JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Breast Cancer Classification using Transfer Learning with Ensemble Method

[1] **Sundari David**, [2] **Prof. Sandeep Kumar**

[1]M.Tech Student, [2] Professor
[1]*Dept of CSE*,
[1] *Cambridge Institute of Technology*, Bangalore, India.

Abstract : Breast cancer is one of the dreadful diseases that affect women globally. The occurrences of cancerous masses in the breast region are the main cause for women to develop a breast cancer. Early detection of breast mass will increase the survival rate of women. In the proposed work, a model is developed by using the Transfer Learning (TL) with ensemble method for classifying the breast cancer masses (BCM) as Benign or Malignant. The proposed system uses techniques such as VGG19, DenseNet201 which are CNN-based and Inception V3 for recognizing images. TL employs the formerly gained knowledge on the newest model and also solves the problem of building from scratch. The proposed system is applied on the MIAS (Mammography Image Analysis Society) dataset. The evaluation is done based on various performance metrics such as F1-score, accuracy, recall and precision.

***IndexTerms* - Breast Cancer masses (BCM), Transfer Learning, Ensemble, VGG1, DenseNet201, Inception V3.**

I. INTRODUCTION

There are two types of tumors in the human body. A benign tumor consists of non-cancerous cells that grow only locally and do not spread in the human body. A malignant tumor consists of cancerous cells that invade surrounding tissues. BC affects approximately 12% of women in India over their lifetime. Every two minutes, a woman in India is diagnosed with BC. This makes BC by far the most common type of cancer in women. Breast cancer is caused by the abnormal growth of breast cells.

It depends on the cells that become cancerous. BC can develop anywhere in the breast. There are three main parts of the breast: lobes, ducts, and connective tissue. Most BC develops in the ducts or lobules. Therefore, early BC detection is significant in increasing patient survival rates. Researchers have been looking for more precise models for cancer detection as cancer is associated with high morbidity and considerable healthcare costs. Mammography and biopsy are two of the most common methods of BC detection. Studies have shown that mammography has resulted in a reduction in death rates caused by breast cancer. Mammography uses a specific type of breast image to detect early signs of cancer in women. It is also possible to perform a biopsy for BC detection. The main challenges in viewing BC images are the variations in size, shape, and location of cancer cells.

II. RELATIVE WORK

Each algorithm performs in a different way depending on the dataset and the parameter selection. For overall methodology, KNN technique has given the best results. Naive Bayes and logistic regression have also performed well in diagnosis of breast cancer. SVM is a strong technique for predictive analysis and owing to the above finding, we conclude that SVM using Gaussian kernel is the most suited technique for recurrence/non-recurrence prediction of breast cancer [1].

For this work, we used at University of Wisconsin Hospital database which is composed of thirty values which characterize the properties of the nucleus of the breast mass. As we showed in result sections, DNN classifier has a great performance in accuracy level (92%), indicating better results in relation to traditional models. Random forest 50 and 100 presented the best results for the ROC curve metric, considered an excellent prediction when compared to other previous studies published [2].

The proposed model in this paper presents a comparative study of different machine learning algorithms, for the detection of breast cancer. Performance comparison of the machine learning algorithms techniques has been carried out using the Wisconsin Diagnosis Breast Cancer data set. It has been observed that each of the algorithms had an accuracy of more than 94%, to determine benign tumor or malignant tumor. it is found that kNN is the most effective in detection of the breast cancer as it had the best accuracy, precision and F1 score over the other algorithms [3].

The effectiveness of applied ML techniques is compared in term of key performance metrics such as accuracy, precision, recall and ROC area. Based on the performance metrics of the applied ML techniques, SVM (Sequential Minimal Optimization Algorithm) has showed the best performance in the accuracy of 96, 9957 % for the diagnosis and prediction from WBC dataset [4].

Comparison with different machine learning methonds such as, Naïve Bayes(NB), Decision tree(DT), Support Vector Machine (SVM), Vote(DT+NB+SVM), Random Forest(RF) and AdaBoost. Experimental data were downloaded from breast cancer Wisconsin dataset and using machine learning tool rapidminer. Using ten-fold cross-validation we found that the high accuracy of 96.99% with deep learning by Exprectifier activation function [5].

