

Predicting Invasive Ductal Carcinoma Tissues in Whole Slide Images of Breast Cancer by Using Convolutional Neural Network Model and Multiple Classifiers in Google Colab

Deepa B G (✉ deepabg08@gmail.com)

REVA University <https://orcid.org/0000-0003-4952-8861>

S. Senthil

REVA University

Research

Keywords: Breast Cancer, Invasive Ductal Carcinoma, Histopathological Images, Convolutional Neural Network, Google Colab, Deep Learning, Keras Sequential Algorithm

Posted Date: May 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-478654/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Breast Cancer (BC) is the common type of cancer found in women which is caused due to the abnormal growth of cells in the breast. An early BC detection helps to increase the survival rate of the patient and 80% BC type was Invasive Ductal Carcinoma (IDC). In this work, a deep learning-based IDC prediction model is proposed with multiple classifiers and CNN (Convolutional Neural Network). The developed deep learning method used a sequential Keras model like conv2D, Maxpooling2D, Dropout, Flatten and Dense. The multiple classifiers are LR (Logistic Regression), RF (Random Forest), K-NN (K-Nearest Neighbors), SVM (Support Vector Machine), Linear SVC, GNB (Gaussian NB) and DT (Decision Tree). The CNN model generated by using SkLearn, Keras and Tensor flow libraries, and results are organized by MatPlot libraries. At the classification stage, a helper function was defined, and Google Colab online browser platform used for developing the proposed model. The performance is analysed in terms of Accuracy, Precision, Recall, F1-score and Support.

1. Introduction

BC is the most common type of cancer found in women, which are categorized based on the nature of affected cells. The three significant parts of the breast are connective tissue, ducts and lobules. Most of the BC found in either ducts or lobules and common type of BC are Invasive Lobular Carcinoma and IDC [1]. According to WHO (World Health Organization), there are 1.38 million BC cases reporting every year. Among them, 458,000 death cases are registered. [2]. In India, every four minutes one woman was diagnosed with BC [2]. There are various factors affecting the treatment and diagnosis of BC [3].

The major types of BC surgeries are Lumpectomy, Mastectomy, Sentinel Node Biopsy, Axillary Lymph Node Dissection and Contralateral Prophylactic Mastectomy[4]. The common types of BC are DCIS (Ductal Carcinoma In-Situ), LCIS (Lobular Carcinoma In-Situ), Invasive Ductal Carcinoma(IDC) and ILC (Invasive Lobular Carcinoma) [5]. The WSI (Whole Slide Imaging) are mainly utilized in the field of clinical education and digital image analysis. These digital images organized with digital slides, and used for segmentation, BCa prognosis, Breast slide classification and immunohistochemical (IHC) quantification [6].

The major limitation of WSI is the quality of the image and high processing time. The estimation of ROI (region of interest) in the malignant sample image is a difficult task. The WSI used in various research field [7]. The deep learning(DL) model utilizes WSI at the stage of processing and used to understand various data features like image, text and sound. The different phases of traditional ML are pre-processing, feature extraction, feature selection and finally, classification [8]. Due to the introduction of various ML (Machine Learning) techniques, physician can easily detect cancer with high level of accuracy [9]. There are lots of DL, and ML approaches are used in the field of BC diagnosis [10].

Our main contributions are given below.

- To design a IDC prediction model in WSI using CNN.

- To compare proposed model with multiple classifiers.

The main objective of this work is to introduce an IDC prediction model using CNN. The remaining section of this paper arranged as follows. Section II explain the related works, Section III discuss the proposed methodology, Sect. 4 represents the result and discussion, and Section V gives the conclusion.

2. Related Works

Hamidreza Bolhasani et al. [11] described about the IDC cancer type and estimated that 80% of cancer cases in women belong to the IDC category. For the IDC analysis purpose, 922 histopathological images selected from 124 patients. The developed dataset was named by Databiox, and image resolution was 2100 x 1574. There are five steps for the preparation of DataBiox, which are the selection of IDC patients, collection of H&E strained slide, specimen labelling, microscopy imaging and finally, dataset organization.

Angel Cruz-Roa et al. [12] introduced an automatic approach for detecting malignancy region on breast. For analysing the part of the tumour, CNN was utilized and estimated with the help of 113 slides. The value of accuracy and F-measure obtained from the introduced method are 84.23% and 71.80%, respectively. RGB (66.64% and 77.24%) and fuzzy colour (67.53% and 78.74%) histograms provided accurate handcrafted features. The IDC diagnosed with the help of The Cancer Institute of New Jersey and the Hospital of the University of Pennsylvania.

Sumaiya Dabeer et al. [13] discussed the role of DL algorithms in the field of BC diagnosis. The proposed method utilized CNN for extracting features and fully connected network for classification. The precision accuracy obtained from the introduced method was 99.86%. The development of ML and DL in the field of computer-aided diagnosis helps to increases the extraction process faster and error -free. For reducing the size of the image, an encoder was developed, which allows to compress the data without reducing prominent features. It helps to maintain 90% of the originality of the images.

Shikha Roy et al. [14] introduced a two-class MI approach for differentiating the stages of IDC. The initial and final stage of IDC determined, and the model trained by using RNA-seq gene from 610 patients. During the stage of classification, data size reduced. The introduced web - based Duct-BRCA-CSP method helps for predicting cancer at the early stage. The proposed technique was based on RNA-seq gene expression. The utilized server was [http://bioinfo.icgeb.res.in/duct- BRCA-CSP](http://bioinfo.icgeb.res.in/duct-BRCA-CSP).

Yibao Sun et al. [15] developed a DL based automatic approach for estimating border of DCIS. For predicting the DCIS tumour border CNN was trained. The probability map is generated by using sliding window method. The DCIS border was estimated at the final stage. The Accuracy, F1-Score, Mean of the proposed method was 98.46% ,97.4% and 95.6%, respectively. Finally,95.6% of tumour tissue was identified.

Aiza M. Romano et al. [16] proposed DL based architecture for predicting BC in IDC ,and CNN trained with 32 filters of size . Each image is used for improving prediction accuracy and F-Score. The accuracy and F-Score of the proposed work was 85.41% and 85.28%, respectively. The model setup was free from the overfitting issue. For achieving better results, the number of filters increased from 32 to 64. During the classification stage, DL method was applied. The Precision, Specificity, Sensitivity and F-Score are the performance evaluation parameters.

Koji Takada et al. [17] discussed the risk factors of invasion and diagnosed from either VAB (Vacuum-Assisted Biopsy) or CNB (Core Needle Biopsy). After the analysis, 25.9% of reported IDC analysis cases are CNB, and 118 patients are analysed.

Laith Alzubaidi et al. [18] discussed the DL models for diagnosing BC and main problem encountered with previous techniques is deficiency of the training data. Hence the transfer learning technique was introduced. The proposed hybrid CNN method consists of 74 layers. During the classification stage, biopsy images are categorized into 4 types- Invasive Carcinoma, Carcinoma in suit, Benign Tumor and Normal Tissue. The accuracy value was 97.4%

Hesham Alghodhaifi et al. [19] introduced two CNN models. One was depth-based, and another was standard convolution. These models used for increasing CNN accuracy and for identifying IDC in histology images. The tested activation functions are ReLU, Sigmoid, Tanh and Gaussian noise, and used to check the robustness of both models. The accuracy, sensitivity and specificity are 87.5%, 93.5% and 71.5% respectively.

3. Proposed Methodology

BC is an invasive type of cancer. The ducts and lobes are the parts of the Breast. Most of the BC found in the duct cells. The presence of red, swollen and warm in the breast area are the indication of BC. The growth process of BC depends on the health status and age of the person. For detecting the various stages of BC, different techniques like- CT Scan, PET Scan, Chest X-Ray and Bone Scan used. Invasive and Non –Invasive are the two categories of BC. Among them, the most dangerous cancer is the Invasive type. Because it initially located at one place and gradually spreads to the other body area. LCIS, ILC, IDC and DCIS are the four different kinds of BC and theses types are represented in Fig. 1.

Figure 1 represents the various type of Breast Cancer. DCIS and LCIS are non-invasive type breast cancer. The DCIS usually found in the duct area, and LCIS is identified at the lobules. The IDC type mainly affects the milk duct, and their surrounding area and ILC generated at the place of lobules.

3.1 Prediction of IDC in the WSI using CNN

In this work, an accurate IDC BC prediction model proposed with CNN and multiple classifiers (LR, RF, KNN, SVM, Linear SVC, GNB, DTC). The accuracy level of these classifiers is estimated with the help of

CNN and diagnosis of cancer achieved with the help of WSI (Whole Slide Image). The Flowchart of the proposed deep learning model shown in Fig. 2.

