

# Cancer Detection Using Deep Learning

Khan Tawheed<sup>1</sup>, Khan Salman<sup>2</sup>, Aadil Farooqui<sup>3</sup>, Prof. Sachin Chavan<sup>4</sup>

<sup>1,2,3</sup>Student, Dept. of Computer Engineering, M.G.M. College of Engineering and Technology, Kamothe, Maharashtra, India

<sup>4</sup>HOD. Dept. of Computer Engineering, M.G.M College of Engineering and Technology, Kamothe,

\*\*\*

**Abstract** - In this paper, we first describe the basics of the field of cancer diagnosis, which includes steps of cancer diagnosis followed by the typical classification methods used by doctors, providing a historical idea of cancer classification techniques to the readers. These methods include Asymmetry, Border, Color and Diameter (ABCD) method, seven-point detection method, Menzies method, and pattern analysis. They are used regularly by doctors for cancer diagnosis, although they are not considered very efficient for obtaining better performance. Moreover, considering all types of audience, the basic evaluation criteria are also discussed. The criteria include the receiver operating characteristic curve (ROC curve), Area under the ROC curve (AUC), F1 score, accuracy, specificity, sensitivity, precision, dice-coefficient, average accuracy, and Jaccard index. Previously used methods are considered inefficient, asking for better and smarter methods for cancer diagnosis. Artificial intelligence and cancer diagnosis are gaining attention as a way to define better diagnostic tools. In particular, deep neural networks can be successfully used for intelligent image analysis. The basic framework of how this machine learning works on medical imaging is provided in this study, i.e., pre-processing, image segmentation and post-processing. The second part of this manuscript describes the different deep learning techniques, such as convolutional neural networks (CNNs), generative adversarial models (GANs), deep auto-encoders (DANs), restricted Boltzmann machine (RBM), stacked auto-encoders (SAE), convolutional auto-encoders (CAE), recurrent neural networks (RNNs), long short-term memory (LSTM), multi-scale convolutional neural network (M-CNN), multi-instance learning convolutional neural network (MIL-CNN). For each technique, we provide Python codes, to allow interested readers to experiment with the cited algorithms on their own diagnostic problems. The third part of this manuscript compiles the successfully applied deep learning models for different types of cancers. Considering the length of the manuscript, we restrict ourselves to the discussion of breast cancer, lung cancer, brain cancer, and skin cancer. The purpose of this bibliographic review is to provide researchers opting to work in implementing deep learning and artificial neural networks for cancer diagnosis a knowledge from scratch of the state-of-the-art achievements.

**Key Words:** Deep learning, Image processing, Image recognition. (KNN) algorithm.

## 1. INTRODUCTION

Cancer is the leading cause of deaths worldwide. Both researchers and doctors are facing the challenges of fighting cancer. According to the American cancer society, 96,480 deaths are expected due to skin cancer, 142,670 from lung cancer, 42,260 from breast cancer, 31,620 from prostate cancer, and 17,760 deaths from brain cancer in 2019 (American Cancer Society, new cancer release report 2019). Early detection of cancer is the top priority for saving the lives of many. Typically, visual examination and manual techniques are used for these types of a cancer diagnosis. This manual interpretation of medical images demands high time consumption and is highly prone to mistakes. For this reason, in the early 1980s, computer-aided diagnosis (CAD) systems were brought to assist doctors to improve the efficiency of medical image interpretation. Feature extraction is the key step to adopt machine learning. Different methods of feature extraction for different types of cancer have been investigated in. However, these methods based on feature extraction have weaknesses. To overcome these weaknesses and to enhance the performance, representation learning has been proposed in. Deep learning has the advantage of generating directly from raw images the high-level feature representation. In addition to deep learning, Graphics Processing Units (GPU) are also being used in parallel, for feature extraction and image recognition. For example, convolutional neural networks have been able to detect cancer with promising performance. To test these algorithms, there are publicly available datasets. These include IN breast and Break His for breast cancer testing; Digital Database for Screening Mammography (DDSM) for mass detection; MITOSTAPIA for mitosis detection; Japanese Society of Radiological Technology (JSRT), The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI), and Danish Lung Cancer Screening Trial (DLCST) for lung nodule classification; multimodal Brain Tumor Segmentation challenge (Bra TS) for brain cancer identification; and Dermoscopy Image Segmentation (DermIS) as well as data

given to the public by International Skin Image Collaboration (ISIC) for skin cancers.

