RCC INSTITUTE OF INFORMATION TECHNOLOGY

Canal South Road, Beliaghata, Kolkata – 700015

BREAST CANCER DIAGNOSIS USING DEEP LEARNING BASED ON ARTIFICIAL NEURAL NETWORK

Soumyamoy Das (11700118036)

Under The Guidance Of

Dr Parama Bagchi

INTRODUCTION

- Breast Cancer is one of the most commonly occurring cancers among the middle-aged women globally
- According to the statistics, 1 in every 8 women are diagnosed with Breast Cancer
- Despite the moderate curable rate, breast cancer affects the physique and development of a woman and can also prove fatal to many
- According to another statistics, more than 40,000 women succumb to breast cancer every year in the US
- Breast cancer, like any other cancer, is highly curable when diagnosed at an early stage

PROBLEM STATEMENT

Cancerous tissues are detected after observing various parameters of the Fine Needle Aspirate Cytology (FNAC) image of the tissue mass, manually under a microscope. Whenever, the observed parameters varies from the desired ones, it signifies the presence of carcinogenic cells. Now, this manual examination of every tissue and tallying the parameters against the standard ones, is both cumbersome and error-prone. Thus, we have tried to develop a model which would speed up the process of cancer cells' detection.

SCOPE

Present Scope

- Presently, we are considering the pre-extracted data (parameters) from the FNAC images to predict whether the cell is cancerous or not.

Future Scope

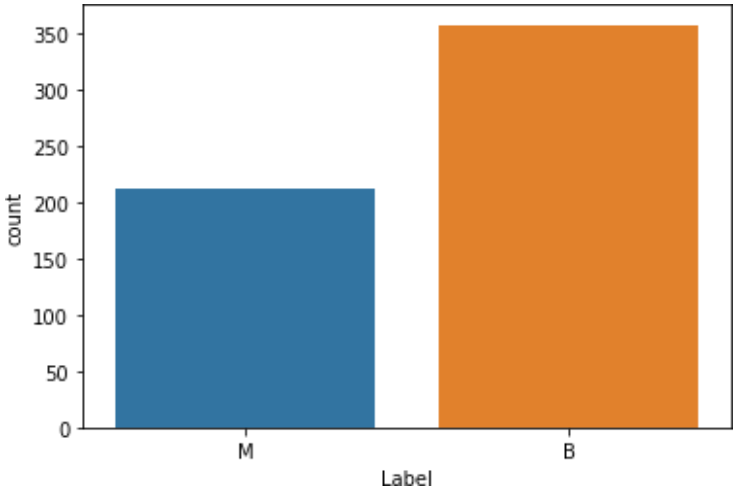
- The model shall work directly on the Breast Mass FNAC image, which would further speed up the entire procedure. Also, the results would be more accurate.

