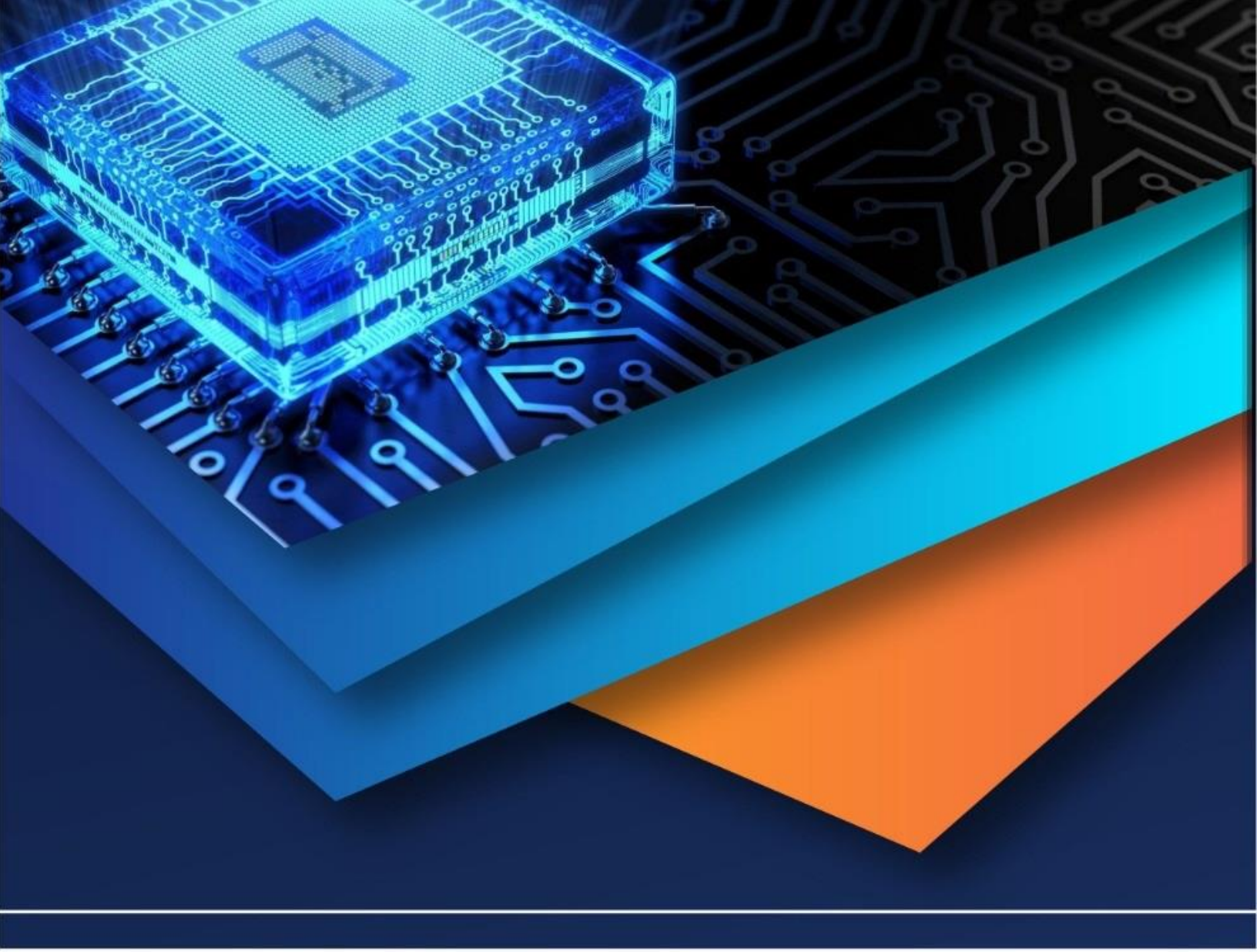
VII.CONCLUSION

The deep learning algorithm we developed has shown remarkable success in solving real-world problems by automating aspects of the medical world that were once thought impossible. Over the past few years, Machine Learning and Deep learning models have shown significant strides in detecting breast cancer and classifying various types of cancer cells [6].

We showed how utilizing a convolution neural network trained on mammogram images can help in classifying benign and malignant cells. The model can predict cancer cells with an extremely high degree of precision. This algorithm can be utilized to increase the accessibility of medical resources in regions where there is a lack of healthcare and medical facilities. This not only enhanced the diagnostic the current capabilities of the Computer Aided Diagnostics (CAD) system but also provided robust solutions for simplifying and automating several clinical practices

REFERENCES

[1] Mamatha Sai Yarabarla, Lakshmi Kavya Ravi, Dr. A. Sivasangari . "Breast Cancer Prediction via Machine Learning" In Third International Conference on Trends in Electronics and Informatics (ICOEI 2019).
[2] Uma Ojha, Dr. Savita Goel."A Study On Prediction Of Breast Cancer Recurrence Using Data Mining Techniques" In International Conference on Cloud Computing, Data Science Engineering.
[3] Li Shen, Laurie R. Margolies, Joseph H. Rothstein, Eugene Fluder, Russell McBride Weiva Sieh "Deep Learning to Improve Breast Cancer Detection on Screening Mammography" In Scientific Reports, Volume 9, 29th August 2019.
[4] Anji Reddy Vakka, Badal Soni, Sudheer Reddy "Breast cancer detection by leveraging Machine Learning" The Korean Institute of Communication and Information Sciences(KICS) 7th May 2020.
[5] Syed Jamal Safdar Gardezi, Ahmed Elazab, Baiying Lei, Tianfu Wang, "Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review", Journal of Medical Internet Research Jul 26, 2019.
[6] Abdullah-Al Nahid and Yinan Kong, "Involvement of Machine Learning for Breast Cancer Image Classification: A Survey", Hindawi - Published on 31 Dec 2017.
[7] J. Guo, S. Xu, D. Yan, Z. Cheng, M. Jaeger and X. Zhang, 'Realistic Procedural Plant Modeling from Multiple View Images,' - Published on 1 Feb. 2020.
[8] D. Li, G. Shi, W. Kong, S. Wang and Y. Chen, 'A Leaf Segmentation and Phenotypic Feature Extraction Framework for Multiview Stereo Plant Point Clouds,' in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 2321-2336, 2020.
[9] "Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis "Alexander Rakhlin, Alexey Shvets, Vladimir Iglovikov, Alexandr A. Kalinin



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY