# Deep Learning Techniques for Breast Cancer Risk Prediction using IBM Cloud

# INDEX

# 1. INTRODUCTION

## 1.1 Overview

In this application in python, we'll build a classifier to train on 80% of a breast cancer histology image dataset. Of this, we'll keep 10% of the data for validation. Using Keras, we'll define a CNN (Convolutional Neural Network), and train it on our images on IBM Watson.

## 1.2 Purpose

Breast cancer is one of the main causes of cancer death worldwide. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible.

The goal is to classify images into two classifications of malignant and benign. As early diagnostics significantly increases the chances of correct treatment and survival. In this application, we are helping the doctors and patients to classify the Type of Tumour for the specific image given with the help of Neural Networks.

# 2. LITERATURE SURVEY

## 2.1. Existing Problem

Many works have been submitted which attempted to diagnose breast cancer using machine learning algorithms. For instance, Sun at al. in year 2005, proposed comparing feature selection methods for a unified detection of breast cancers in mammograms. Another approach, introduced by Malek at al. in year 2009, proposed a method using wavelet and proposed a design of automated detection, segmentation, and classification of breast cancer nuclei using a fuzzy logic for feature exaction and classification respectively. Zheg at al. in year 2014 combined support vector machine (SVM) and K-means algorithm for breast cancer diagnosis. Aličković and Subasi in year 2017 applied a genetic algorithm for feature extraction and rotation for classification. Another approach is conducted by Bannaie in year 2018 based on the dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) technique to attain output of interest. There are several other works performed based on clustering and classification . Alireza Osarech, Bita Shadgar achieved 98.80% and 96.63% accuracies upon using SVM classification technique on two different benchmark datasets for breast cancer. Mandeep Rana, Pooja Chandorkar, Alishiba Dsouza applied KNN, SVM, Gaussian Naïve Bayes, and Logistic Regression techiniques programmed in MATLAB to diagnose and predict recurrence of breast cancer. The classification techniques were applied on two dataset from UCI depository. One dataset was used for identification of diseases (WBCD), and other is used for prediction of recurrence. Vikas Chaurasia, BB Tiwari and Saurabh Pal build predictive models on breast cancer and compared their accuracies using famous algorithms viz. J48, Naïve Bayes, and RBF. The results indicated that Naïve Bayes predicted well among them with 97.36% accuracy. Haifeng Wang and Sang Won Yoon developed a powerful model for breast cancer prediction by using and comparing Naive Bayes Classifier, Support Vector Machine (SVM), AdaBoost tree and Artificial Neural Networks (ANN). They implemented PCA for dimensionality reduction. S.Kharya proposed Artificial Neural Networks (ANN) while working on breast cancer prediction. The paper highlighted advantages of using machine learning methods like SVM, Naive Bayes, Neural network and Decision trees . Naresh Khuriwal and Nidhi Mishra used Wisconsin Breast Cancer database to work on breast cancer diagnosis.

Based on their experiments they concluded that, ANN and Logistic Algorithm worked better and achieved an accuracy of 98.50 percent .

## 2.2. Proposed Solution

Our proposed method has the following steps:

- Data collection: the dataset was taken from link.
- Image preprocessing: The ImageDataGenerator accepts the original data, randomly transforms it, and returns only the new, transformed data
- Model Building:The neural network model is to be built by adding different network layers like convolution, pooling, flattening, dropout, and neural layers. We start building our model by:
    1. Initializing the mode
    2. Adding Convolution layers
    3. Adding Pooling layers
    4. Flatten layer
    5. Full connection layers which include hidden layers