Most women's are suffered from breast cancer due to their unconsciousness. We used two ML algorithms here. By applying the algorithms we have got accuracies for example 74.73% accuracy of Random forest and 73.63% accuracy for XGBoost. We compared the results of our study with other existing systems and found that our system performed better than the existing system [6].

2.1 MACHINE LEARNING:

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

2.2 Transfer Learning:

Transfer learning is the process of applying a prior learnt model to a new situation. It's especially relevant in deep learning currently as it allows in training deep learning models with very little input. This is very useful in data science because most true scenarios don't really necessitate lakhs of labeled points of data to train complex systems.

Traditional ML vs. Transfer Learning:

Until now, traditional ML & DL algorithms were developed to function in seclusion. These methods have been programmed to address certain problems. When the feature-space composition alters, the systems must be reconstructed from the ground up. During transfer learning, the information of an earlier trained ML system is transferred to a separate but nearly related task.

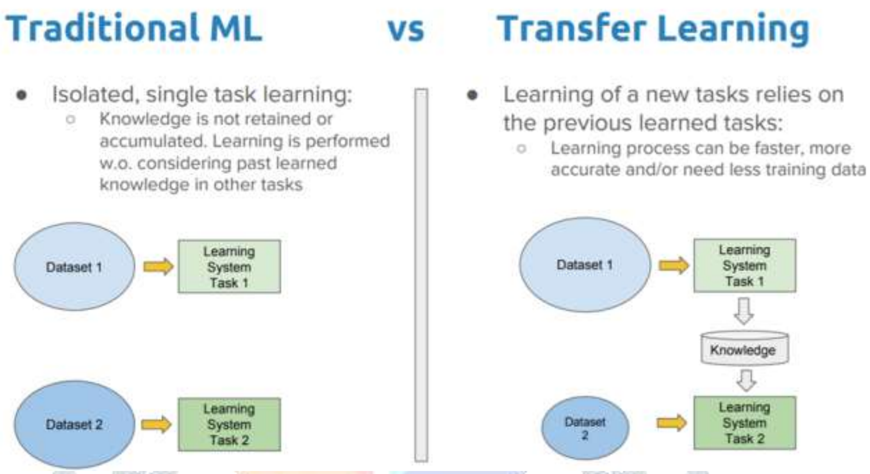

Fig 1: Traditional ML vs. Transfer Learning

Conventional learning systems are secluded and take place solely centered on particular activities, sets of data & training distinctively secluded systems on them. There is no expertise which could be passed on between one system to the other. Transfer learning allows one to use formerly trained models' information (properties, weighting, and so on) to train new models & also solve challenges like possessing less information for the newest assignment.

III. ALGORITHMS

3.1VGG19

VGG19 is a convolution neural network (CNN) architecture which was used to win ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2014. The ILSVRC is an annual computer vision competition. Each year, teams compete on two tasks. The first is to detect objects within an image coming from 200 classes, which is called object localization. The second is to classify images, each labeled with one of 1000 categories, which is called image classification.

VGG 19 was proposed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group Lab of Oxford University in 2014 in the paper "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION". This model won the 1st and 2nd place on the above categories in 2014 ILSVRC challenge.

