WORLD JOURNAL OF PHARMACEUTICAL RESEARCH

SJIF Impact Factor 8.084

Volume 8, Issue 11, 720-730. Research Article ISSN 2277– 7105

ADVANCED DETECTOR FOR INVASIVE DUCTAL CARCINOMA USING DEEP LEARNING

N. L. Anbarivan*[1], V. Geethanjali[2] and V. Yeshwanth[3]

[*1]Master of Computer Science, VIT, Vellore, India.

[2]Bachelor of Technology, Peri Institute of technology, Chennai.

[3]California Software Limited.

Article Received on 26 July 2019,

Revised on 16 August 2019,
Accepted on 04 Sept. 2019,

DOI: 10.20959/wjpr201911-15822

***Corresponding Author**

N. L. Anbarivan

Master of Computer Science, VIT, Vellore, India.

ABSTRACT

The utilization of deep learning intelligence system to redefine medical assistance and support for decision making is shifting the paradigm of disease diagnosis. Clinical grade systems that are able to automate the process of detection of disease will help in effective decision making. Traditional methods of clinical screening and analysis suffers from various complexities varying from specificity to image noise. The objective of this research is to utilize deep learning intelligence system to redefine medical assistance and support for decision making and shift the paradigm of disease diagnosis. Clinical grade systems that are able to automate the process of detection of disease will help in effective decision making. We propose a reliable model that is able to capture cancer cells in images and support in making the process of treatment faster. Neural networks and transfer learning has been employed to detect the presence of cancer cells. In addition, automatic feature extraction and classification has been exploited to support the objective. We have achieved an accuracy of 95% in the test set and 94% in the training set and we have taken a step ahead to transform the idea of accessing care using our unique medbots.

KEYWORD: Breast cancer, intelligence prognostic tool, retrained models, InceptionV3, Automatic feature extraction.

INTRODUCTION

Breast cancer is the second deadliest disease that affects 8% of women in the world. It occurs due to the uncontrolled mutations in the genes present in the breast. It occurs commonly in lobules and ducts. Breast cancer show little or no symptoms at the early stages. Some of the symptoms include pain in the breast area, swelling, bloody discharge, scaling, peeling, inversion of the nipple, etc. Due to the heterogeneous structure, different types of carcinogens have different structural characteristics.

Several methods have been proposed for retrieving the cell information. Traditional methods of diagnosis involve the use of mammogram and biopsy. Mammogram used by pathologists still show low reliability as it depends on the size and degrades accuracy. Structure of the tissue can be unveiled using tissue biopsy performed using a microscope.

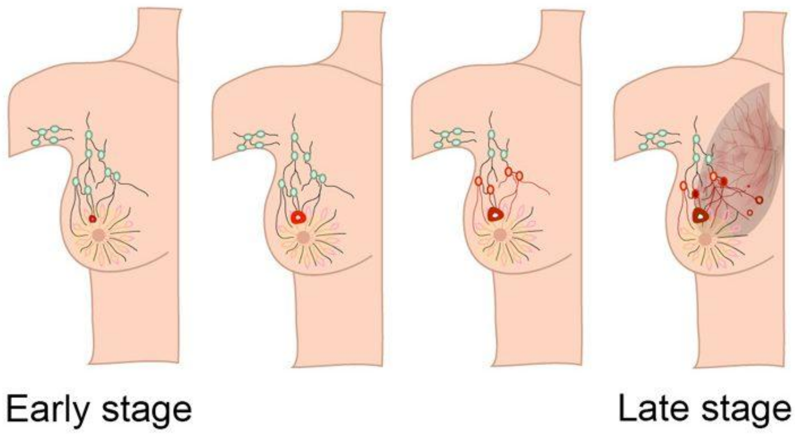

Figure 1: Illustrative Diagram of Tumor In The Breast.

Design of advanced intelligence system that have extensible diagnostic capabilities for recognition and classification of malignant cells from healthy ones is required for progression in biomedical research. Rate of survival depends upon the pace at which the disease is diagnosed. Also, the healthy and malignant cells are very identical and adds complexity to disease detection. Various kinds of transformations, segmentation as well as clustering techniques are exploited for the use case.

Evolving deep learning methods provide a platform to deliver clinical insights that clearly defines the objective. Neural networks based pre-trained model ensure reliability as well as accuracy when exploited properly. Using transfer learning will aid in development of a better tool to efficiently extract information from image and pave way to futuristic methods that can serve as a medium to save millions.