Fig.2 represents the flowchart of the proposed model. The initial stage is the loading process of the image dataset. After that dataset divided into two sections-IDC+, and IDC-. The IDC- is used for representing the class zero image and IDC+ is used for indicating class one image. IDC+ denoted by one, and IDC- is marked by 0. The image is resized, and labelled. These processes belongs to the pre-processing phase. The next stage of the proposed model is a data processing, and it is composed of data splitting, data normalization, data flattening and data rebalancing. The dataset divided into two parts (testing and training) at the stage of data splitting, and it helps to increase the accuracy of the model. Then data is organized for providing compatibility for considering classifiers (RF, LR, SVM, KNN, GNB, CNN, Linear SVC, DTC and CNN). The data flattening stage used for converting the 2D array into the 1D array. After that data sampled for balancing imbalanced datasets. Random Undersampling and Random Oversampling are the two methods used for resampling. In this proposed work, IDC+ and IDC- are balanced to deal with the imbalanced classes by reshaping of data. The representation of imbalanced class and balanced class shown in Fig 3 and Fig.4.

After the data balancing, K-fold cross-validation utilized for checking the accuracy of multiple classifiers. The multiple classifiers fed with the input of the different size, and helper function is defined. Finally, plotting the accuracy value and comparison chart for the numerous classifiers. Further, the model accuracy and loss of CNN plotted with the help of epochs.

3.2 The proposed algorithm

The proposed algorithm described the below section.

Step 1

Input the histology IDC image data set. To retrieve the files of the specified pattern `glob ()` python module used. The rules of `glob` are similar to the Unix path, which identify matched pathnames.

Step 2

Data Pre-processing

2.1: Data Separation: Separating the data set into IDC- and IDC + classes. The Python class used in this process is `fnmatch. filter ()`. This function used to return the Boolean value based on the matching pattern.

2.2: Image Resizing: Resizing of the image by using python class `cv2.resize ()`, in terms of dimensions, height, width, by preserving the image's originality in the resized image.

2.3: Image Labelling: If there is a IDC presence, the value is "1" else "0". The `pd.DataFrame ()` is used for creating data frames.

Step 3

Data Processing

3.1: Data Splitting: Splitting of data into test and train subsets are done using the python class, `sklearn.model_selection.train_test_split()`, which helps to provide random partitioning of subsets and increase the model accuracy by reducing the overfitting.

3.2: Data Normalization: The data from 0 to 256 scaled to 0 to 1, which helps in data compatibility for a wide variety of classification algorithms.

3.3: Data Flattening: Conversion of the 2D array into 1D array done using python class, `numpy.ndarray.flatten()`. A copy of the input array flattened using this class. Another method used for data flattening are `np.ravel()` and `np.reshape()`.

3.4: Data Rebalancing: This method used for resampling the dataset for imbalanced classes by using two methods-Random Oversampling and Random Undersampling. The python class used is `imblearn.under_sampling.RandomUnderSampler()`. During the Random oversampling, the minority class examples are duplicated, which results in overfitting. In the case of Random Undersampling eliminating minority class examples.

Step 4

Data Modelling

4.1: Data Model Fitting: By using K-cross validation, the model fits with different classification models and CNN, for checking the model accuracy.

4.2: Data Plotting: The python class `matplotlib.pyplot.figure()` is used for plotting data. The `matplotlib` is a mathematical extension of `Numpy`. Accuracy and loss of the plotting model are made with the help of epochs number. All class-zero images is defined by IDC- and class -one image are defined as IDC+.

4. Result And Discussion

The Cloud Platform Google Colab and Histopathological Image Dataset used for the implementation purpose. The dataset consists of 277,524 RGB images with a resolution of pixels which derived from 162 H&E -stained samples of breast histopathology. The CNN trained using a large volume of image patches which obtained from the WSI of BC. The image patches are small and mined from the digital samples of tissues. The actual data scaled from 0 to 256. The proposed work is scaled from 0 to 1 for analysing various classification algorithm. The data set is divided into two parts - test and training. The splitting of data helps to increase the model accuracy by reducing the overfitting. The proposed model analysed with two various splitting 70 – 30 and 80 – 20. After analysis a few of the tissues are cancerous cells. The dataset exported by using <https://www.kaggle.com/paultimothymooney/breast-histopathology->

images".The dimension transformation procedure referred to from <https://www.geeksforgeeks.org/python-flatten-a-2d-numpy-array-into-1d-array/> .

4.1 Simulation Results

The proposed model is carried out with two different splittings (training and testing) and uses image patches of 1000,2000,3000,4000 and 5000 for both splits. The model accuracy and model loss predicted based on epoch numbers for various data size. The epoch used to measure the number of times the training vector used before updating the weights. In this proposed neural network, eight epochs used, and the relative analysis of results with a different dataset on CNN shown in Table.1.

Table 1. Relative Analysis of results with different data set on CNN (70 – 30)

Dataset Prediction	Image Count	Accuracy of CNN (%)
70 – 30 Training-Testing data set split	1000	86.08
	2000	85.03
	3000	85.15
	4000	78.71
	5000	80.39

From Table.1. we can see that CNN gives better accuracy (86.08%)at the lowest image count (1000). The accuracy level of 2000,3000,4000 and 5000 image counts are 85.03,85.15,78.71 and 80.39 respectively. The considered data split is 70 – 30(train -test).

Figure 5 represents the performance evaluation of CNN for 70 – 30 Split. Totally, 8 epochs used. The precision, recall, F1-score and support of IDC (-) in 70 – 30 split are 0.90,0.81,0.85 and 115, respectively. The IDC (+) gives recall, precision, support, and F1-score are 0.91,0.90,115 and 0.87 respectively.

Figure 6 represents the value accuracy curve of epochs for 70 – 30 split. The initial validation value is 0.50, and the training initial value is 0.54. The training value of the accuracy curve shows a larger spike after the epoch number is 3. The validation value gradually increases after the epoch number is 1. The validation value of 1,2,3,4,5 and 6 are 0.50,0.50,0.50,0.84,0.50 and 0.85 respectively. At the final epoch, the training value is 0.6727, and the validation value is 0.8609.

Figure 7.represents the model accuracy and model loss curve of the test-train with 1000 images. The number of epochs used for model accuracy and model lose is 8. The initial training value of the model loss is 0.78, and the initial testing value is 0.76. The model loss of 2, 3, 4, 5, 6 and 7 are 0.67, 0.64, 0.68, 0.60, 0.64 and 0.56 respectively. As the number of epochs is increases, loss of the model is reduced. The 70 – 30 data split is used for the model lose and model accuracy.

Table 2
Relative Analysis of results with different data set on CNN (80 – 20)

Dataset Prediction	Image Count	Accuracy of CNN (%)
80 – 20 Training-Testing data set split	1000	80.51
	2000	82.41
	3000	83.06
	4000	84.54
	5000	84.98

Table.2 represents the relative analysis of result with different data set on CNN with an 80 – 20 data split. The CNN gives a better accuracy value (84.98%) at the 5000 image count. When the image count is 1000, the accuracy of CNN is 80.51%. The model accuracy increased as the image count increase. The CNN accuracy of 200, 3000 and 4000 image count are 80.51, 82.41, 83.06 and 84.98%.

Figure 8 represents the performance evaluation of CNN for 80 – 20 Split. There are eight epochs are used. The precision-recall, F1-score and support value of IDC (+) are 0.94, 0.74, 0.83 and 323, respectively. The IDC (-) gives recall, precision, support and F1-score are 0.96, 0.79, 323 and 0.86 respectively.

Figure 9 represents the value accuracy curve of epochs for 80 – 20 split. The initial accuracy value of training is 0.62, and the validation value is 0.52. The accuracy value of validation for 1,2,3,4,5 and 6 epochs are 0.79,0.85,0.75,0.89,0.86 and 0.87 respectively. The final epoch value of validation and testing are 0.8557.

Figure 10 represents the model accuracy and model loss curve of the test-train with 5000 images. Totally eight epochs are used for training and testing the model accuracy and model loss. The testing model loss value of 1,2,3,4,5,6,7 and 8 epochs are 0.77, 0.45, 0.36, 0.50, 0.31, 0.33, 0.31 and 0.35 respectively. The model loss decreases as the number of epochs increases. The 70 – 30 split with 1000 images gives better accuracy performance (86.08) than the80-20 split with 5000 images. The 70 – 30 split is suitable for less image dataset, and the 80 – 20 split suitable for a larger dataset.