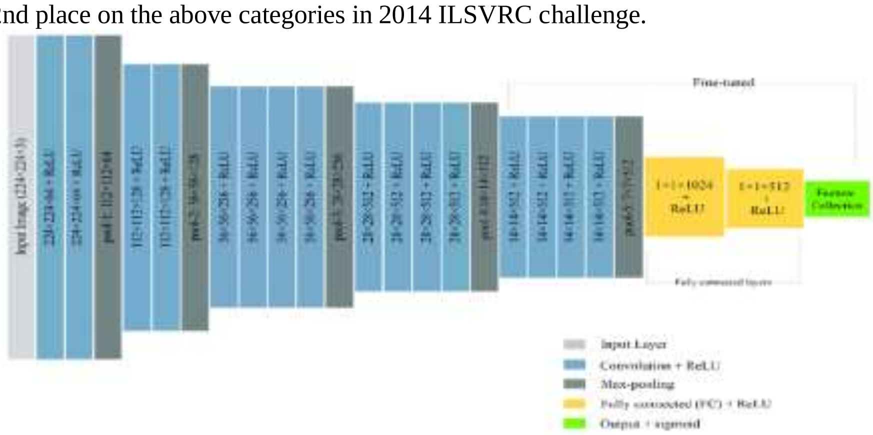

Fig 2: VGG19 Architecture

3.2DenseNet201

The DenseNet-201 comprises 201-layerings of CNN. One can import a pre - trained variant of such networks out from the ImageNet sets of data, which has been trained with over lakhs of photos. The system could categorize photos over thousand different object categories, including keyboards, desktops, pencils, and a variety of animals. As an outcome, the system has learnt about various high characterized forms for a variety of photos. The channel's picture intake size is 224 × 224 pixels.

Every layers of DenseNet receives extra info by all the prior layering and then sends its unique feature-maps to every following layers. The term "concatenation" is being applied. Every layer will receive "shared learning" from the layers above it.

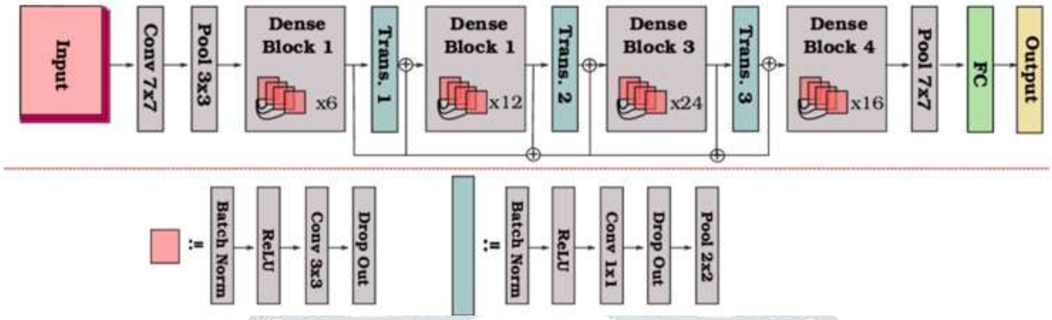

Fig 3: DenseNet201 Architecture

3.3InceptionV3

On the ImageNet datasets, Inception v3 is a picture recognizing system which has successfully proven to achieve higher than 78.1 percent accuracy. The system represents the result of several insights explored across time by a several scholars. It is inspired on Szegedy, et al's work, "Rethinking the Inception Architecture for Computer Vision."

Convolutional layers, average & max pooling, sequences, dropouts, & fully-connected layers are among the symmetrical & asymmetrical structural components in the system. Batch normalising is done to activating input datas & is utilized heavily across the system. Softmax is used to calculate the losses.

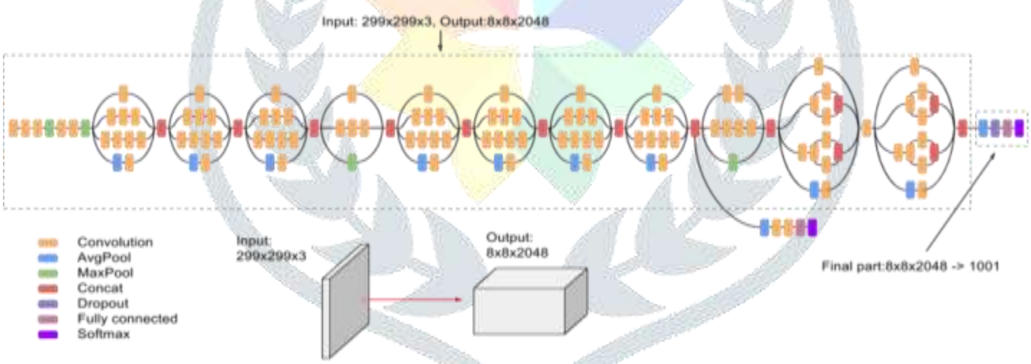

Fig 4: Inception V3 Architecture

IV. PROPOSED SYSTEM

The proposed system is used for detecting & classifying the BCM using the Transfer Learning algorithms VGG19, DenseNet201 and Inception V3 going through Ensemble Process. The classification is done on the MIAS Dataset