DATASET

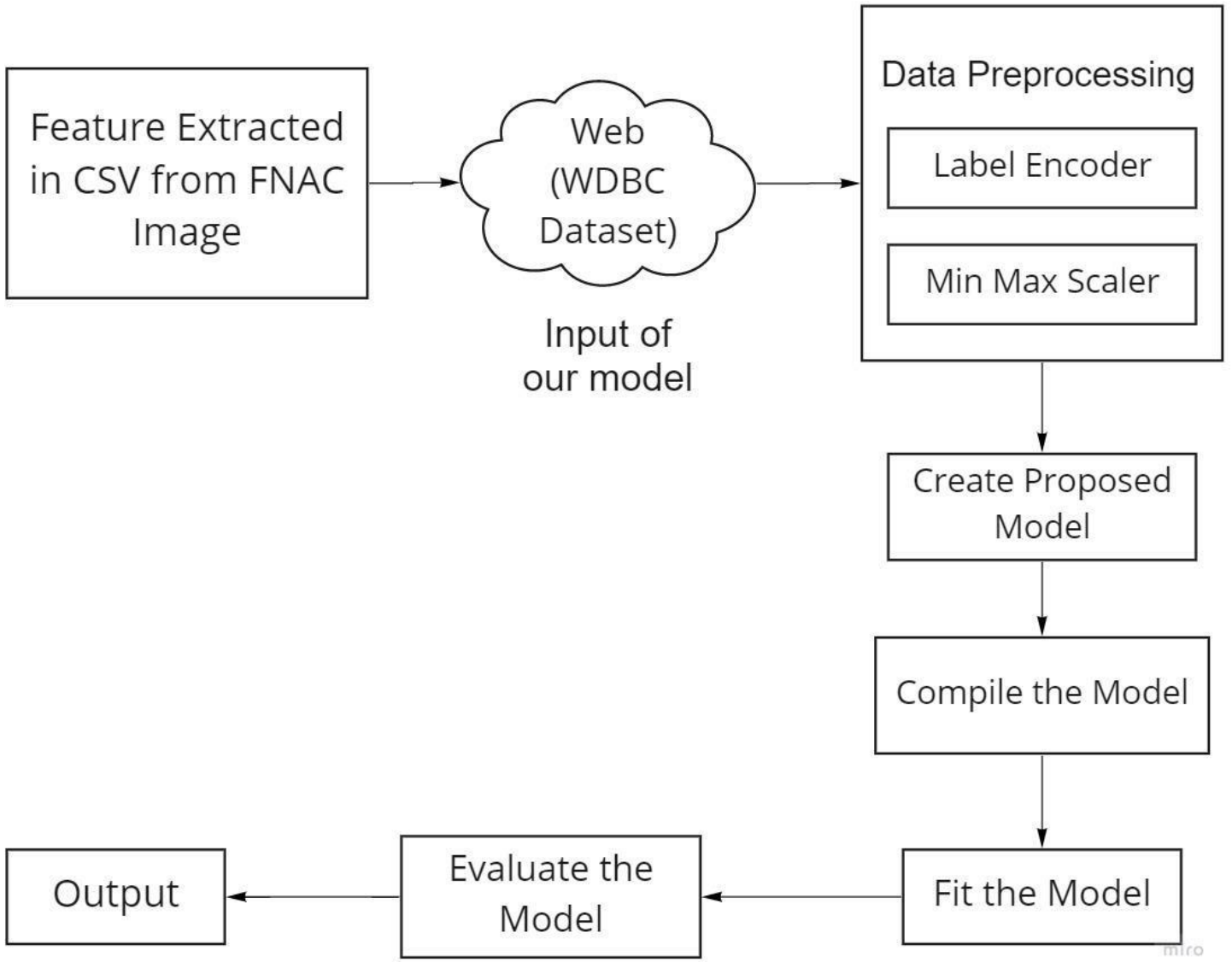
UCI BREAST CANCER WISCONSIN (DIAGNOSTIC) DATA SET (WDBC)

DATASET – DESCRIPTION

- The Dataset is in the form a CSV file and comprises of **32 Attributes** for **569 Unique Entries**
- The attributes are pre-extracted from the Breast Tissue FNAC images
- The Dataset comprises of 357 Benign and 212 Malignant entries
- 10 real valued parameters extracted from the FNAC image are: *radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry* and *fractal dimension*
- These 10 parameters are present in 3 groups: *mean, standard error* and *worst*
- We are using these 30 attributes to train and test our model
- The other 2 attributes are sample id and result, which are not used directly in building the model



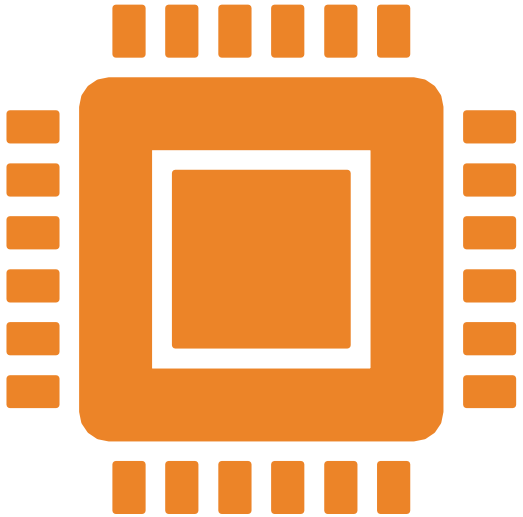
WORKFLOW


Feature Extracted in CSV from FNAC Image
Web (WDBC Dataset)
Input of our model
Data Preprocessing
Label Encoder
Min Max Scaler
Create Proposed Model
Compile the Model
Fit the Model
Evaluate the Model
Output