## 2. PROBLEM STATEMENT

Breast Cancer is one of the leading cancers developed in many countries including India. Though the endurance rate is high — with early diagnosis 97% women can survive for more than 5 years. Statistically, the death toll due to this disease has increased drastically in last few decades. The main issue pertaining to its cure is early recognition. Hence, apart from medicinal solutions some Data Science solution needs to be integrated for resolving the death causing issue. This analysis aims to observe which features are most helpful in predicting malignant or benign cancer and to see general trends that may aid us in model selection and hyper parameter selection. The goal is to classify whether the breast cancer is benign or malignant. To achieve this is have used machine learning classification methods to fit a function that can predict the discrete class of new input.

In addition to breast cancer, deep learning has found its use in lung cancer as well. Some of the studies which have applied deep learning for this purposed are discussed in this section. Patients' survival time was successfully predicted using deep convolutional neural networks by Zhu et al. directly from the lung cancer pathological images. A pretrained CNN, which was trained on a large-scale data, was adopted by Paul et al., for the detection of lung cancer by extracting features from the CT images. On the raw images of the lung, the DBN and CNN were applied with end-to-end learning. They used 2D CT images for pulmonary node classifications, whereas, in, researchers used 3D CT images on multi-view CNN, which were could be used for end-to-end training. They extracted the 2D patches from the 3D images and used them on CNN for feature extraction. The features were fed to the classifier after fusing them together. As observed in the research study by Dou et al., they formed a model with CNN, which dealt with 3D images directly instead of mapping them into a 2D model.

Brain cancer has an uncontrolled growth and it may occur in any part of the brain. It has been quite challenging to detect which part of the brain contains cancer. Consequently, the biggest challenge for brain cancer is the segmentation of the brain from the healthy part. Several challenges have been conducted by BRATS for this purpose. Here, we include some of the research work in which deep learning has been successfully applied to brain

images. Two algorithms based on 2D CNN and 3D CNN were proposed by Gao et al., working on 2D sliced images and 3D images, respectively. The final result was obtained by fusing the output from these two models. This hybrid model outperformed the 2D and 3D scale-invariant features swift (SIFT) and Kaze features. The automatic magnetic resonant image segmentation method based on CNN was discussed by Author1. They investigated the intensity normalization and augmentation for brain tumor detection. By exploring the local and global contextual features in the CNN model, Havaej et al. used a fully connected layer in the final layer of CNN to increase the speed of the system and detected the brain cancer successfully. A fully connected convolutional neural network (FCN) and conditional random field (CRFs) were used in for the brain cancer segmentation. First, the image patches were used to train the FCN model and the training of CRF was done. In the end, the system was finely tuned using the image slices directly. Adjacent image patches were joined together into one pass using a dense training scheme in the CNN model. The false positives were removed by using the 3D fully connected random field, after the 3D segmentation of the images using modality of CNN. Zhao et al. combined the multi-modality information from T1, T1C, T2, and fluid-attenuated inversion recovery (FLAIR) images and trained the proposed CNN from this information. The algorithm proposed by them was a 3D voxel classification based on CNN. Different scaled 2D patches were extracted from 2D slices obtained by slicing the 3D dataset and these 2D patches were fed to multiple CNNs for the learning process.

## 3. EXISTING SYSTEM

In this project we will use Data Mining and Machine Learning Algorithms to detect cancer, based off of data. Cancer is a common cancer for women around the world. Early detection of BC can greatly improve prognosis and survival chances by promoting clinical treatment to patients. We will use the UCI Machine Learning Repository for breast cancer dataset. The dataset used in this story is publicly available and was created by Dr. William H. Walberg, physician at the University of Wisconsin Hospital at Madison, Wisconsin, USA. To create the dataset Dr. Walberg used fluid samples, taken from patients with solid breast masses and an easy-to-use graphical computer program called Xcyt, which is capable of perform the analysis of cytological features based on a digital scan. The program uses a curve-fitting algorithm, to compute ten features from each one of the cells in the sample than it

calculates the mean value, extreme value and standard error of each feature for the image, returning a 30 real-valued vector Attribute Information:

1. ID number 2) Diagnosis (M = malignant, B = benign) 3—32) Ten real-valued features are computed for each cell nucleus:

2. radius (mean of distances from centre to points on the perimeter)

3. texture (standard deviation of Gray-scale values)

4. perimeter.

5. area.