Table 3. Comparison of multiple classifier accuracy for 70 – 30 split

Classifiers	Accuracy				
	1000 images	2000 images	3000 images	4000 images	5000 images
LR	87.66	84.0	86.77	87.25	78.4
RF	88.33	86.5	88.66	85.75	83.93
K-NN	85.33	84.0	86.66	87.16	82.86
SVM	85.33	84.83	81.66	82.33	79.13
Linear SVM	87.66	82.5	85.44	86.83	75.73
GNB	81.0	84.16	67.88	77.66	71.0
DT	81.66	80.83	82.11	80.33	77.6

Table 3 represents the multiple classifiers accuracy on the different sized dataset for 70 – 30 split. After the analysis, RF gives better performance than other classifiers like LR, K-NN, SVM, Linear SVM, GNB and DT. The maximum accuracy achieved is 88.66 with 3000 images.

Figure 11. represents the accuracy comparison of multiple classifiers for 3000 images. From Fig. 11 we can see that RF gives the best accuracy of 88.66% for 70 – 30 split.

Table 4
Comparison of multiple classifier accuracy for 80 – 20 split

Classifiers	Accuracy				
	1000 images	2000 images	3000 images	4000 images	5000 images
LR	89.0	88.25	82.5	82.0	85.3
RF	88.0	87.75	86.5	84.75	86.6
K-NN	86.5	87.0	82.66	82.5	83.7
SVM	85.0	83.25	85.0	84.87	86.2
Linear SVM	89.0	86.5	80.5	80.5	84.0
GNB	84.0	82.0	77.66	76.12	78.3
DT	83.0	78.0	80.33	78.5	78.1

Table 4 represents the multiple classifiers accuracy for different size. Among them RF provide better accuracy than other classifiers. The maximum accuracy achieved is 89.0 (LR and Linear SVM) with 1000 images.

Figure 12 describes the accuracy comparison of multiple classifiers for 1000 images. In this LR resulted with the best accuracy of 89.0% for 80 – 20 split.

Figure 13 represents the performance comparison of the 80 – 20 split. The blue bar indicates the absence of IDC (IDC-), and the red bar indicates the lack of IDC (IDC+). The precision, recall and F1-score of IDC- are 0.79, 0.96 and 0.86 respectively. The value of precision, recall and F1-score of IDC + are 0.94, 0.74 AND 0.83, respectively.

Figure 14 represents the performance comparison of 70 – 30 split. The blue bar indicates the absence of IDC (IDC-) and red bar indicates the absence of IDC(IDC+). The precision, recall and F1-score of IDC- are 0.90, 0.81 and 0.85 respectively. The value of precision, recall and F1-score of IDC + are 0.83, 0.91 and 0.87 respectively.

5. Conclusion

An IDC prediction model is introduced in WSI using CNN and Multiple classifiers with statistical comparisons. Rebalancing and dimensionality reduction are the general methods to improve classification accuracy. The CNN provide an accuracy of 86.08% with 70 – 30 data split for 1000 images dataset. For the 80 – 20 split, CNN resulted in an accuracy of 84.98% for the 5000 images dataset. Eight epochs used to get the best results from the CNN. The proposed method check the multiple classifier for both 70 – 30 split and 80 – 20 split. In the 70 – 30 split RF gives higher accuracy of 88.66%. For the 80 – 20 split, multiple classifiers resulted with accuracy of 89.0% for LR and Linear SVM classifiers with 1000 image dataset. The early BC prediction increases the survival rate of the patient.

6. Declarations

Funding: No funding is provided for the preparation of manuscript.

Declaration of Interest: Authors Deepa B G, S. Senthil, declares that they have no conflict of interest.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate: Two authors have equal contributions

Consent to Publish: Reviewer and Editors can publish this work

Authors Contributions: All authors are equal contributions in this work

Availability of data and materials: No data Availability

Acknowledgement: I sincerely thanks S. Senthil for their guidance and encouragement in carrying out this research work.

7. References

1. Shaikh K, Krishnan S, Thanki R. Artificial Intelligence in Breast Cancer Early Detection and Diagnosis. *Springer International Publishing AG*; 2020.
2. <https://>.
3. Maurya AP, Brahmachari S. (2020). Current Status of Breast Cancer Management in India. *Indian Journal of Surgery*, 1–6.
4. <https://www.healthline.com/health/breast-cancer#types>.
5. Ray R, Abdullah AA, Mallick DK, Dash SR. (2019). Classification of benign and malignant breast cancer using supervised machine learning algorithms based on image and numeric datasets. *In Journal of Physics: Conference Series, IOP Publishing, 1372(1): 012062*.
6. Gandomkar Z, Brennan PC, Mello-Thoms C. (2016). Computer-based image analysis in breast pathology. *Journal of pathology informatics 7*.
7. Kumar A, Prateek M. Localization of Nuclei in Breast Cancer Using Whole Slide Imaging System Supported by Morphological Features and Shape Formulas. *Cancer Management Research*. 2020;12:4573.
8. Alzubaidi L, Al-Shamma O, Fadhel MA, Farhan L, Zhang J, Duan Y. Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. *Electronics*. 2020;9(3):445.
9. Hamed G, El-Rahman Marey MA, El-Sayed Amin S, Tolba MF. (2020). Deep learning in breast cancer detection and classification. *In Joint European-US Workshop on Applications of Invariance in Computer Vision, Springer, Cham, 322–333*.
10. Tiwari M, Bharuka R, Shah P, Lokare R. (2020). Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques. Available at SSRN 3558786.
11. Bolhasani H, Amjadi E, Tabatabaeian M, Jassbi SJ. A histopathological image dataset for grading breast invasive ductal carcinomas. *Informatics in Medicine Unlocked*. 2020;19:100341.
12. Cruz-Roa A, Basavanhally A, González F, Gilmore H, Feldman M, Ganesan S, Shih N, Tomaszewski J, Madabhushi A. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. *In Medical Imaging 2014: Digital Pathology, International Society for Optics and Photonics, 9041: 904103*.
13. Dabeer S, Khan MM, Islam S. Cancer diagnosis in histopathological image: CNN based approach. *Informatics in Medicine Unlocked*. 2019;16:100231.
14. Roy S, Kumar R, Mittal V, Gupta D. Classification models for Invasive Ductal Carcinoma Progression, based on gene expression data-trained supervised machine learning. *Scientific reports*. 2020;10(1):1–15.
15. Sun Y, Xu Z, Strell C, Moro CF, Wörnberg F, Dong L, Zhang Q. (2018). Detection of Breast Tumour Tissue Regions in Histopathological Images using Convolutional Neural Networks. *In 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS), 98–103*.

16. Romano AM, Hernandez AA. (2019). Enhanced deep learning approach for predicting invasive ductal carcinoma from histopathology images. *In 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), IEEE*, 142–148.
17. Takada K, Kashiwagi S, Asano Y, Goto W, Morisaki T, Takahashi K, Fujita H, et al. Factors predictive of invasive ductal carcinoma in cases preoperatively diagnosed as ductal carcinoma in situ. *BMC Cancer*. 2020;20:1–9.
18. Alzubaidi L, Al-Shamma O, Fadhel MA, Farhan L, Zhang J, Duan Y. (2020). Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model. *Electronics* 9(3): 445.
19. Alghodhaifi H, Alghodhaifi A, Alghodhaifi M. (2019). Predicting Invasive Ductal Carcinoma in breast histology images using Convolutional Neural Network. *In 2019 IEEE National Aerospace and Electronics Conference (NAECON)*, 374–378.