FEATURES

Speed up cancer detection from the pre-extracted parameters of the breast mass FNAC image

Eliminate any manual intervention and thus, to improve detection accuracy



TECH STACK & DEPENDENCIES

- **Programming Language: Python**
  - Python Dependencies:
    - Pandas
    - Matplotlib
    - Seaborn
    - Numpy
    - Sklearn
    - Keras

The code has been written and compiled in Google Colaboratory

METHODOLOGY

Data Preprocessing

- Eliminating null valued records
- Label Encoding: Encoding of categorical data from text (B and M) to integers (0 and 1)
- Min-Max scaling/Normalization: Scales and translates each feature individually such that it is in the given range on the training set, here, between 0 and 1
- Splitting the data for training and testing: We have used 75% of the data for training and the rest for testing

METHODOLOGY

Building the Model

We are building the model using Keras. Sequential is the easiest way to build a model in Keras since it allows us to build a model layer by layer.

Our neural network model comprises of 5 layers: 1 input layer, 3 hidden layers and 1 output layer.

Activation function: Activation function used in all the hidden layers is relu (Rectified Linear Unit). Activation function used in the output layer is sigmoid.

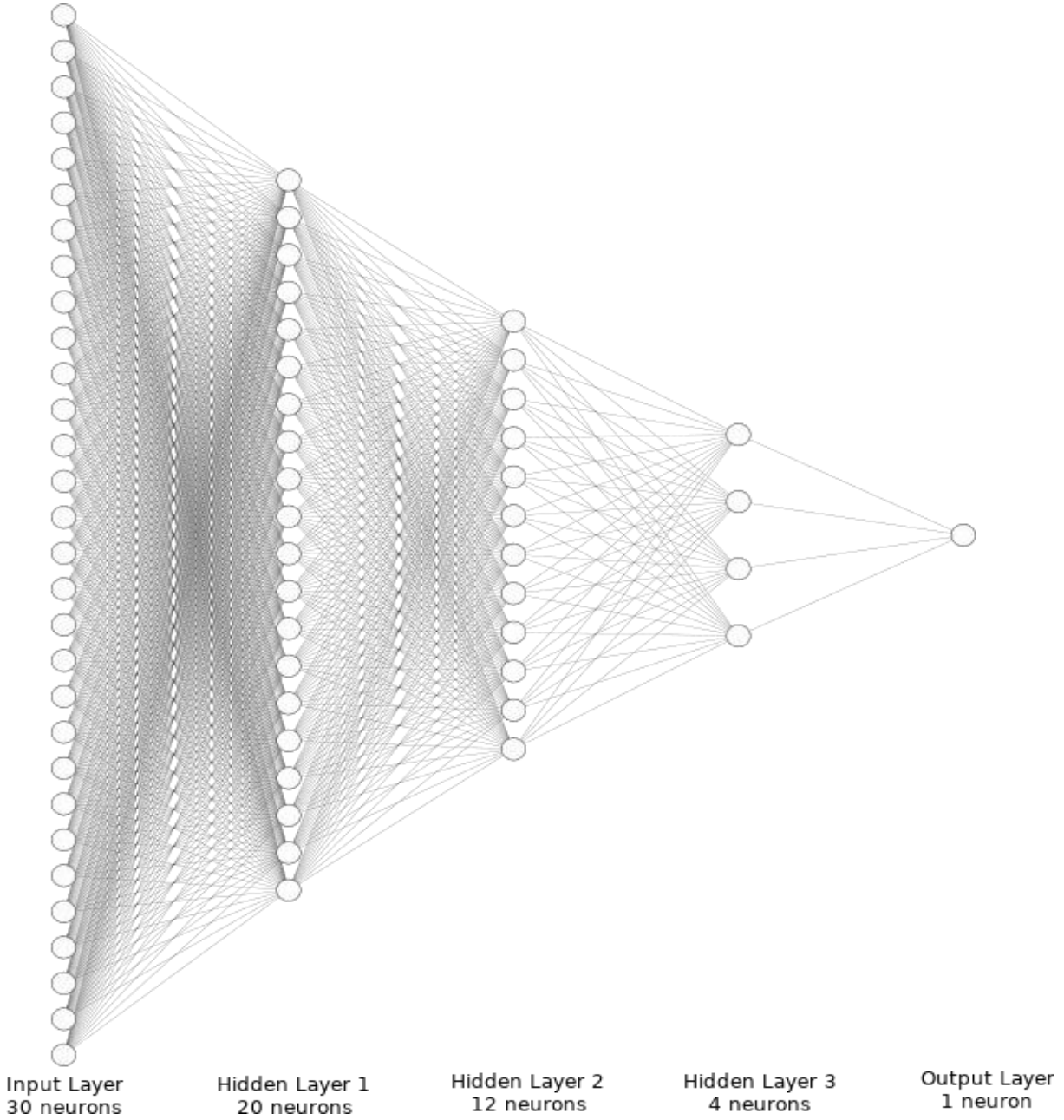
Dropout: Dropout is a simple way to prevent neural networks from overfitting. Hence, we are dropping out 10% of the connections at every layer.