6. smoothness (local variation in radius lengths)

7. compactness ( $\text{perimeter}^2 / \text{area} - 1.0$ )

8. concavity (severity of concave portions of the contour)

9. concave points (number of concave portions of the contour)

10. symmetry.

11. fractal dimension (coastline approximation — 1) The mean, standard error and worst or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

#### 4. LITERATURE SURVEY

Twenty-four recent research articles have been reviewed to explore the computational methods to predict breast cancer. The summaries of them are presented below. Chaurasia et al. developed prediction models of benign and malignant breast cancer. Wisconsin breast cancer data set was used. The dataset contained 699 instances, two classes (malignant and benign), and nine integer-valued clinical attributes such as uniformity of cell size. The researchers removed the 16 instances with missing values from the data set to become the data set of 683 instances. The benign were 458 (65.5%) and malignant were 241 (34.5%). The experiment was analysed by the Waikato Environment for Knowledge Analysis (WEKA). Naive Bayes, RBF Network, and J48 are the three most popular data mining algorithms were used to develop the prediction models. The researchers used 10-fold cross-validation methods to measure the unbiased estimate of the three prediction models for performance comparison purposes. The models' performance evaluation was presented based on the methods"

effectiveness and accuracy. Experimental results showed that the Naive Bayes had gained the best performance with a classification accuracy of 97.36%; followed by RBF Network with a classification accuracy of 96.77% and the J48 was the third with a classification accuracy of 93.41%. In addition, the researchers conducted sensitivity analysis and specificity analysis of the three algorithms to gain insight into the relative contribution of the independent variables to predict survival. The sensitivity results indicated that the prognosis factor Class was by far the most important predictor.

#### 5. PROPOSED SYSTEM

##### 5.1 Data Mining and Machine Learning

The term "data mining" is a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence (e.g., machine learning) and business intelligence. The book Data mining: Practical machine learning tools and techniques with Java [8] (which covers mostly machine learning material) was originally to be named just Practical machine learning, and the term data mining was only added for marketing reasons. Often the more general terms (large scale) data analysis and analytics — or, when referring to actual methods, artificial intelligence and machine learning — are more appropriate. In this project we use the following machine learning algorithms: Decision tree algorithms: Decision tree algorithms are successful machine learning classification techniques. They are the supervised learning methods which use information gained and pruned to improve results. Moreover, decision tree algorithms are commonly used for classification in many researches, for example, in the medicine area and health issues. There are many kinds of decision tree algorithms such as ID3 and C4.5. However, J48 is the most popular decision tree algorithm. J48 is the implementation of an improved version of C4.5 and is an extension of ID3. K-nearest-neighbours

(KNN) algorithm: It is a simple supervised learning algorithm in pattern recognition. It is one of the most popular neighbourhood classifiers due to its simplicity and efficiency in the field of machine learning. KNN algorithm stores all cases and classifies new cases based on similarity measures; it searches the pattern space for the k training tuples that are closest to the unknown tuples. The

performance depends on the optimal number of neighbours (k) chosen, which is different from one data sample to another.

Support Vector Machine (SVM): It is a supervised learning method derived from statistical learning theory for the classification of both linear and nonlinear data. SVM classifies data into two classes over a hyperplane at the same time avoiding over-fitting the data by maximizing the margin of hyperplane separating.

Naïve Bayes (NB) It is a probabilistic classifier: It is one of the most efficient classification algorithms based on applying Bayes' theorem with strong (naïve) independent assumptions. It assumes the value of the feature is independent of the value of any other features,

Neural networks are capable of performing the tasks of complex computation because of the nonlinear processing of neurons. An artificial neural network. As the artificial neural network has the power of prediction, it can be used for medical images. In a general artificial neural network, test images are given to the neurons for training. To train neurons, back-propagation algorithm is used, with the flow in the forward direction. This error propagates in the backward direction. Weights are adjusted for error reduction. This processing is repeated until the error becomes zero. There is a layered structure in the neural network with the number of interconnected nodes and an activation function among them. These activation functions are tangent hyperbolic function, sigmoid function, piece-wise linear function, and threshold function. Input patterns are presented to the network through an input layer, which then connects to the hidden layer, and this hidden layer connects to the output layer.