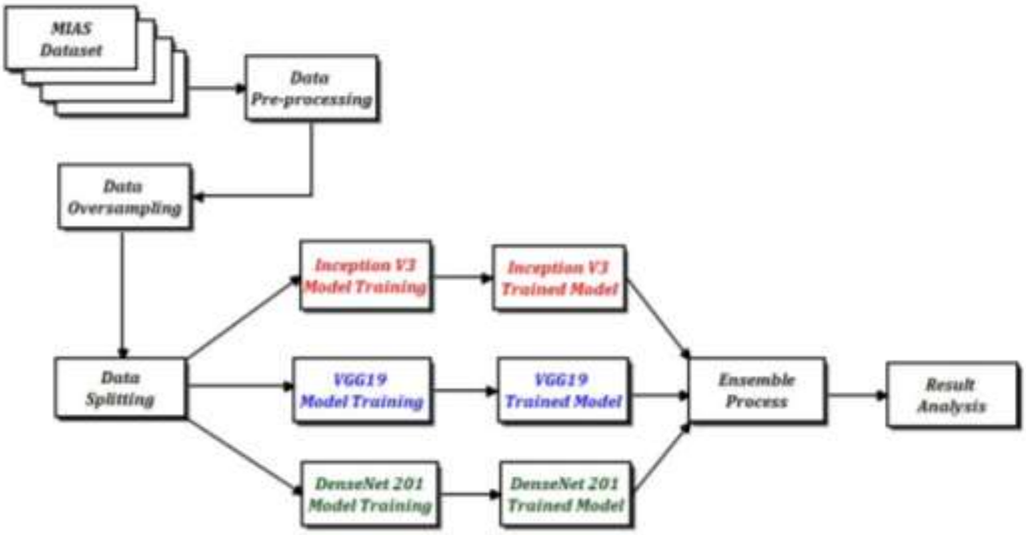

Fig 5: Proposed System Architecture

Ensemble learning method is a machine learning technique that combines several base models in order to produce one optimal predictive model and most of the times avoids misclassification. The result analysis is given by obtaining the classification report and confusion matrix for these algorithms based on the performance metrics such as F1-score, accuracy, recall and precision. The following are the steps involved:

4.1 Image Acquisition & Preprocessing

We are collecting the breast cancer image and storing it into our system. During image pre-processing, there may be artifacts in the images that should be corrected prior to feature measurement and analysis. In this module we will get the data from the online source. Further we will resize the image for future use.

4.2Data Oversampling

This step is used for imbalanced set of data; the oversampling process makes the minor category i.e., the Malignant one have the same no. of images such as Benign. For this we use the SMOTE oversampling algorithm which is referred as Synthetic Minority Oversampling Technique

4.3Data Splitting

We will divide the data into two sets in this step: the training set and the testing set.

4.4Model Training

An Ensemble learning approach is used in this proposed approach. We are Transfer learning algorithms. After splitting the dataset, the Train dataset will be used for model training. After model construction it is time for model training. We were able to build an ensemble learning based classifier that can recognize the breast cancer image either as benign or malignant.

V. DATASET

Here, we discuss about the set of data we have used. Lower viability X-ray pictures are employed in Mammogram to visualize mammalian existence in assessment and diagnosis.