METHODOLOGY

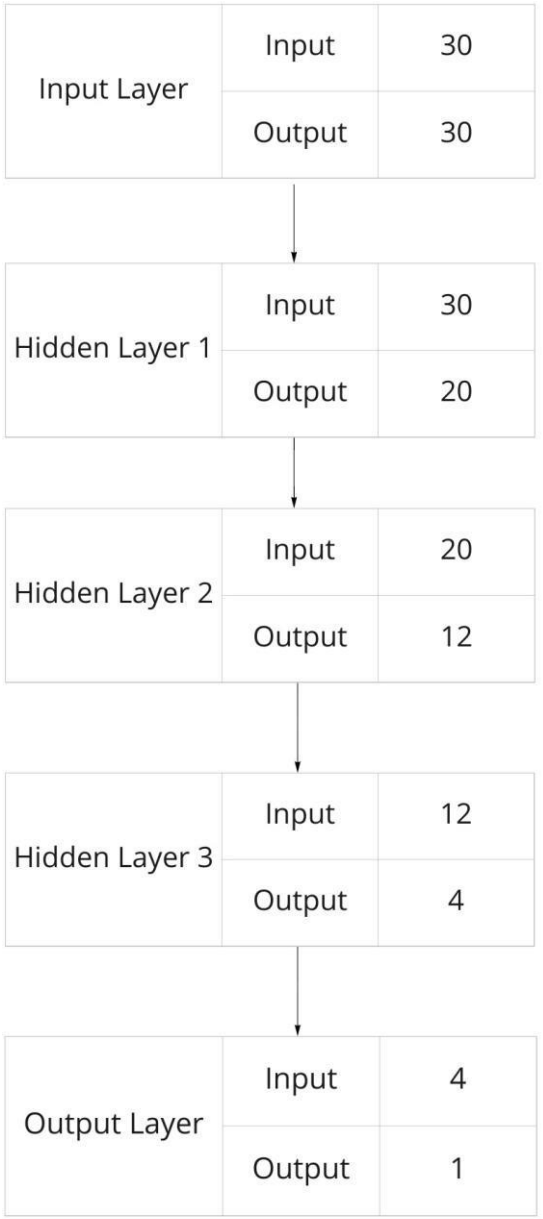
Compiling the Model

Binary Cross Entropy: The model is compiled using binary cross entropy loss function. Loss/cost functions are used to optimize the model during training. Binary cross entropy is used as a loss function for binary classification model.

DEEP LEARNING MODEL BASED ON ANN



MODEL OVERVIEW



PERFORMANCE ANALYSIS

Training Loss

Training loss is a metric used to assess how a deep learning model fits the training data, i.e., it assesses the error of the model on the training set.