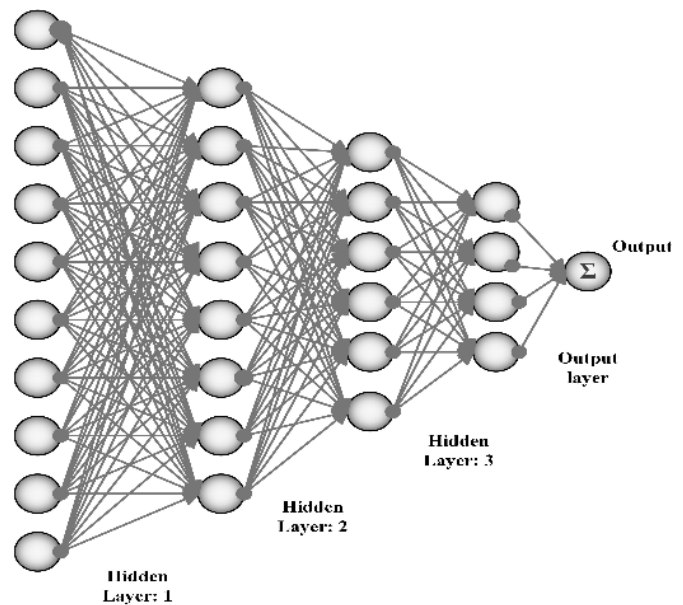


Fig - 1: Neural Network

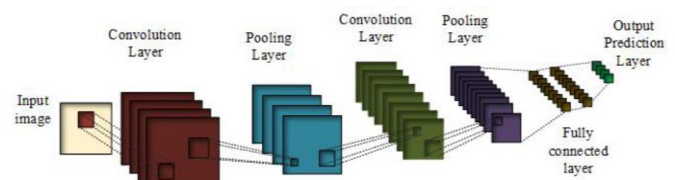


Fig - 2: Convolution Neural Network

The basic purpose of pooling in CNN is the task of subsampling i.e., it summarizes the nearby neighbourhood pixels and replaces them in the output at a location with summarized characteristics. Pooling reduces the dimensionality and performs the invariance of rotational transformations and translation transformations. There are many pooling functions one of the most famous is max pooling, in which the output is the maximum value of the rectangular pixel neighbourhood. In average pooling function, the output becomes the average of the rectangular neighbourhood. Another type consists of the weighted average based on the distance from the central pixel. Pooling helps to make the representation invariant to small changes to the translation in the input.

## 6. IMPLEMENTATION

We will first go with importing the necessary libraries and

We can examine the data set using the pandas head () method.

```
df.head(7) {first 7 rows of the data}
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compa
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	
3	84348301	M	11.42	20.38	77.58	366.1	0.14250	
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	
5	843786	M	12.45	15.70	82.57	477.1	0.12780	
6	844359	M	18.25	19.98	119.60	1040.0	0.09463	

Fig - 3: Patient Data 1

perimeter_se	area_se	smoothness_se	compactness_se	concavity_se	concave points_se	symmetry_se	fractal_dimension_se	radius_worst	texture_worst	perimeter_worst	area_worst	smoothness_worst	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst	radius_worst	texture_worst	perimeter_worst	area_worst	smoothness_worst	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst	radius_worst	texture_worst	perimeter_worst	area_worst	smoothness_worst	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst	
8.589	153.40	0.006389	0.04804	0.05373	0.01587	0.03003	0.006193	25.38	17.33	184.60	2019.0	0.00532	0.03652	24.99	23.41	158.80	1856.0	0.004571	23.57	25.53	152.50	1789.0	0.005208	0.03236	14.91	26.50	98.87	567.7	0.005115	22.54	16.67	152.20	1575.0	0.005892	15.47	23.75	103.40	741.6
3.389	74.98	0.005225	0.01308	0.01660	0.01340	0.01389	0.003532	24.99	23.41	158.80	1856.0	0.00532	0.03652	24.99	23.41	158.80	1856.0	0.004571	23.57	25.53	152.50	1789.0	0.005208	0.03236	14.91	26.50	98.87	567.7	0.005115	22.54	16.67	152.20	1575.0	0.005892	15.47	23.75	103.40	741.6
4.585	94.03	0.008150	0.04006	0.03832	0.02058	0.02250	0.004571	23.57	25.53	152.50	1789.0	0.005208	0.03236	14.91	26.50	98.87	567.7	0.005115	22.54	16.67	152.20	1575.0	0.005892	15.47	23.75	103.40	741.6											
3.445	27.23	0.009110	0.07458	0.05661	0.01687	0.05863	0.005208	14.91	26.50	98.87	567.7	0.005115	22.54	16.67	152.20	1575.0	0.005892	15.47	23.75	103.40	741.6																	
5.438	94.44	0.011480	0.02461	0.05688	0.01885	0.01756	0.005115	22.54	16.67	152.20	1575.0	0.005892	15.47	23.75	103.40	741.6																						
2.217	27.19	0.007510	0.03345	0.03872	0.01137	0.02165	0.005892	15.47	23.75	103.40	741.6																											
3.180	53.91	0.004214	0.01382	0.02254	0.01039	0.01389	0.002179	22.88	27.86	153.20	1606.0																											

Fig - 4: Patient Data 2

concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst	Unnamed: 32
0.7119	0.2654	0.4601	0.11890	NaN
0.2416	0.1860	0.2750	0.08902	NaN
0.4504	0.2430	0.3613	0.08758	NaN
0.6869	0.2575	0.6638	0.17300	NaN
0.4000	0.1625	0.2364	0.07678	NaN
0.5355	0.1741	0.3985	0.12440	NaN
0.3784	0.1932	0.3063	0.08368	NaN

Fig - 5: Patient Data 3