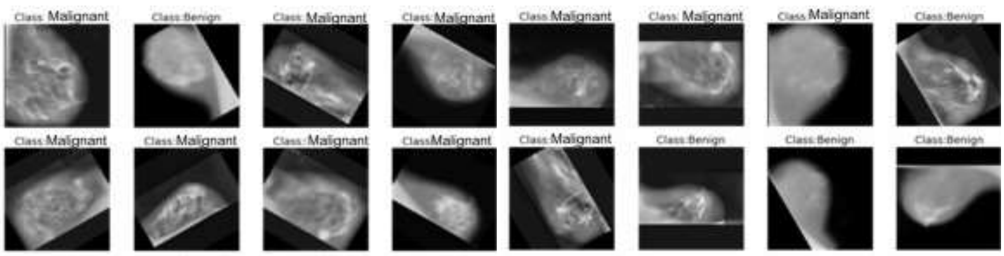

Fig 6: Sample Dataset Images for both Benign and Malignant Class

The MIAS (Mammography Image Analysis Society) is a scanned film of Mini Mammographic database developed by UK research groups for understanding of mammograms by generating a database of digital mammograms.

VI. RESULT AND ANALYSIS

The machine model will have to be tested against the validation dataset. This helps assess the accuracy of the model. Identifying the measures of success based on what the model is intended to achieve is critical for justifying correlation.

Table 1: Classification Report

Algorithm & class	Precision	Recall	F1-score	Accuracy
VGG19	82%	89%	73%	68%
DenseNet 201	88%	99%	93%	93%
Inception V3	75%	95%	84%	81%

The Classification Report and the Confusion Matrix for Transfer Learning algorithms VGG19, DenseNet201 and Inception V3 are obtained based on the various performance metrics such as F1-score, Recall, Precision, and Accuracy. The Table 1 shows the classification report for the proposed system.

The confusion matrix is a chart which shows how well a classifier performs on a tested set where the real numbers are revealed.

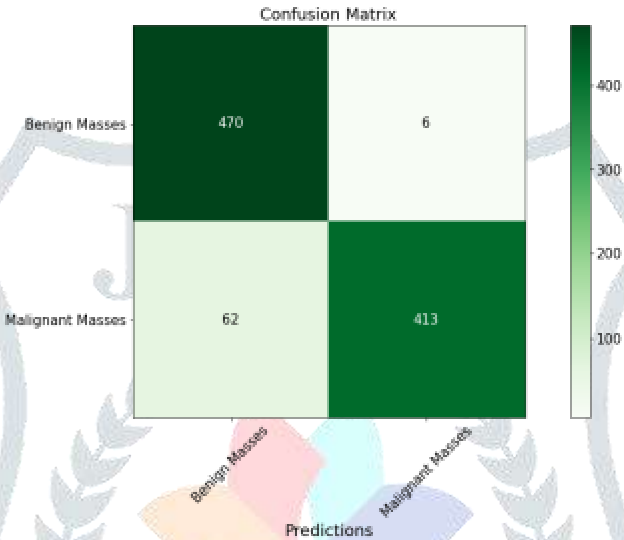

Fig 7: Confusion matrix for DenseNet201

The model accuracy and model loss graph for the trained model is obtained. The model accuracy contains training accuracy versus validation accuracy. The model loss also contains training loss versus validation loss. The Fig 8 shows the model accuracy graph for DenseNet201 and Fig 9 shows the Model Loss graph for DenseNet201.