Figures

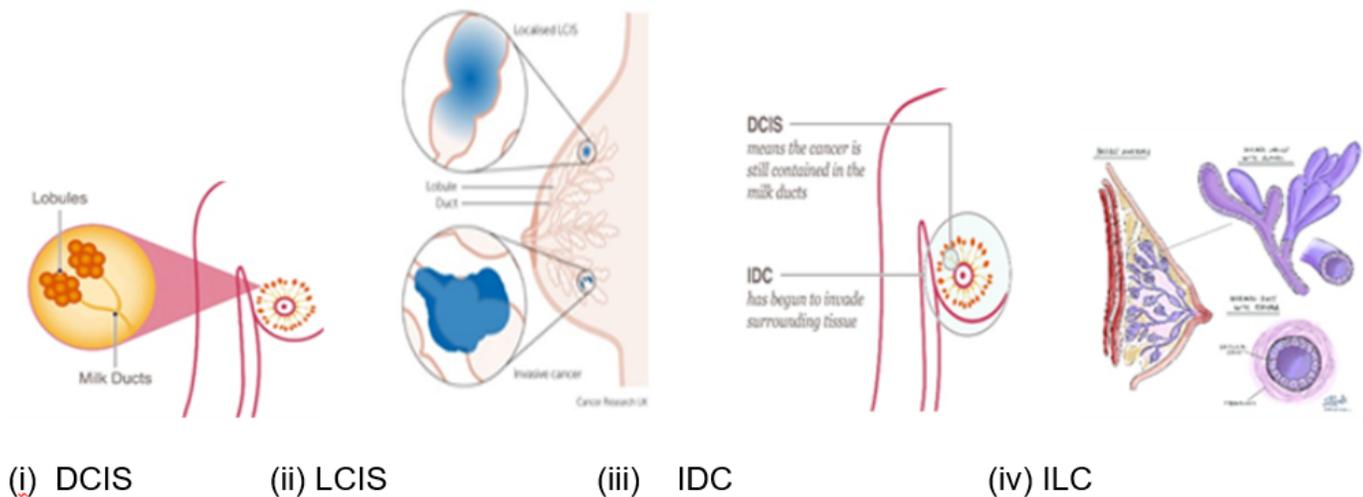


Figure 1

Various Type of Brest Cancer

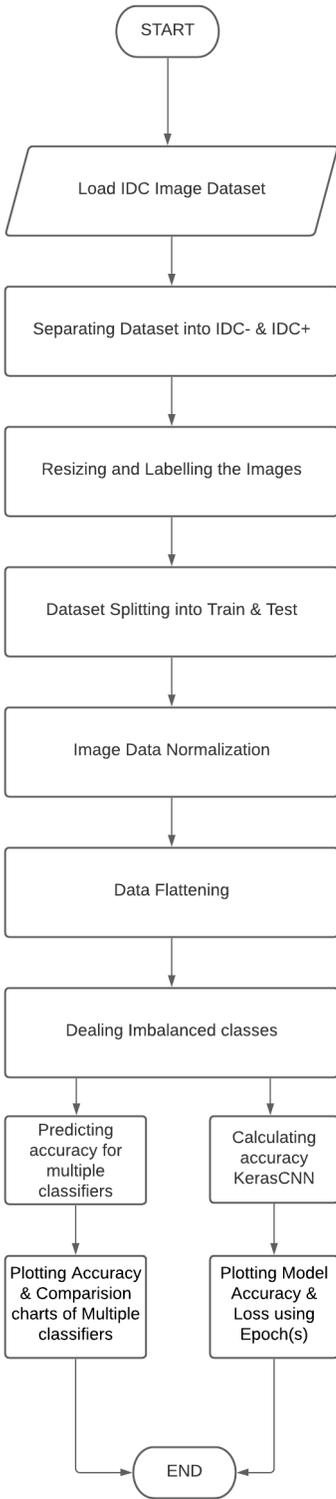


Figure 2

Flowchart of the proposed model

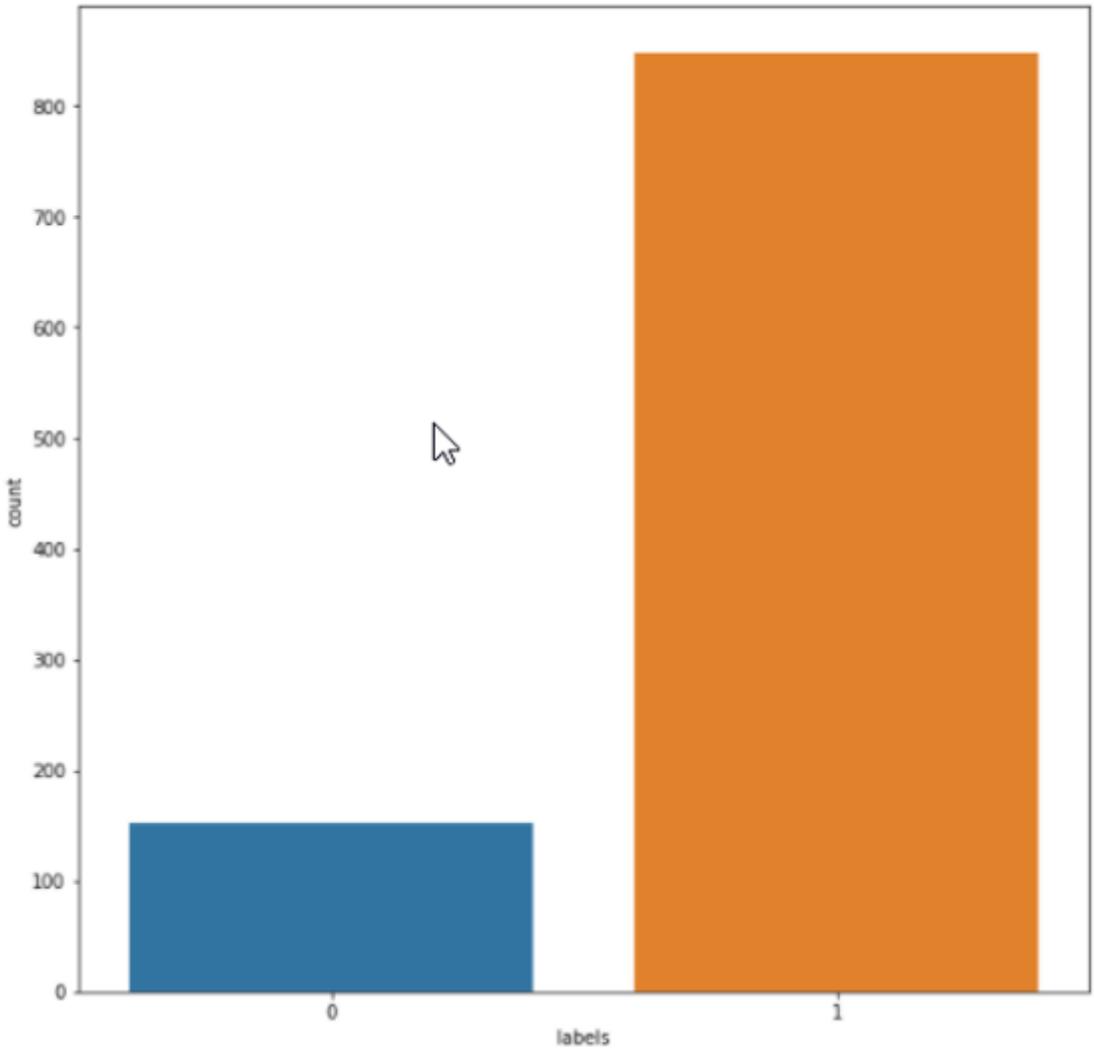


Figure 3

Imbalanced Class

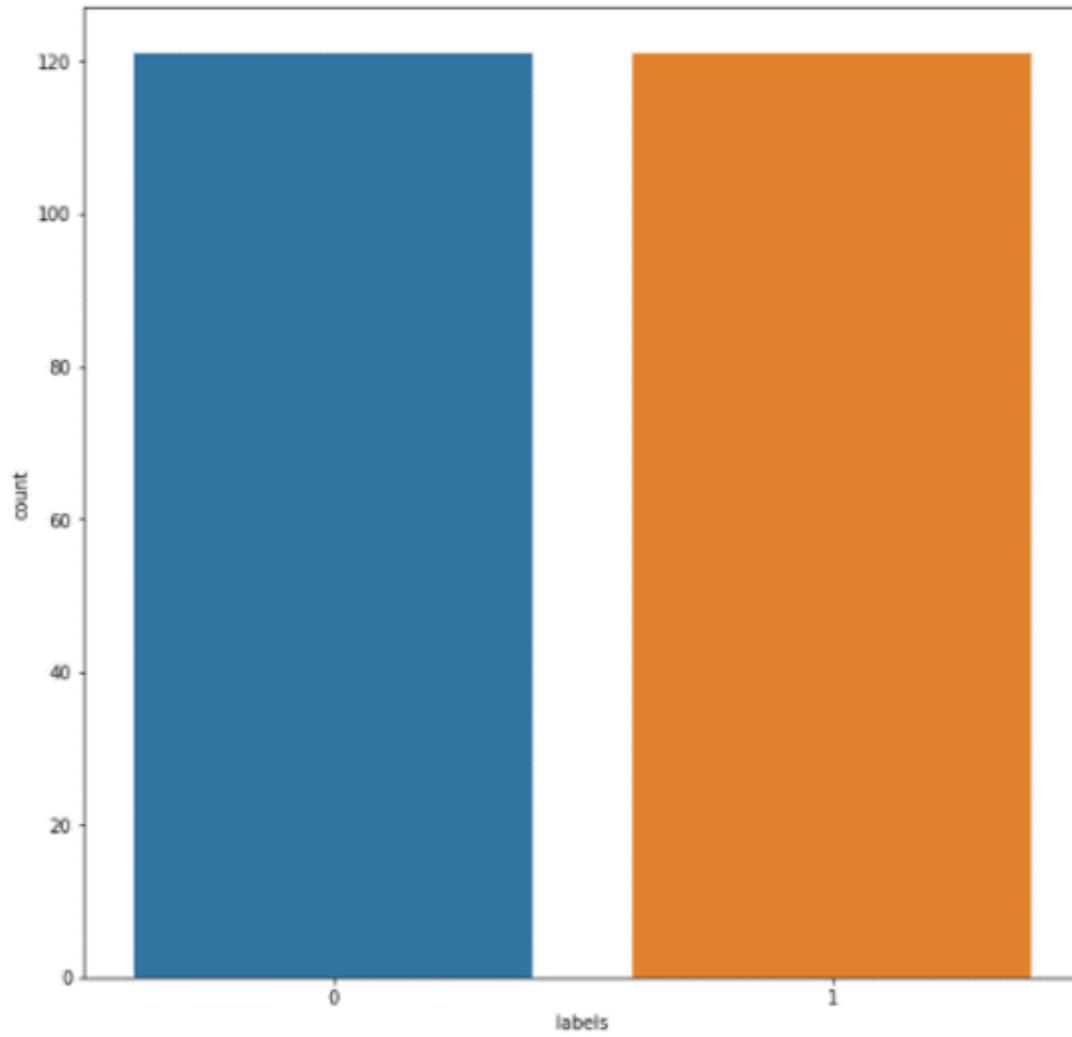


Figure 4

Balanced Class

```

Epoch 1/8
16/15 [=====] - 2s 139ms/step - loss: 0.7687 - acc: 0.5400 - val_loss: 0.7600 - val_acc: 0.5000
Epoch 2/8
16/15 [=====] - 1s 93ms/step - loss: 0.6990 - acc: 0.5084 - val_loss: 0.6718 - val_acc: 0.5000
Epoch 3/8
16/15 [=====] - 2s 102ms/step - loss: 0.6742 - acc: 0.5772 - val_loss: 0.6452 - val_acc: 0.5000
Epoch 4/8
16/15 [=====] - 1s 93ms/step - loss: 0.6946 - acc: 0.6178 - val_loss: 0.6823 - val_acc: 0.5000
Epoch 5/8
16/15 [=====] - 2s 95ms/step - loss: 0.6799 - acc: 0.6299 - val_loss: 0.6047 - val_acc: 0.8478
Epoch 6/8
16/15 [=====] - 1s 93ms/step - loss: 0.6657 - acc: 0.5925 - val_loss: 0.6469 - val_acc: 0.5087
Epoch 7/8
16/15 [=====] - 2s 98ms/step - loss: 0.6840 - acc: 0.6211 - val_loss: 0.5659 - val_acc: 0.8522
Epoch 8/8
16/15 [=====] - 2s 95ms/step - loss: 0.6303 - acc: 0.6811 - val_loss: 0.4808 - val_acc: 0.8609

Keras CNN #1C - accuracy: 0.860869565217

      precision    recall  f1-score   support

 IDC(-)         0.90      0.81      0.85        115
 IDC(+)         0.83      0.91      0.87        115

 avg / total         0.86      0.86      0.86        230