LITERATURE SURVEY

Table 1: Detailed View of Works Done In The Domain.

TITLE	FIRST AUTHOR	LITERATURE REVIEW	JORUNAL
Using deep learning for image-based plant disease detection	Mohanty	This paper details about the usage of deep learning model to predict crop disease. This model achieves an 99.35% accuracy for an apprehended test set and 31.4% accuracy when tested with online sources.	*Frontiers in plant science*
Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning	Shin	Several architectures of CNN as well as transfer learning are studied. Different comparative analysis is made in terms of layers, no of parameters, etc. It is trained with ImageNet model which is a pretrained one. Major CAD problems are also studied.	*IEEE transactions on medical imaging*
Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning	Rajpurkar	This paper primarily focusses on Pneumonia detection using Chest X-rays. They have developed their own architecture which is named as chestXNet that is trained using frontal view X-rays. It is able to detect 14 kinds of chest related disease and uses F1 metrics	*arXiv*
Deep learning models for plant disease detection and diagnosis	Ferentinos	In this paper, they have used CNN to classify healthy leaf image from diseased ones. About 87,848 images are used and 58 kinds of diseases were found. An accuracy of 99.53% was achieved using this model.	*.Computers and Electronics in Agriculture*
Automated identification of diabetic retinopathy using deep learning	Gargeya	This paper targets the leading disease- of diabetic retinopathy counteraction using deep learning techniques. Fundus images were used for the model building. They achieved an 0.97 AUC(area under the roc curve) and sensitivity of 94%	*Ophthalmology*

MATERIALS AND METHODS

Datasets

The dataset contains patches of the most common type of breast cancer, Invasive ductal carcinoma. The dataset is available in kaggle website and was created by Janowczyk and

Madabhushi and Roa et al. A total of 162 slide images were captured and converted to 277,524 patches of 50 by 50 dimensions.

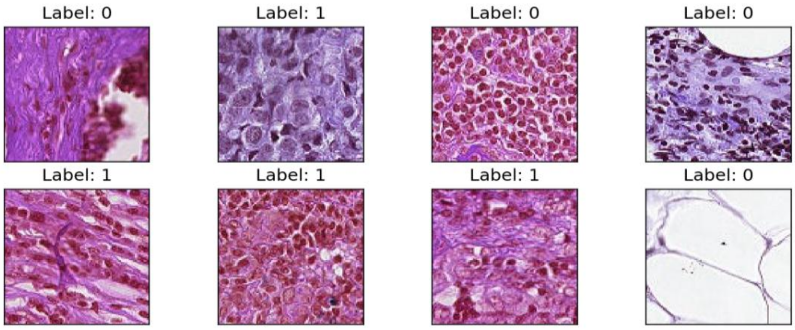

Figure 2: Sample Dataset (Label 0-Benign Cells and Label -1 Malignant Cells).

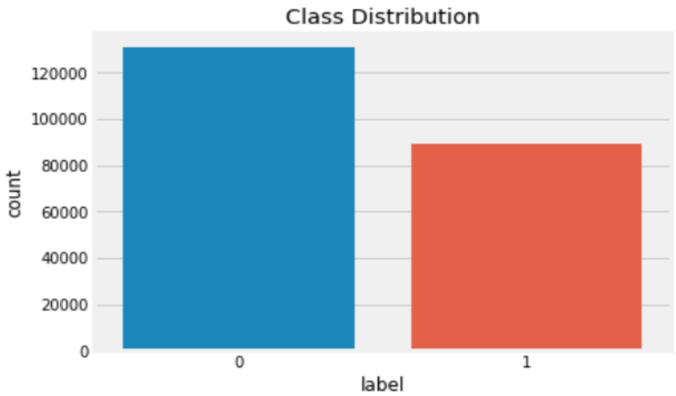

Figure 3: Distribution of the Classes.

Benign images were 198,738 in number and malignant images were about 78,786. They were categorized into different folders. Analytics of the image data indicates the presence of imbalanced dataset with twice the number of benign samples compared to malignant ones.

Preprocessing

This step is crucial for laser focus of the region of interest(ROI) and better qualitative image. The size and quality of dataset is enhanced using data augmentation techniques. This step will increase the reliability as well as performance of the model. Also, this will add strength to the generalization of the problem and reduces overfitting as well regularize the problem. Following are the augmentation techniques that were used.