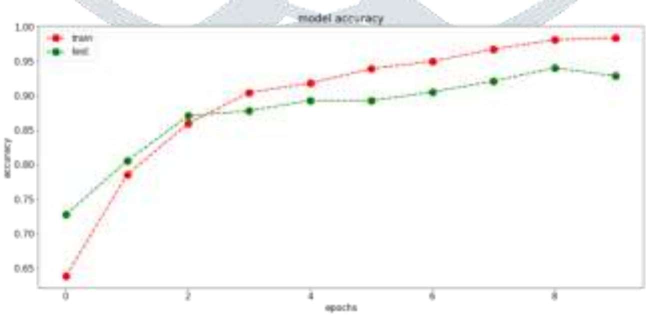

Fig 8: Model Accuracy for DenseNet201

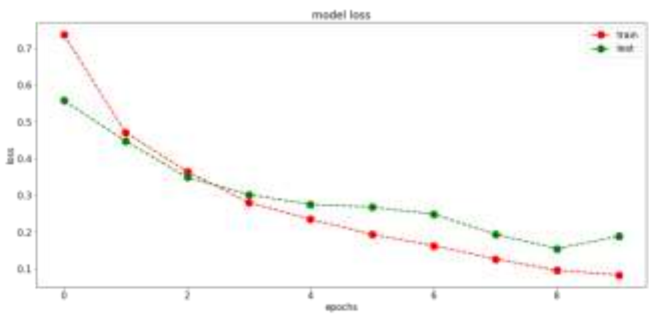

Fig 9: Model Loss for DenseNet201

VII. CONCLUSION

The output after the prediction shows that the Transfer Learning algorithms VGG19, DenseNet201 and Inception V3 after going through ensemble process are effective and can be utilized in classifying and detecting of BCM as Malignant or Benign Class. The MIAS Dataset which is a Mini Mammographic Database was used for classifying the BCM. The DenseNet201 algorithm performed well for the given dataset with good accuracy and classification results.

REFERENCES

[1] Bharat, A., Pooja, N. and Reddy, R.A., 2018, October. Using machine learning algorithms for breast cancer risk prediction and diagnosis. In 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C) (pp. 1-4). IEEE.

[2] Teixeira, F., Montenegro, J.L.Z., da Costa, C.A. and da Rosa Righi, R., 2019, September. An analysis of machine learning classifiers in breast cancer diagnosis. In 2019 XLV Latin American computing conference (CLEI) (pp. 1-10). IEEE.

[3] Sharma, S., Aggarwal, A. and Choudhury, T., 2018, December. Breast cancer detection using machine learning algorithms. In 2018 International conference on computational techniques, electronics and mechanical systems (CTEMS) (pp. 114-118). IEEE.

[4] Bayrak, E.A., Kırcı, P. and Ensari, T., 2019, April. Comparison of machine learning methods for breast cancer diagnosis. In 2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT) (pp. 1-3). IEEE.

[5] Mekha, P. and Teeyasuksaet, N., 2019, January. Deep learning algorithms for predicting breast cancer based on tumor cells. In 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON) (pp. 343-346). IEEE.

[6] Kabiraj, S., Raihan, M., Alvi, N., Afrin, M., Akter, L., Sohagi, S.A. and Podder, E., 2020, July. Breast cancer risk prediction using XGBoost and random forest algorithm. In 2020 11th international conference on computing, communication and networking technologies (ICCCNT) (pp. 1-4). IEEE.

[7] Ara, S., Das, A. and Dey, A., 2021, April. Malignant and benign breast cancer classification using machine learning algorithms. In 2021 International Conference on Artificial Intelligence (ICAI) (pp. 97-101). IEEE.

[8] Basunia, M.R., Pervin, I.A., Al Mahmud, M., Saha, S. and Arifuzzaman, M., 2020, June. On predicting and analyzing breast cancer using data mining approach. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1257-1260). IEEE.

[9] Khasana, U., Sigit, R. and Yuniarti, H., 2020, September. Segmentation of Breast using ultrasound image for detection breast cancer. In 2020 International Electronics Symposium (IES) (pp. 584-587). IEEE.

[10] Khuriwal, N. and Mishra, N., 2018, March. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In 2018 IEEMA engineer infinite conference (eTechNxT) (pp. 1-5). IEEE.

[11] Nagpure, R., Chandak, S. and Pathak, N., 2020, February. Breast Cancer Detection using Neural Network Mammogram. In 2020 International Conference on Convergence to Digital World-Quo Vadis (ICCDW) (pp. 1-6). IEEE.

[12] Nemissi, M., Salah, H. and Seridi, H., 2018, November. Breast cancer diagnosis using an enhanced Extreme Learning Machine based-Neural Network. In 2018 International Conference on Signal, Image, Vision and their Applications (SIVA) (pp. 1-4). IEEE.

[13] Sethi, A., 2018, August. Analogizing of evolutionary and machine learning algorithms for prognosis of breast cancer. In 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 252-255). IEEE.