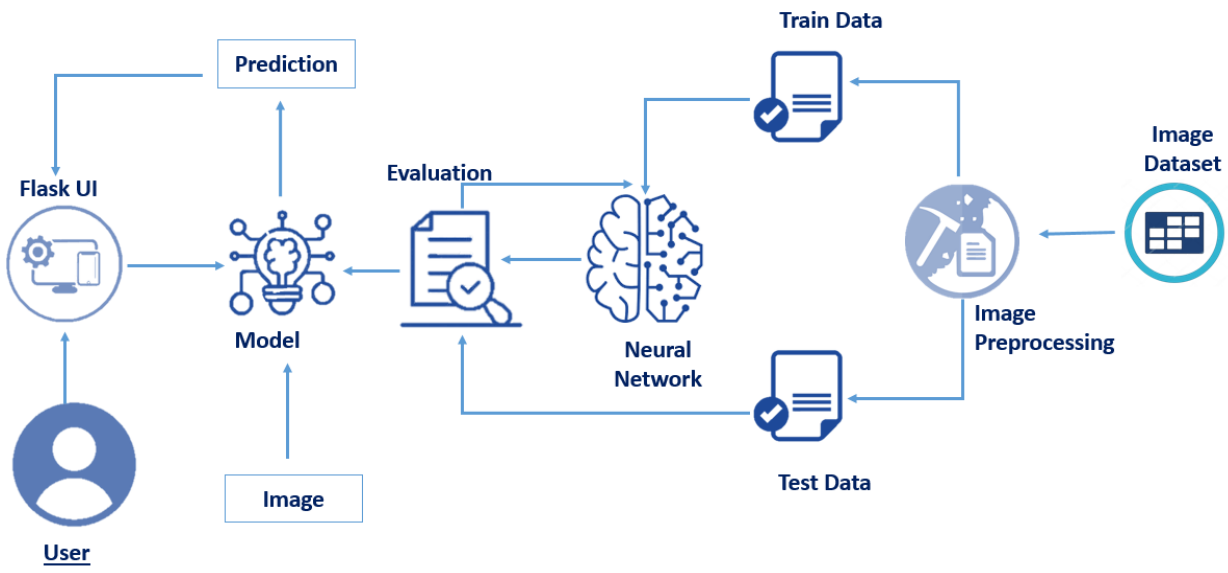
At last, we compile the model with layers we added to complete the neural network structure

- Model Testing: The model is to be tested with different images to know if it is predicting correctly.

- Application Building: After the model is built, we will be integrating it into a web application so that normal users can also use it. The users need to give the X-ray images to know the predictions.

# 3. THEORITICAL ANALYSIS

## 3.1. Block Diagram



## 3.2. Hardware / Software designing

We need the following hardware requirements:

1.60GB RAM
2. Nvidia K80 (24GB) graphics card
3. 20GB hard disk space
4. Testing can be done on PC: 8GB RAM
5. intel i7 processor

We used the following software requirements:

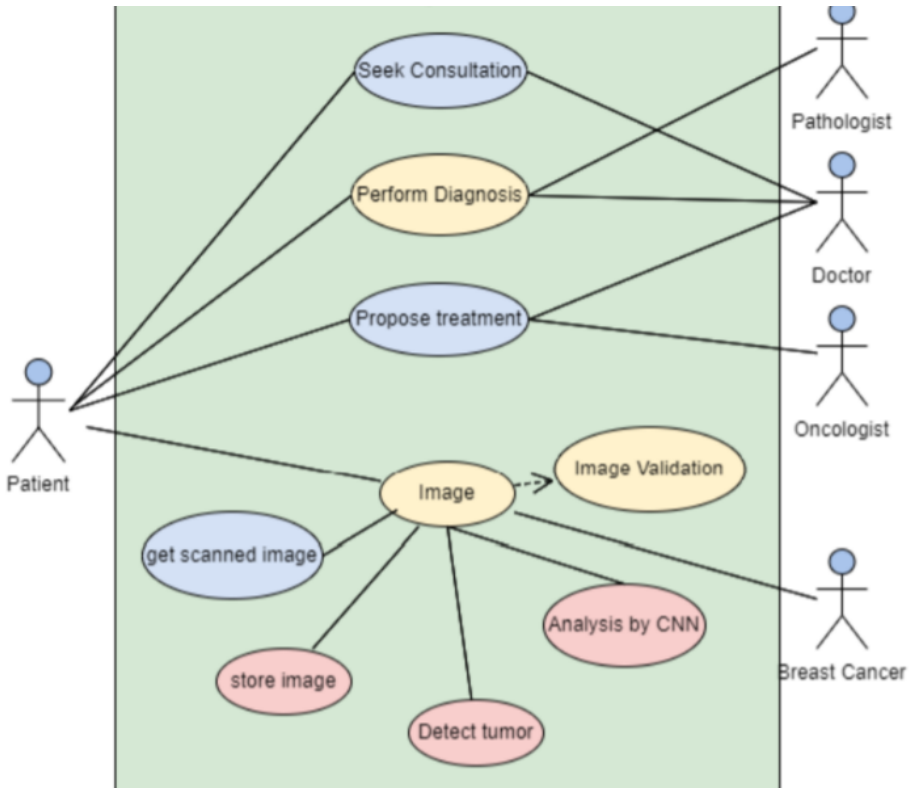
1. Anaconda Navigator
2. TensorFlow
3. Keras
4. Flask