```

Figure 5

Performance evaluation of CNN for 70-30 Split

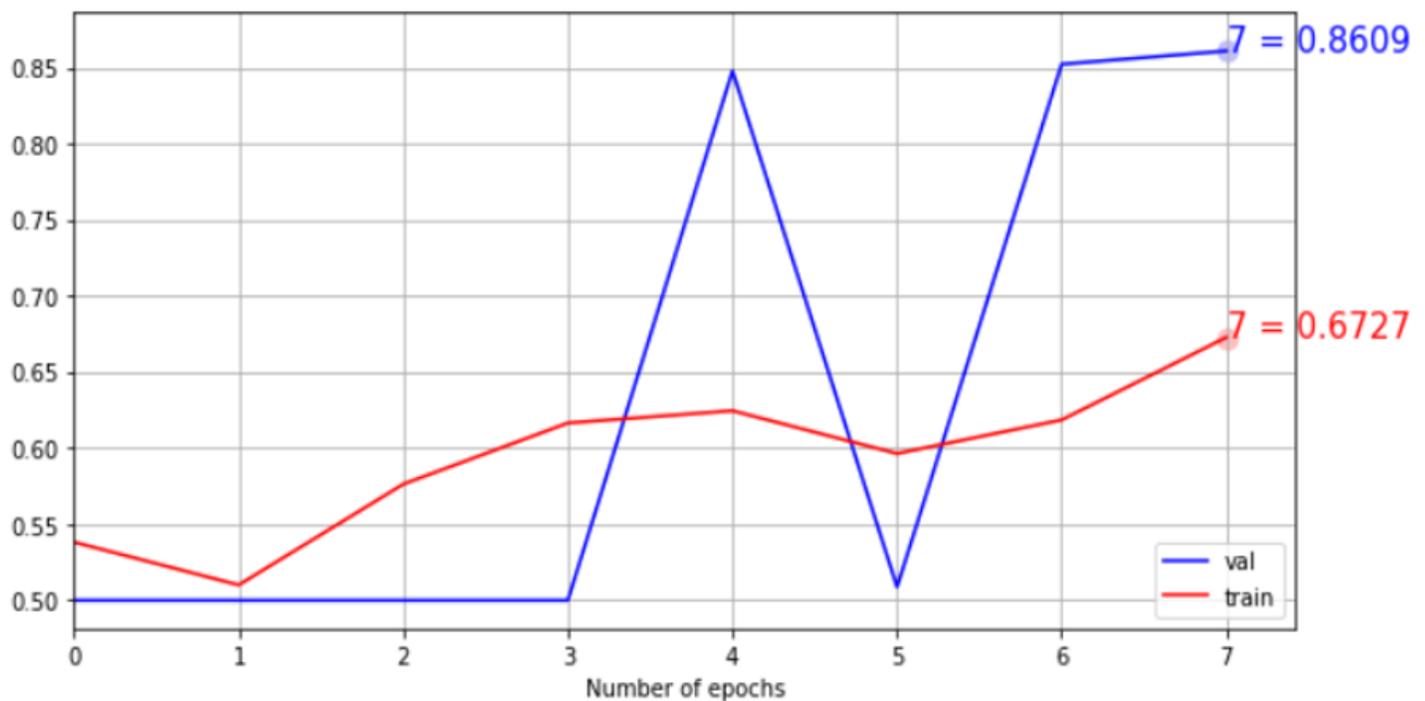


Figure 6

Value accuracy curve of epochs for 70-30 split

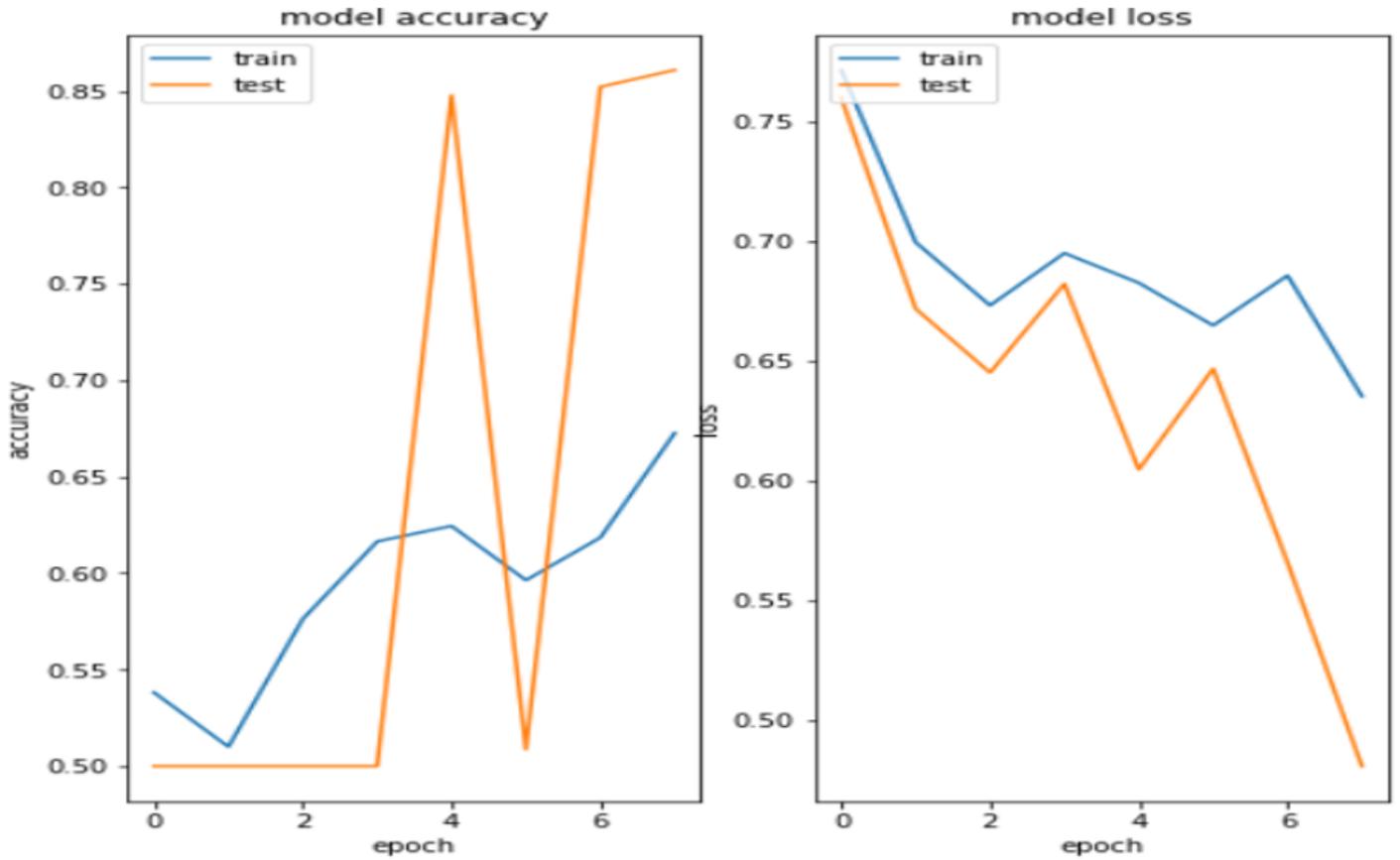


Figure 7

Model accuracy & Model loss curve of test –train with 1000 images

```

Epoch 1/8
80/79 [=====] - 7s 93ms/step - loss: 0.6717 - acc: 0.6125 - val_loss: 0.7758 - val_acc: 0.5248
Epoch 2/8
80/79 [=====] - 7s 84ms/step - loss: 0.5990 - acc: 0.7074 - val_loss: 0.4592 - val_acc: 0.7926
Epoch 3/8
80/79 [=====] - 7s 88ms/step - loss: 0.4687 - acc: 0.7941 - val_loss: 0.3624 - val_acc: 0.8576
Epoch 4/8
80/79 [=====] - 7s 91ms/step - loss: 0.4303 - acc: 0.8223 - val_loss: 0.5085 - val_acc: 0.7554
Epoch 5/8
80/79 [=====] - 7s 85ms/step - loss: 0.3939 - acc: 0.8422 - val_loss: 0.3139 - val_acc: 0.8793
Epoch 6/8
80/79 [=====] - 7s 88ms/step - loss: 0.3866 - acc: 0.8461 - val_loss: 0.3378 - val_acc: 0.8622
Epoch 7/8
80/79 [=====] - 7s 86ms/step - loss: 0.3569 - acc: 0.8520 - val_loss: 0.3147 - val_acc: 0.8700
Epoch 8/8
80/79 [=====] - 7s 86ms/step - loss: 0.3479 - acc: 0.8570 - val_loss: 0.3535 - val_acc: 0.8498

Keras CNN #1C - accuracy: 0.849845201054

      precision    recall  f1-score   support

 IDC(-)         0.79      0.96      0.86         323
 IDC(+)         0.94      0.74      0.83         323

 avg / total         0.87      0.85      0.85         646