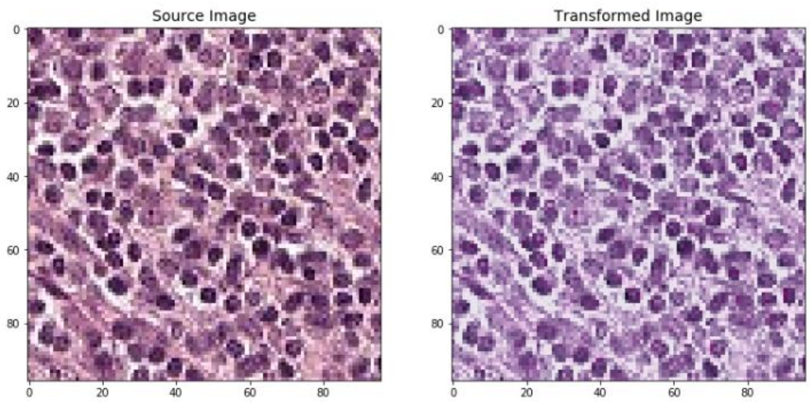

Figure 4: Depiction of the Original Image And The Transformed Image.

Image data generator, a data pipeline was used for this model. A series of flipping, rotation, zooming, shearing, etc. Were done in the pipeline process. Random directional rotation was done using the rotation range parameter. Horizontal and vertical translation of the image was done using width and height shift respectively. Shear range is used to decide the panoramic distortion angle. If there were images with empty border, then fill mode assigned to reflect was used. This ensures borders are filled using the image's reflection. Also image noise must be removed. Denoising techniques were carried out to remove background noise with inbuilt capabilities of preserving as much information possible. Standard normal distribution of the images was implemented using sample_wise_normaization=True. Vertical flip was done to increase the size of the images. This step increases the size of training data as well as the quality of the results due to its highly variant image processing methodology.

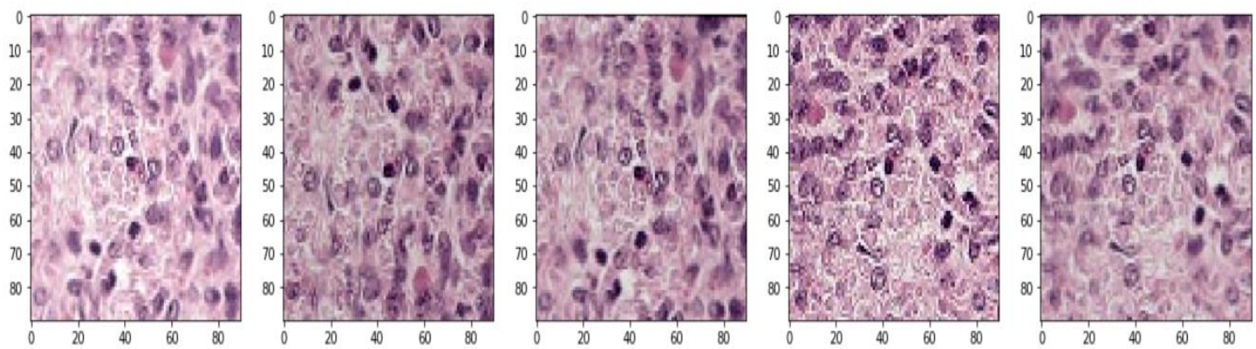

Figure 5: Random Augmentations of the Same Image.

Feature Extraction

Pretrained models served two purposes-classification and feature extraction. Features were the entities that clearly distinguished the healthy from the cancerous cells. We used modified inceptionv3 for this process available in Keras framework.2048 channeled convolution layer in the end of the network was converted to one dimensional vector using global average pooling and then images were fed for automatic feature extraction.

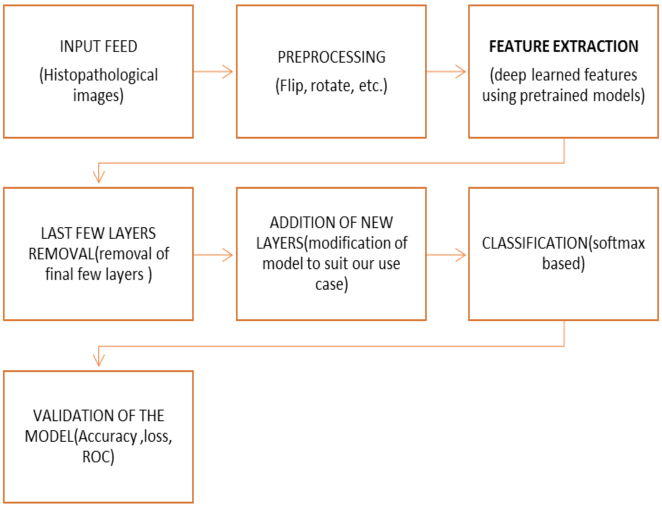

Figure 6: Flowchart Depicting The Steps Involved In Breast Cancer Image Classification.