```

id 0
diagnosis 0
radius_mean 0
texture_mean 0
perimeter_mean 0
area_mean 0
smoothness_mean 0
compactness_mean 0
concavity_mean 0
concave points_mean 0
symmetry_mean 0
fractal_dimension_mean 0
radius_se 0
texture_se 0
perimeter_se 0
area_se 0
smoothness_se 0
compactness_se 0
concavity_se 0
concave points_se 0
symmetry_se 0
fractal_dimension_se 0
radius_worst 0
texture_worst 0
perimeter_worst 0
area_worst 0
smoothness_worst 0
compactness_worst 0
concavity_worst 0
concave points_worst 0
symmetry_worst 0
fractal_dimension_worst 0
Unnamed: 32 569
dtype: int64

```

Fig - 6: Column

We can find the dimensions of the data set using the panda dataset shape attribute.

```
df.shape (569, 33)
```

We can observe that the data set contain 569 rows and 33 columns. Diagnosis is the column which we are going to predict which says if the cancer is M = malignant or B = benign. 1 means the cancer is malignant and 0 means benign.

### 7. RESULT

We can identify that out of the 569 persons, 357 are labelled as B (benign) and 212 as M (malignant).

Each row represents a patient and 33 features on the 569 patients.

The last column Unnamed 32 has NaN values so we need to remove that column with empty values.

So, we count the number of empty columns and drop the columns with empty values.

So, column Unnamed: 32 has 569 missing values so we drop it. So, the new shape of the data is (569, 32) which means 569 rows and 32 columns. Now we can see the number of Malignant (M) (harmful) or Benign (B) cells (not harmful) cells and plot it in a graph.

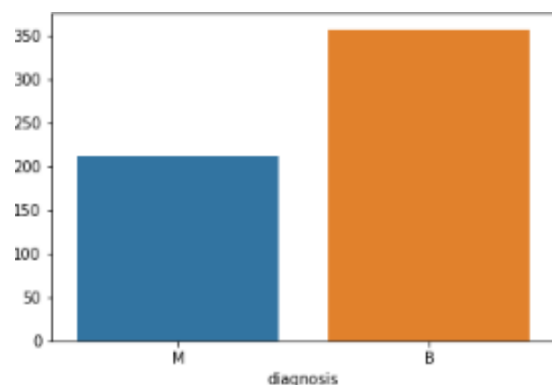


Fig - 7: Graph Presentation



So, we can see that the Decision Tree Classifier has the best accuracy among all the models ie.100%. We will now predict the test set results and check the accuracy with each of our model:

To check the accuracy, we need to import confusion matrix method of metrics class. The confusion matrix is a way of tabulating the number of mis-classifications, i.e., the number of predicted classes which ended up in a wrong classification in based on the true classes.

```

[[86  4]
 [ 4 49]]
Model[0] Testing Accuracy = "0.9440559440559441!"

[[89  1]
 [ 5 48]]
Model[1] Testing Accuracy = "0.958041958041958!"

[[87  3]
 [ 2 51]]
Model[2] Testing Accuracy = "0.965034965034965!"

[[85  5]
 [ 6 47]]
Model[3] Testing Accuracy = "0.9230769230769231!"

[[84  6]
 [ 1 52]]
Model[4] Testing Accuracy = "0.951048951048951!"