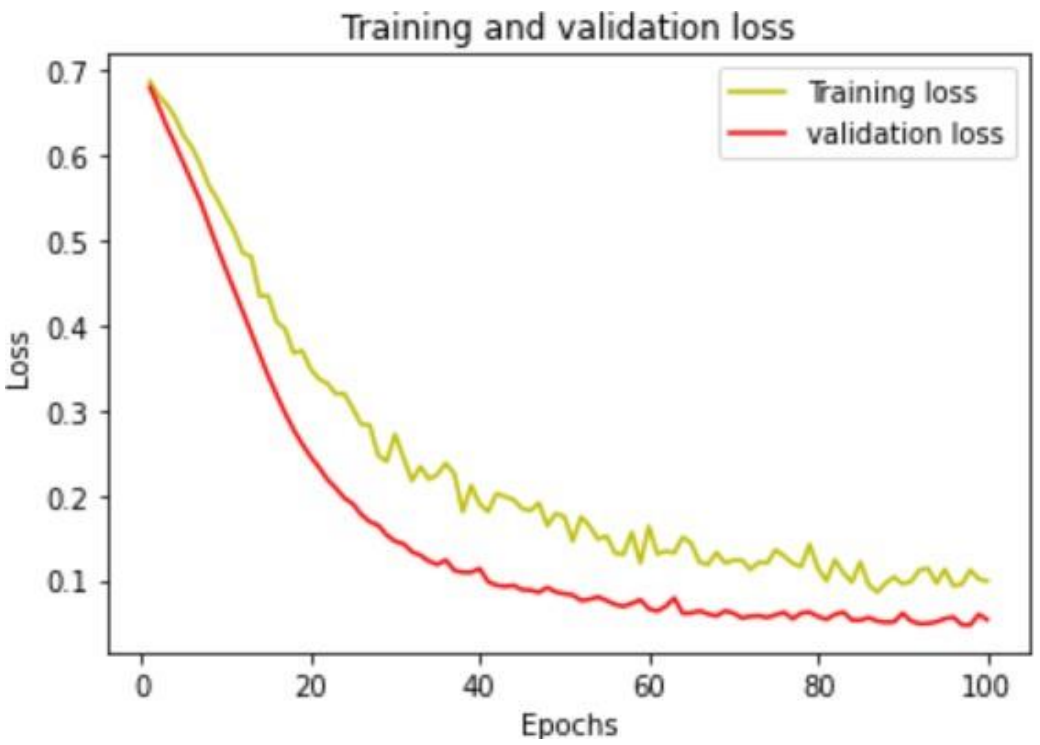
The training set is a portion of a dataset used to initially train the model.