Model Building

Initially models were built from scratch using convolutional neural networks (CNN). Weights were initialized randomly and this led to over fitting as there was no learning after few epochs. Fine tuning of the hyper parameters also did not show difference in the model performance.

Since building CNN's from scratch was computationally expensive, hence transfer learning was adopted for this method. Transfer learning involved use of pre trained models with assigned weights that could be used for a specific problem. Keras framework was used for this task. Keras runs on top of tensor flow and is a high level API (Application program interface) which is easy, fast and reliable.

Modified Model

The binary class classification problem was solved using a modified inceptionV3 based network that has additional layers to improve efficiency and for reliable outcome delivery. After image augmentation is carried out, the patches are fed into our model. GPU'S (Graphical processing units) were used to fine tune the model where batches of images were fed into the model. Replacement of output layers is done and preassigned weights were implemented. Model is trained for 20 epochs with incremental learning rates.

Adam optimizer and sigmoid function were used. The inceptionV3 model is one of the complex model requiring high computational power. It takes more time to train the model. So we use transfer learning that only uses the predefined weights of InceptionV3. The initial layers are frozen and only the last few layers that were added are trained again. The additional layers include an alternate set of dropout layers and fully connected layers. Additionally, only the learning rates of newly added layers are modified.

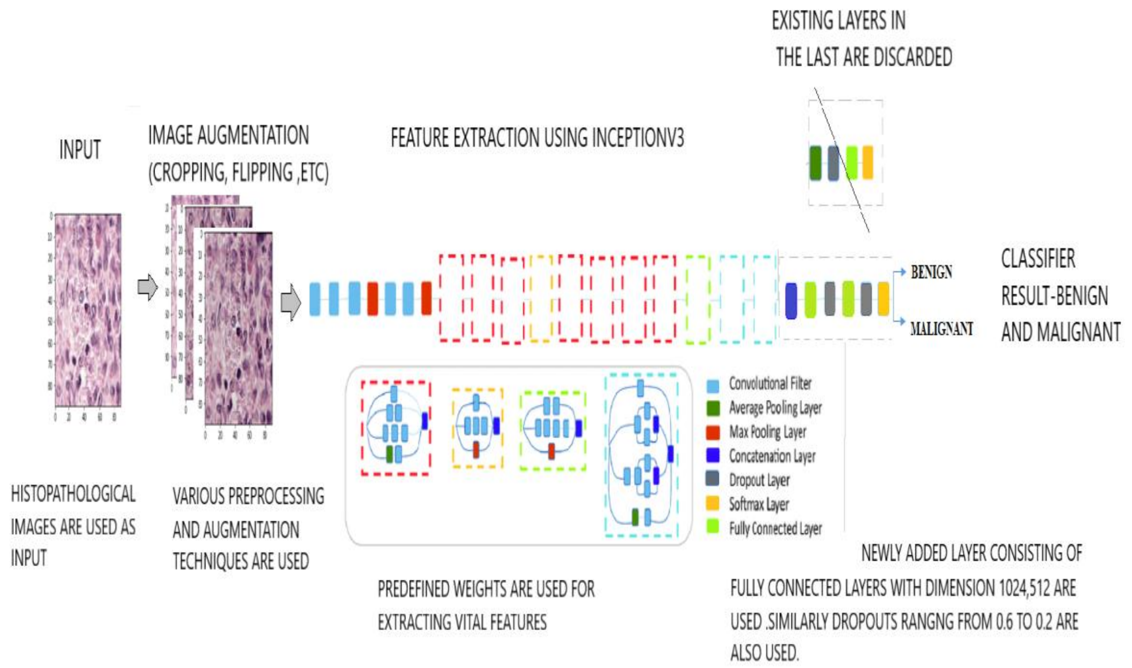

Figure 7: Detailed Illustration of Our Model.