[[87  3]
 [ 2 51]]
Model[5] Testing Accuracy = "0.965034965034965!"

```

Fig - 12: Result

Here the matrices are of form [TP FP] [FN TP] were

TP is true positive: A true positive is an outcome where the model correctly predicts the positive class TN is true negative: A true negative is an outcome where the model correctly predicts the negative class. FN is false negative: A false negative is an outcome where the model incorrectly predicts the negative class. FP is false positive: A false positive is an outcome where the model incorrectly predicts the positive class. Based on the test data we can see that Model 5 ie Random forest classifier has 96.5% accuracy on the test data so we can use it to predict the actual outcome whether a patient has cancer or not.

Prediction of Model

```
pred = model [5]. predict(X_test) print(pred)
```

```

[1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 1 1 1 1 1 1 0 0 1 0 0 1 0 1 0 1 0 1 0 1 0
1 0 1 0 0 1 0 0 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 0 1 0 1 1 1 0 0 1 0 0
1 0 0 0 0 0 1 1 1 0 1 0 0 0 1 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 1 1 1 0
1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 1]

[1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 0 0 1 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0 1 1 0 1 0 0 1 0 0 0 1 1 1 1 0 0 0 0 0 0 1 1 1 0 0 1 0 1 1 1 0 0 1 0 1
1 0 0 0 0 0 1 1 1 0 1 0 0 0 1 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 1]

```

Fig - 13: Prediction of Data

So here we printed the predictions. The first data shows the actual result of which patient had cancer and the second data is the one predicted by the model.

The accuracy of the model is 96.5% so we can see a few wrong predictions but mostly this model is successful in predicting a tumour Malignant (M) (harmful) or Benign (B) (not harmful) based upon the features provided in the data and the training given.

### 8. CONCLUSIONS

This review focuses on providing all the necessary information to the beginners of this field, starting from the main concepts of cancer diagnosis, evaluation criterion and medical methods. As this manuscript mainly focuses on the deep learning for cancer diagnosis, the most important things to introduce to our readers are all the possible techniques of deep learning that can be used for diagnostic purposes in this document. Furthermore, to facilitate the audience, the respective practice codes for each technique, which are easily available online, are put together in a table. One of the major issues that one can encounter in implementing any algorithm is the dataset availability, therefore all possible access links to the datasets are presented in this work. Different architectures of CNN are also described in this manuscript. The implementation of the deep learning algorithms for brain cancer, lung cancer, breast cancer, and skin cancer is the focus of this manuscript. The performance measures for different studies are provided. In this review, different deep learning algorithms for classifying different types of cancers are presented. In this review, fifteen studies used Histopath model with CNN for classification and detection of different types of cancers as provided. Six of these studies provided the source of data while nine studies did not publish the source of data. Two research studies used mammographs for detection along with CNN and published data source. Eight studies used CT Slices, three of which used data from PROMISE and LIDC. Five studies used volumetric computed tomography. Seven studies were for brain cancer classification.

In this project in python, we learned to build a breast cancer tumour predictor on the Wisconsin dataset and created graphs and results for the same. It has been observed that a good dataset provides better accuracy. Selection of appropriate algorithms with good home dataset will lead to the development of prediction systems. These systems can assist in proper treatment methods for a patient diagnosed with breast cancer. There are many

treatments for a patient based on breast cancer stage; data mining and machine learning can be a very good help in deciding the line of treatment to be followed by extracting knowledge from such suitable databases.

## REFERENCES

- [1] Sharma, S.V., Haber, D. A. & Dettleman, J. Cell Line-based platforms to evaluate the therapeutic agents. *Nat. Rev. Cancer* 10(4),241-253 (2010).
- [2] Liu, Z., Delavan, B., Roberts, R. & Tong, W. Lessons learned from two decades of anticancer drugs. *Trends Pharmacol. Sci.* 38(10), 852-872 (2017).
- [3] Masters, J. R. W. Cell line misidentification: the beginning of the end. *Nat. Rev. Cancer* 10(6), 441-448 (2010).
- [4] Lohar, P.S. *Textbook of Biotechnology* (MJP Publishers, New Delhi, 2019)
- [5] Michaelis, M., Wass, M. N. & Cinatl, J. Drug-adapted cancer cell lines as preclinical models of acquired resistance. *Cancer Drug Resist.* 2, 447-456 (2019).
- [6] Engelman, J. A. et al. Met amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* (80-) 316(5827), 1039-1043 (2007).