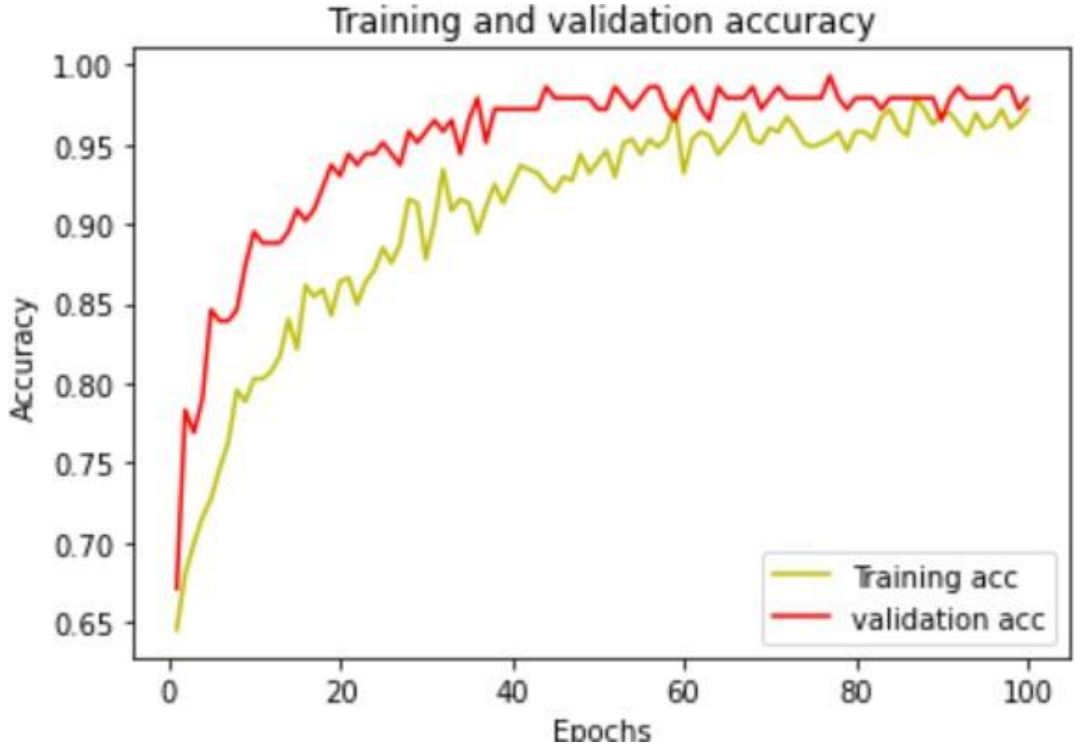
Validation Loss

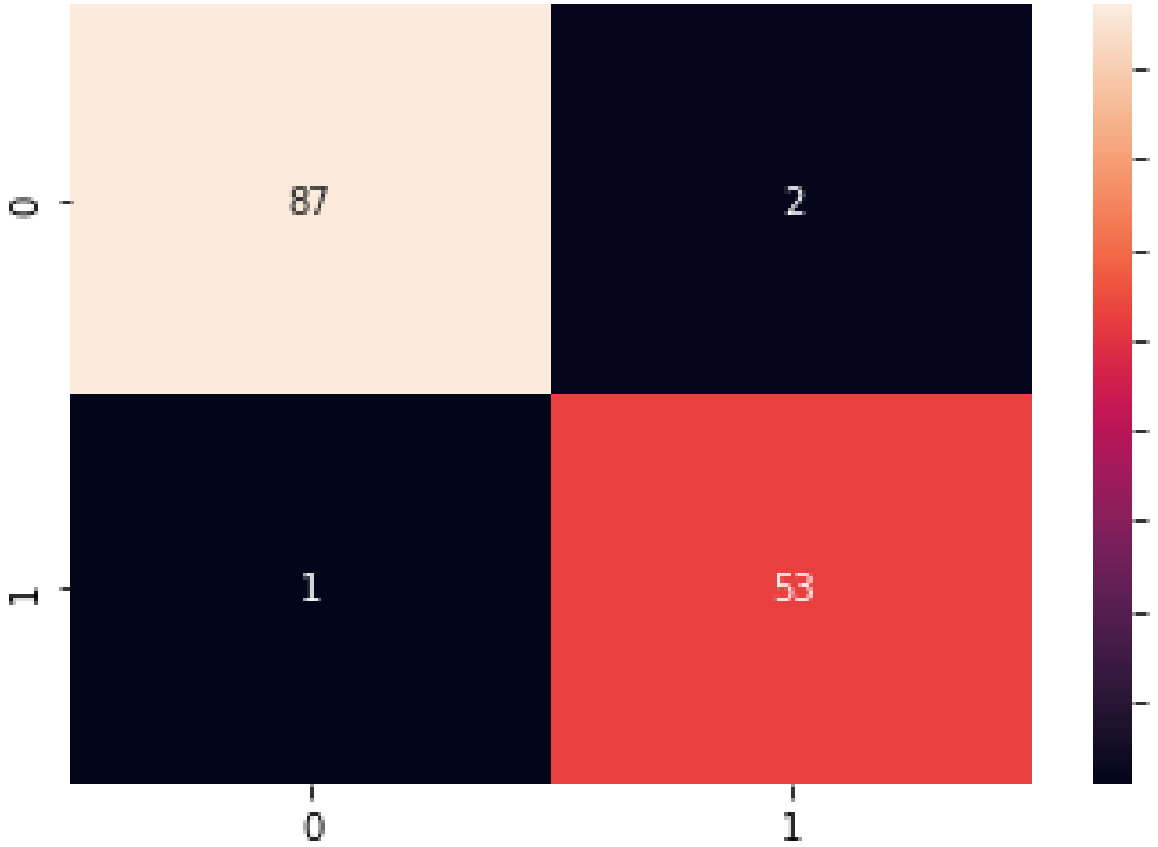
Validation loss is a metric used to assess the performance of a deep learning model on the validation set.