# 4. EXPERIMENTAL INVESTIGATIONS

One of major part of Artificial Intelligent (AI) algorithms to imitate human brain regarding data processing, pattern creation, and decision making is deep learning. It is a subset of machine learning in AI, that has high ability of leaning unstructured data. Deep learning is also called as deep neural network or deep neural learning. Comparing with AI and machine learning algorithms, deep learning algorithms are highly capable to learn large volume of unstructured data automatically. Hence, mainly deep learning methods have been used for bigdata analytics. Every region of the world using deep learning method for evolving and exploring digital era. The bigdata is obtained from

various resources like internet, social websites, online movies, e-commerce and search engines, and etc. These data are particularly available and can be accessed by any internet applications belongs to cloud computing. Since the data is unstructured, it is highly difficult to extract the relevant information from the data. Most of the IT industries trust that AI methods can incredibly extract the information from the data and it will be an automated support. Most of the earlier data processing industries are using machine learning method for bigdata analytics, and they are not fully automatic and self-adaptive. Deep learning is fully automatic, learn the data deeply, uses a hierarchical level of ANN to complete a task based on the machine learning process. ANN is created in accordance to the human brain, with neuron nodes interconnected together like www. Conventional methods analyse the data linearly. But the hierarchical method of deep learning process enables machine to process data using a non-linear approach. In terms of complex dataset like fraud detection, spatial, temporal and other geo-space data, deep learning methods can fetch the important features like time, geographical locations, IP addresses, and the type of retailer. More number of layers in the deep learning method performs their functions efficiently and classify the data. Due to more advantages, this paper utilizes Convolution Neural Network to analyse the breast data.