```

Figure 8

Performance evaluation of CNN algorithm for 80-20 Split

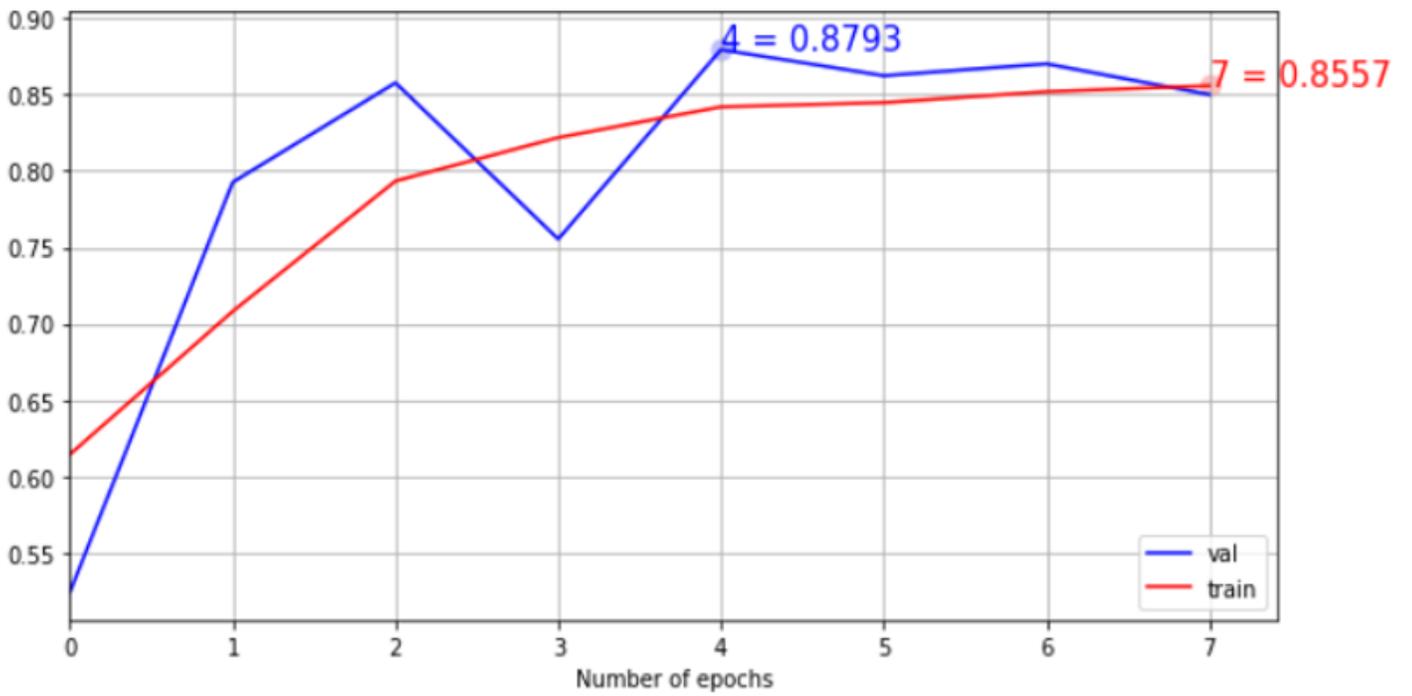


Figure 9

Value accuracy curve of epochs for 80-20 split

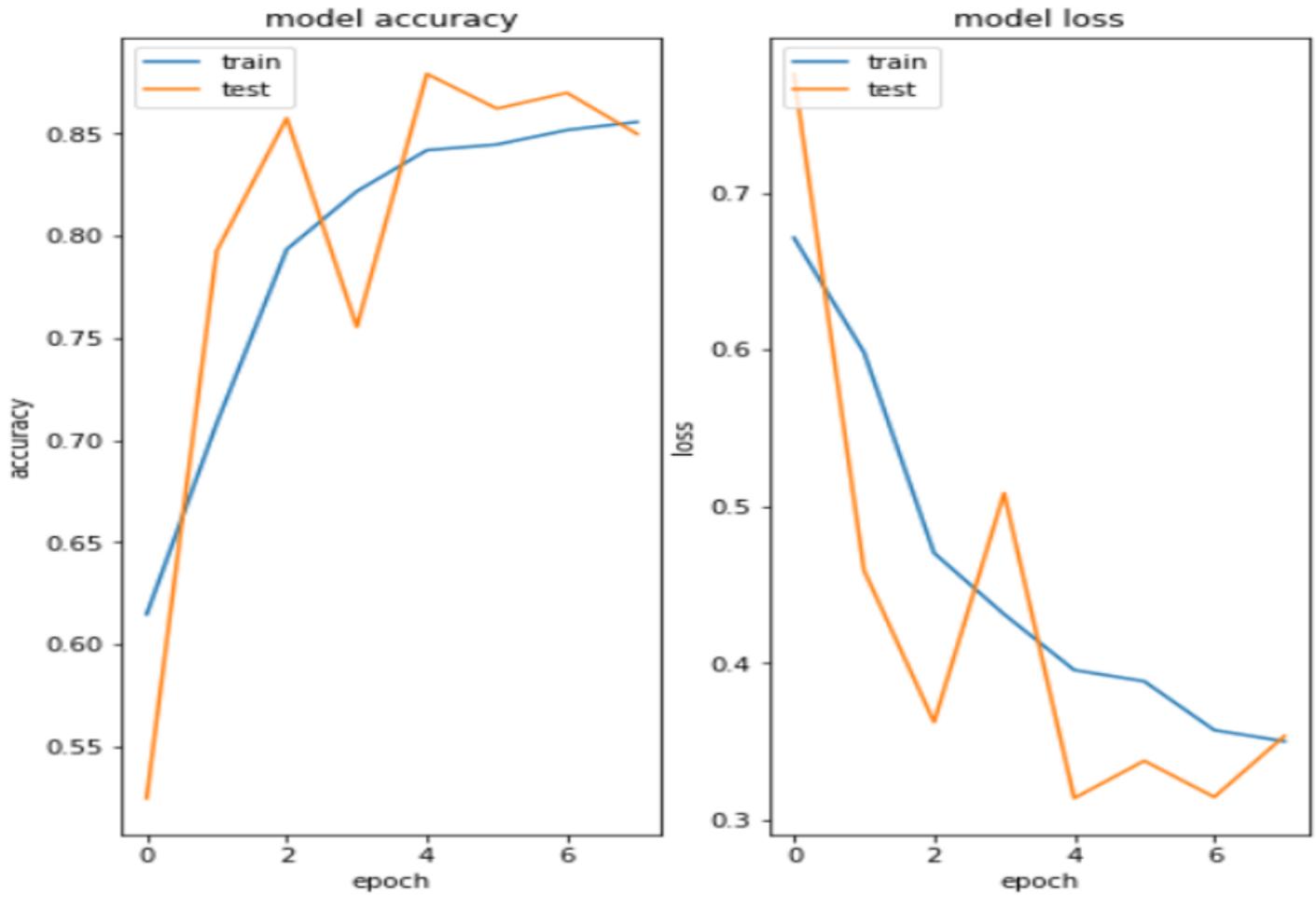


Figure 10