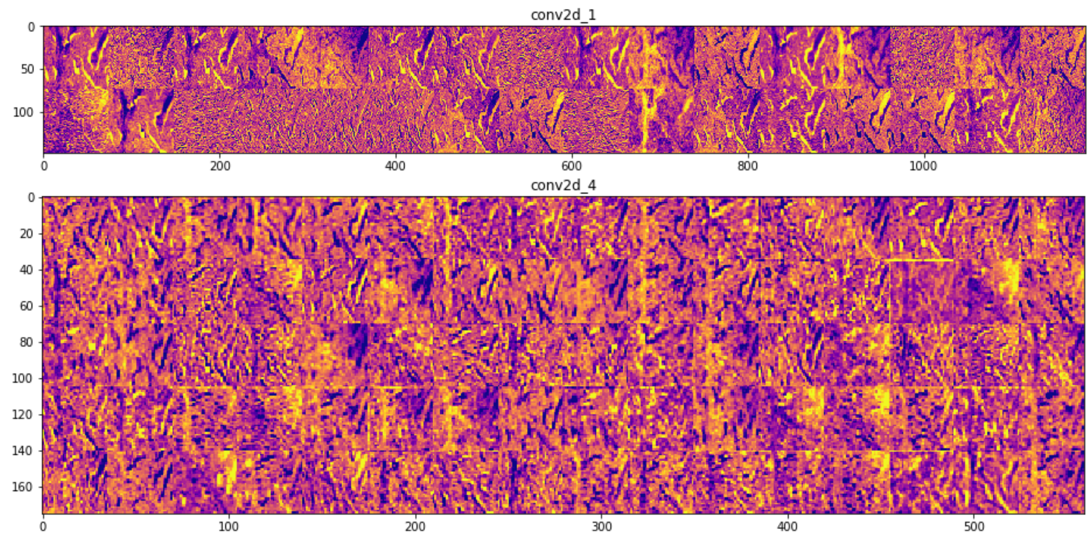

Figure 8: Deep Virtualization of Various Layers In The Model.

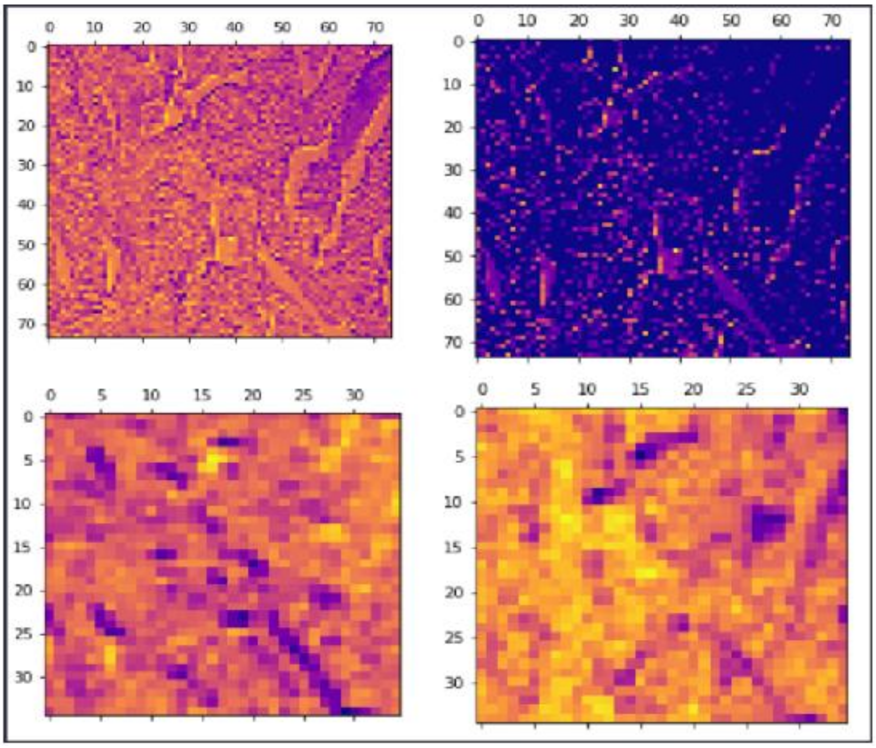

Figure 9: Visualization of Various Layers.

RESULTS

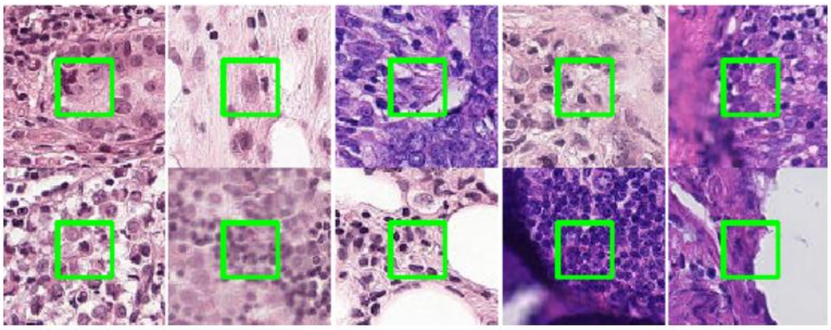

Figure 10: Cancer Images that are Classified by the Model.