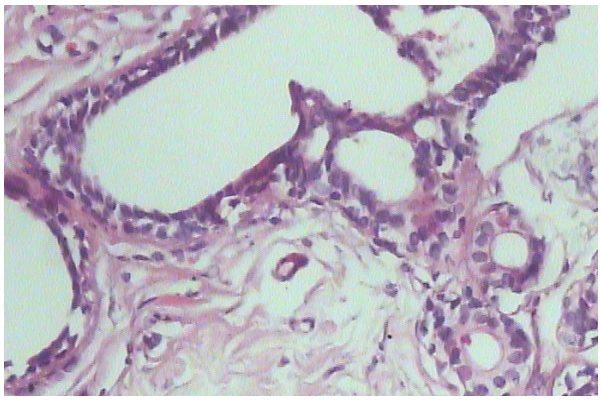
# 5. FLOWCHART



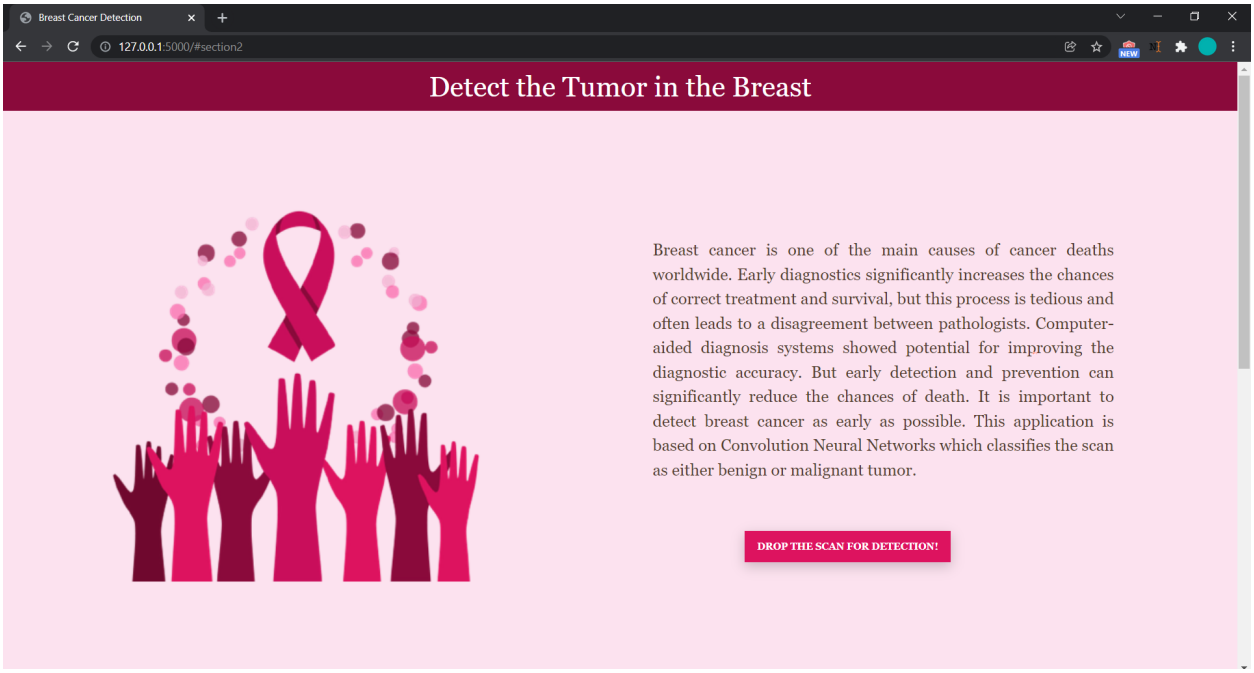
# 6. RESULT

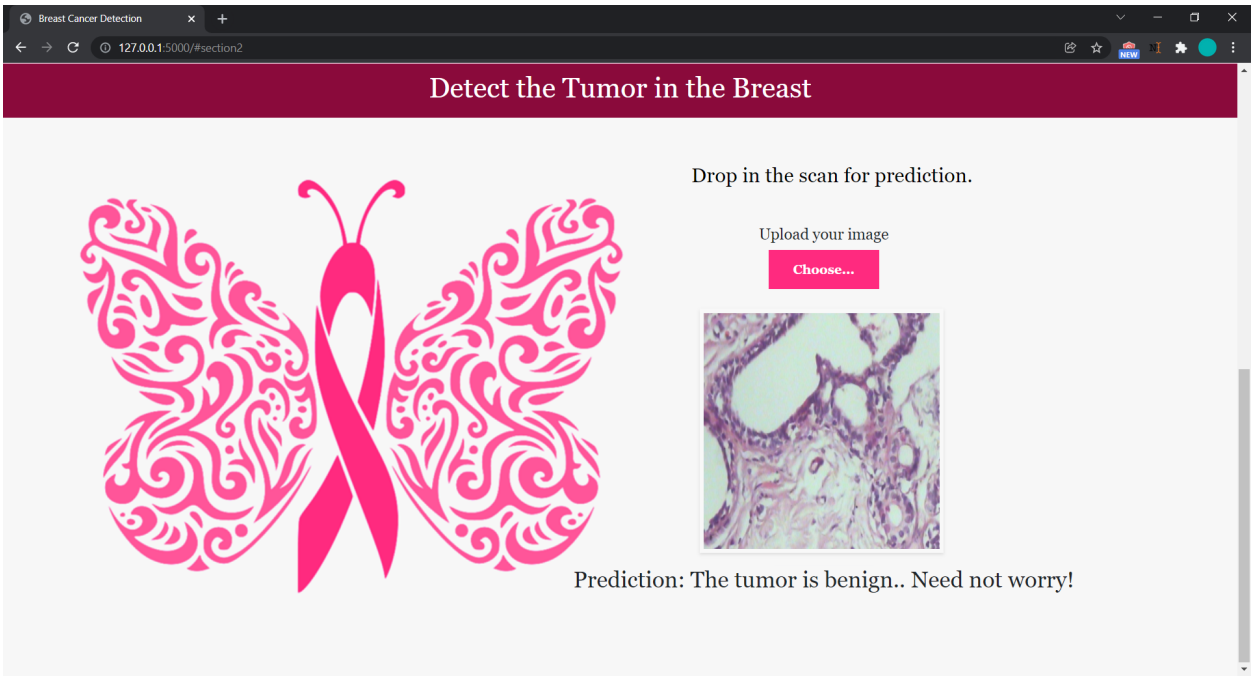
1. The overall accuracy for breast cancer diagnosis achieved equal to 28.77 percent .
2. The dataset showed improved accuracy obtained by CNN when compared to traditional learning techniques.
3. Deep CNN diagnosis provides a second option for image diagnosis which can improve the reliability of expert decision making.
4. The following snapshots define the results or outputs that we will get after step by step execution of each proposed protocol.