Model accuracy & Model loss curve of the test –train with 5000 images

Algorithm Comparison: Accuracy

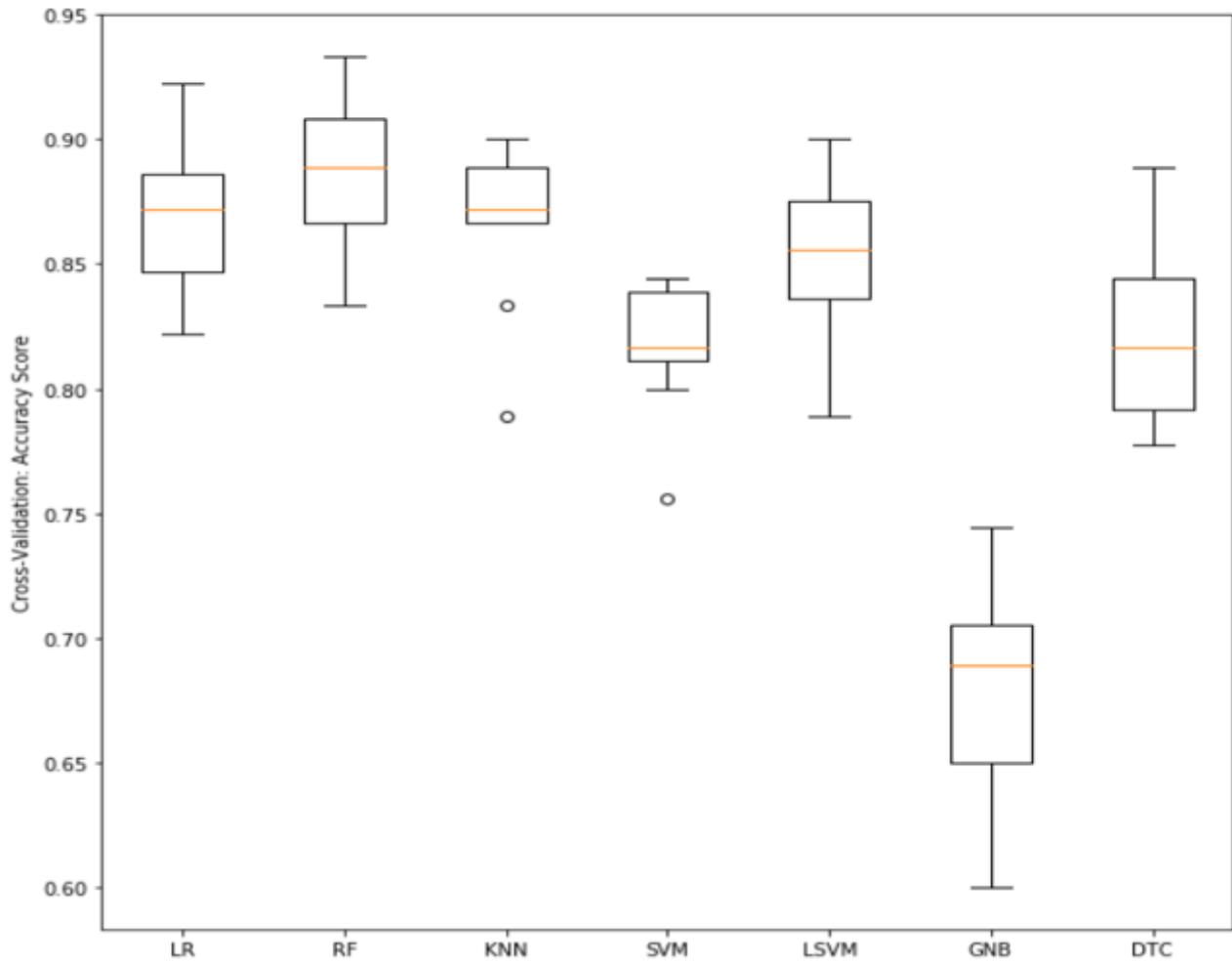


Figure 11

Accuracy comparison chart of multiple classifiers

Algorithm Comparison: Accuracy

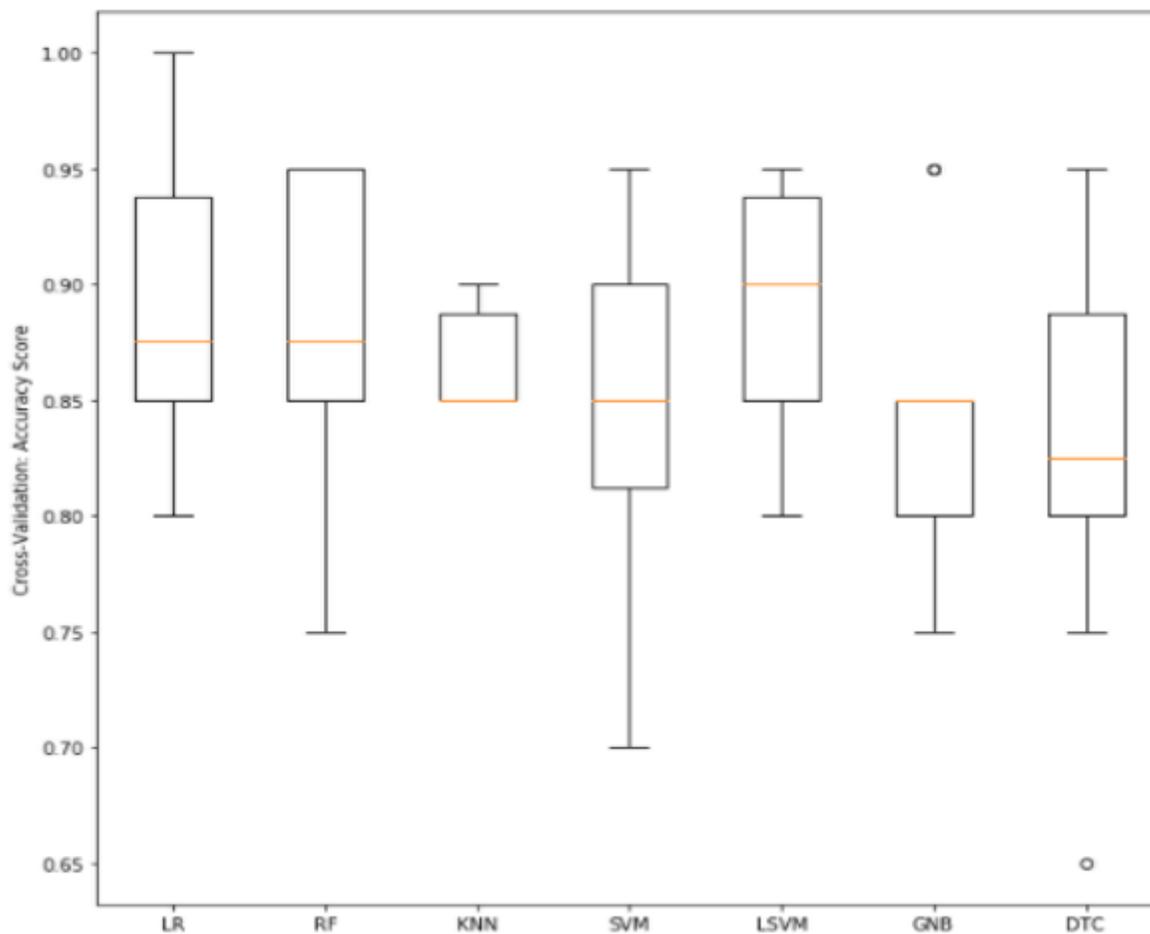