The validation set is a portion of the dataset set aside to validate the performance of the model.

PERFORMANCE ANALYSIS – CURVES







Confusion Matrix

OUTPUT

- The Model is trained and tested taking the number of epochs as 100 and batch size as 64
- Therefore, overall accuracy of our model from the Confusion Matrix is 0.97902 or **more than 97%**

COMPARATIVE ANALYSIS

Algorithm	Accuracy	Paper Reference
Feedforward Neural Network (FNN)	92%	[1]
Convolutional Neural Network (CNN)	95.6%	[2]
Deep Neural Network (DNN)	87%	[3]
Convolutional Neural Network (CNN)	96.7%	[4]
Convolutional Neural Network (CNN)	84.6%	[5]
Convolutional Neural Network (CNN)	90%	[6]
Artificial Neural Network (ANN)	96.5%	[7]
Convolutional Neural Network (CNN)	95.4%	[8]
Convolutional Neural Network (CNN)	94.5%	[9]
Convolutional Neural Network (CNN)	95.6%	[10]
Support Vector Machine (SVM)	95.25%	[11]
k-Nearest Neighbour (*k*-NN)	91.03%	[11]
Our Proposed Model (based on ANN)	**97.90%**	

CONCLUSION

This technique is used practically in several cancer diagnostic systems to detect the cancerous cells automatically by studying only the parameters from a FNAC image.

Our model is quite accurate since we have tested its accuracy to be more than 97%. So, our model can also be used for practical purposes. However, presently we are implementing the model on the pre-extracted parameters from the FNAC image to detect cancer. We have future endeavors to implement the model directly on the FNAC images.

BIBLIOGRAPHY

- WDBC Dataset
- Code Repository
- Towards Data Science
- Machine Learning Mastery
- Free Code Camp
- Scikit Docs
- Keras Docs
- Introduction to machine learning with Scikit-learn and Tensorflow

REFERENCES

- [1] Umamaheswari TS, Sumathi P. Enhanced firefly algorithm (EFA) based gene selection and adaptive neuro neutrosophic inference system (ANNIS) prediction model for detection of circulating tumor cells (CTCs) in breast cancer analysis. Cluster Comput 2019
- [2] Mostavi M, Chiu Y-C, Huang Y, Chen Y. Convolutional neural network models for cancer type prediction based on gene expression. BMC Med Genomics 2020
- [3] Chen H, Gao M, Zhang Y, Liang W, Zou X. Attention-Based Multi-NMF Deep Neural Network with Multimodality Data for Breast Cancer Prognosis Model. Biomed Res Int 2019
- [4] Lancashire LJ, Powe DG, Reis-Filho JS, Rakha E, Lemetre C, Weigelt B, et al. A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks. Breast Cancer Res Treat 2010
- [5] Rawat RR, Ruderman D, Macklin P, Rimm DL, Agus DB. Correlating nuclear morphometric patterns with estrogen receptor status in breast cancer pathologic specimens. Npj Breast Cancer 2018
- [6] Turkki R, Linder N, Kovanen PE, Pellinen T, Lundin J. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin-stained breast cancer samples. J Pathol Inform 2016
- [7] Mert A, Kiliç N, Bilgili E, Akan A. Breast cancer detection with reduced feature set. Comput Math Methods Med 2015
- [8] Murtaza G, Shuib L, Mujtaba G, Raza G. Breast Cancer Multi-classification through Deep Neural Network and Hierarchical Classification Approach. Multimed Tools Appl 2020
- [9] Saha M, Arun I, Ahmed R, Chatterjee S, Chakraborty C. HscoreNet: A Deep network for estrogen and progesterone scoring using breast IHC images. Pattern Recognit 2020
- [10] Mostavi M, Chiu YC, Huang Y, Chen Y. Convolutional neural network models for cancer type prediction based on gene expression. BMC Med Genomics 2020
- [11] A. Mert, N. Kılıç, E. Bilgili, and A. Akan, "Breast cancer detection with reduced feature set," Computational and mathematical methods in medicine, vol. 2015, 2015.

THANK YOU

17TH JUNE 2022