This section is a concise delivery of results of our modified model. Also, the results from our model is compared with two other models. The models that are compared are model built from scratch and pretrained model built using vgg16.

Table II: Comparison of train/test accuracy of the models.

Model	Training accuracy (in %)	Testing accuracy (in %)
Simple CNN	64	60
VGG-16	89	86
Our model	**95**	**94**

The results clearly define our model as a better model to find cancer cells in the breast. Simple CNN AND VGG-16 suffers from overfitting problem and hence cannot be categorized as a generalized model.

Graphs representing the test and train accuracy is shown below.

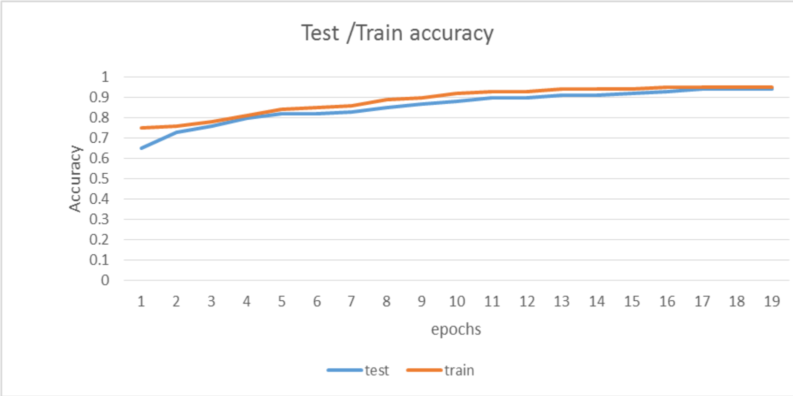

Figure 11: Graphical Representation of Test/Train Accuracy of Our Model.

Red line represents the train accuracy and blue line represents the test accuracy.

Loss –

Figure 12: Graphical Representation of Test/Train Loss of our Model.

Red line represents the train accuracy and blue line represents the test accuracy.

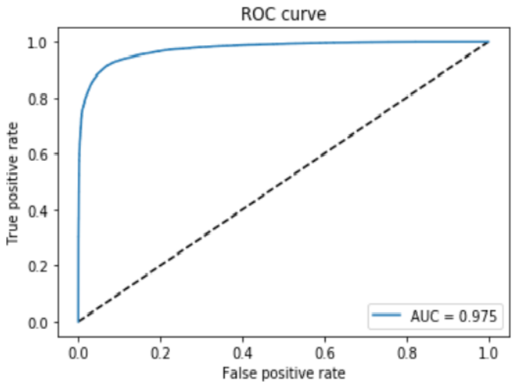

Figure 13: ROC for the Given Model.

ROC (receiver operating characteristic curve) is one of the important metrics used in this domain. It validates the reliability of binary classification system. Here the model achieves an ROC of 97.5 %.

DISCUSSION

Results show an accuracy of 95%. High performance of the model is obtained because of robust feature extractor, agile layers of the modified model and optimized preprocessing techniques. Key driver for high accuracy is the image processing techniques. Huge amounts of computational power as well as time is saved using pretrained model.

The model outperforms other CNN models in terms of reliability, efficiency and accuracy. In future, we plan to ensemble high performing pretrained models so we can achieve an automated system with increased reliability and efficiency.

CONCLUSION

The diagnostic capabilities of the model in cancer cell detection has helped in development of a systematic tool that not only detects disease but also extracts key features and is better than human analytics system. Effective interpretation of the model as well as the promising accuracy of 95% will guarantee timely treatment of the disease and primarily focuses on the people living in remote, village areas. The tailored approach used for image based disease identification has helped in decision support for doctors and also serves as a resource for cancer control.

REFERENCES

1. Mohanty, S. P., Hughes, D. P., & Salathé, M. Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 2016; *7*: 1419.
2. Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I. & Summers, R. M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 2016; *35*(5): 1285-1298.
3. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T. & Lungren, M. P. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.
4. Ferentinos, K. P. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 2018; *145*: 311-318.
5. Gargeya, R., & Leng, T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 2017; *124*(7): 962-969.