Figure 12

Accuracy comparison chart of multiple classifiers

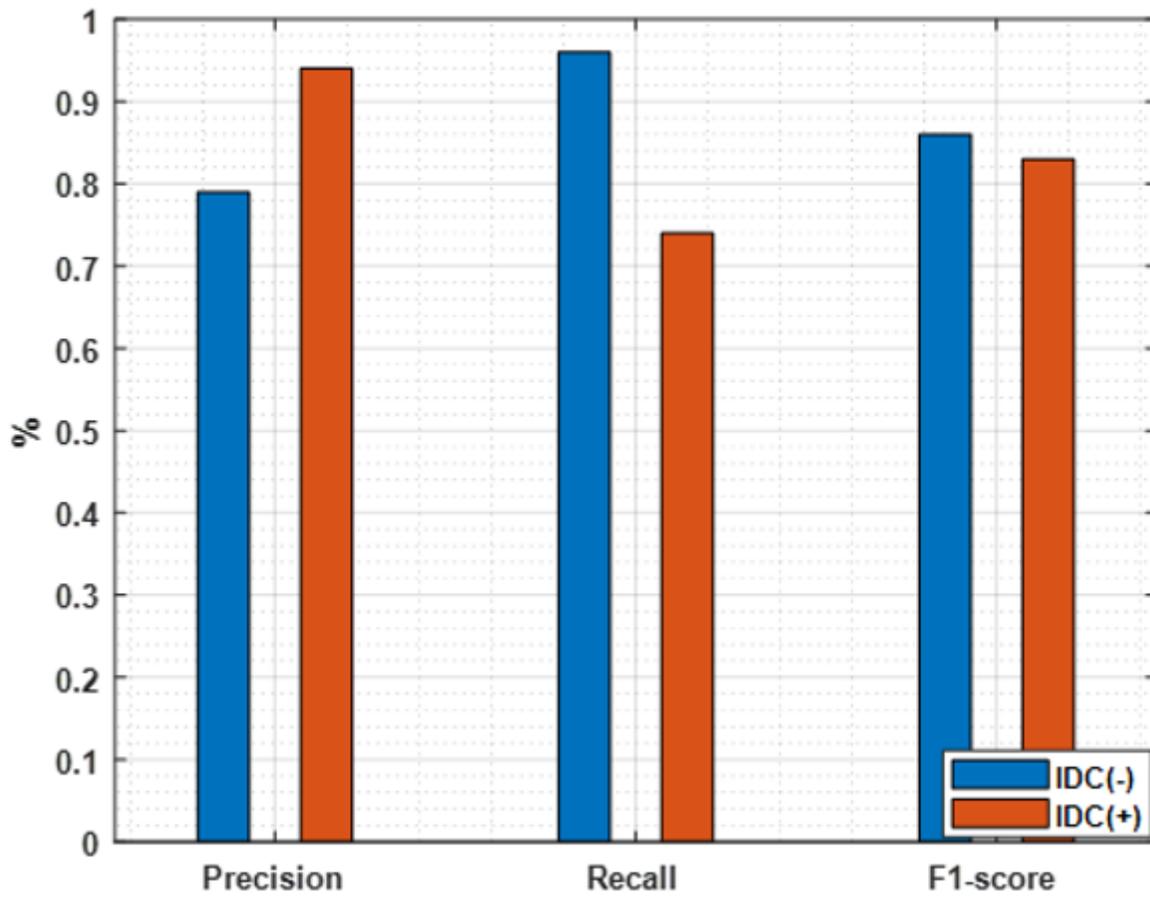


Figure 13

Performance comparison of 80-20 split

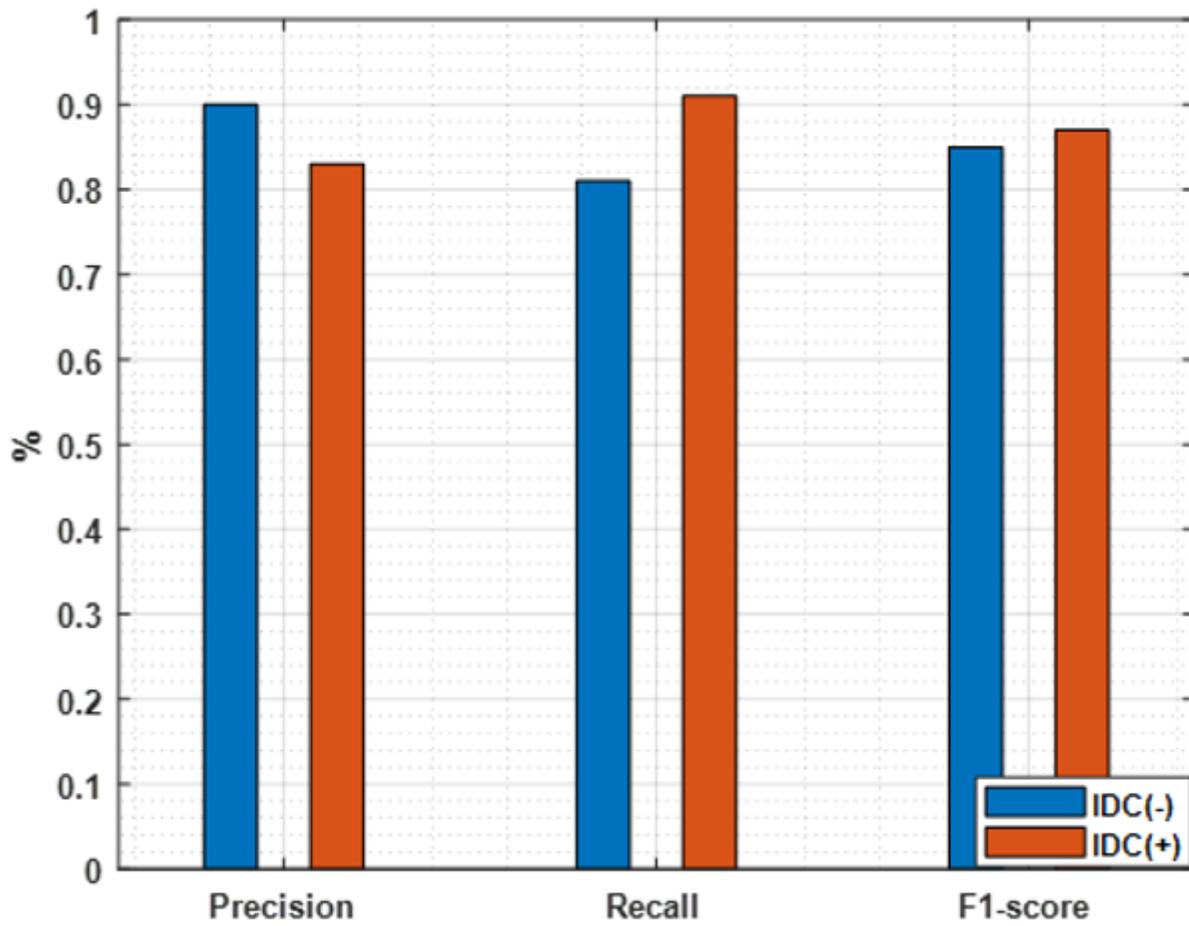


Figure 14

Performance comparison of 70-30 split