Input Image:



Website:






# 7. ADVANTAGES & DISADVANTAGES

Advantage:

- The proposed system is found to be successful, achieving results with high accuracy, which could reduce human mistakes in the diagnosis process.
- Breast Cancer detection Using Convolutional Neural Networks (BCDCNN) is aimed to speed up the diagnosis process by assisting specialist to diagnosis and classification the breast cancer.

Disadvantage:

- The accuracy must be high otherwise it can lead to false diagnosis and can delay treatment.

# 8. APPLICATIONS

This technology can be applied in several other cases:

1. Decoding Facial Recognition:

Facial recognition is broken down by a convolutional neural network into the following major components -

- Identifying every face in the picture
- Focusing on each face despite external factors, such as light, angle, pose, etc.
- Identifying unique features

Comparing all the collected data with already existing data in the database to match a face with a name.

2. Analyzing Documents

Convolutional neural networks can also be used for document analysis. This is not just useful for handwriting analysis, but also has a major stake in recognizers. For a machine to be able to scan an individual's writing, and then compare that to the wide database it has, it must execute almost a million commands a minute. It is said with the use of CNNs and newer models and algorithms, the error rate has been brought down to a minimum of 0.4% at a character level, though it's complete testing is yet to be widely seen.

3. Historic and Environmental Collections

CNNs are also used for more complex purposes such as natural history collections. These collections act as key players in documenting major parts of history such as biodiversity, evolution, habitat loss, biological invasion, and climate change.

4. Understanding Climate

CNNs can be used to play a major role in the fight against climate change, especially in understanding the reasons why we see such drastic changes and how we could experiment in curbing the effect. It is said that the data in such natural history collections can also provide greater social and scientific insights, but this would require skilled human resources such as researchers who can physically visit these types of repositories. There is a need for more manpower to conduct deeper experiments in this field.

5. Grey Areas

Introduction of the grey area into CNNs is posed to provide a much more realistic picture of the real world. Currently, CNNs largely function exactly like a machine, seeing a true and false value for every question. However, as humans, we understand that the real world plays out in a

1000 shades of grey. Allowing the machine to understand and process fuzzier logic will help it understand the grey area us humans live in and strive to work against. This will help CNNs get a more holistic view of what human sees.

6. Advertising

CNNs have already brought in a world of difference to advertising with the introduction of programmatic buying and data-driven personalized advertising.

7. Other Interesting Fields

CNNs are poised to be the future with their introduction into driverless cars, robots that can mimic human behavior, aides to human genome mapping projects, predicting earthquakes and natural disasters, and maybe even self-diagnoses of medical problems. So, you would not even have to drive down to a clinic or schedule an appointment with a doctor to ensure your sneezing attack or high fever is just the simple flu and not symptoms of some rare disease. One problem that researchers are working on with CNNs is brain cancer detection. The earlier detection of brain cancer can prove to be a big step in saving more lives affected by this illness.

# 9. CONCLUSION

The present work is based on the classification of breast cancer using capsule net architecture. From this work, it is clear that the performance of the conventional architectures can be improved by data pre-processing and parameter tuning.

The results show that this method can be used as an automated tool to assist doctors in disease diagnosis, which may lead to higher concentration in the treatment at early stages rather than diagnosis and can increase the cancer survival rate.

As it is difficult to detect the breast cancer in early stages, doctors can use CNN as second opinion. CNN provides accuracy and quality compared to other methods used to predict the breast cancer tumours as Benign and Malignant. CNN can be used as a second opinion by the doctors to diagnose the patients.

# 10. FUTURE SCOPE

- One direction is to improve the recognition accuracy of features using patches, size and the overlapping of patches.
- stain normalization, deeper architectures, and splitting the network before the last fully-connected layer could be investigated.
- Additional data with increased number of patients should be introduced.
- A better investigation on feature and classifier selection could also improve performance.

# 11. BIBLIOGRAPHY


- https://www.kaggle.com/paultimothymooney/breast-histopathology-images/
- Nrea, Simon & Gezahegn, Yaecob & Sinamo, Abiot & Hagos, Gebrekirstos. (2020). Breast Cancer Detection Using Convolutional Neural Networks.
- Breast cancer histopathological image classification using convolutional neural networks," in 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, jul 2016, pp. 2560–2567.


# 12. APPENDIX

**Source Code:**

*Link*