Breast Cancer Prediction Using Data Mining Classification Techniques

Abdul Karim Kazi*[1], Waseemullah[2], Mirza Adnan Baig[3] *and* Shahzaib Khan[4]

Department of Computer Science and Information Technology, NED University, Karachi, 75270, Pakistan

Abstract
In today's world Breast Cancer has become the major source of mortality among women especially in underdeveloped countries like Pakistan, Sri Lanka, and Bangladesh. This is a highly alarming situation and needs the attention of the research community as there are not enough resources and health facilities. The rate of incidence could be reduced if the cancer is diagnosed at an early stage instead of late stages. Breast cancer occurs when some breast cells begin to rise abnormally. This research study intends to predict breast cancer by analyzing a set of attributes that have been selected from several classifications so that prevention can be done in time before it becomes incurable. This research work focuses on different classification techniques of data mining to predict Breast Cancer such as Decision Trees, Random Forest, Logistic Regression, Support Vector Machine, and Linear Discriminant Analysis and their comparative analysis for accurate disease detection. The dataset of breast cancer histopathology images was acquired from online recourses consisting of 277,524 images. The experimental results show that Random Forest performs better than all other algorithms used in this research study with an accuracy of 88.80 %, precision of 83.71 %, and recall 94.28 %. Python programming language to implement and perform the comparative analysis of algorithms used in this research work.
Keywords:
Data Mining; Breast Cancer; Computer Vision; Deep Learning; Disease Prediction.

1. Introduction

In today's world Cancer is a major health issue. As stated by the IARC (International Agency for Research on Cancer) of the World Health Organization (WHO), in 2012 there were 8.2 million deaths due to cancer and around 27 million new cases of cancer are expected till 2030 [1]. Specifically, Breast Cancer (BC) is the most common cancer among women. Its mortality rate is quite high as compared to other types of cancer. Breast cancer is the deadliest cancer among women. For the treatment of these patients, screening has been done. The various method has been used to identify the disease including mammography, histopathology, clinical exam, or self-exam. Imaging procedures can achieve the detection and diagnosis of BC such as ultrasound (sonography), magnetic resonance imaging, diagnostic mammograms (X-rays), and thermography [2]. Cancer screening via Imaging has been investigated for more than four decades [3].

Manuscript received September 5, 2022
Manuscript revised September 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.9.91

In this research study histopathology images are used. Therefore, it's important to discuss why histopathology image? A biopsy is the only method to diagnose cancer. Among biopsy methods, a histopathology image is one of the techniques that is the gold standard in diagnosing all kinds of cancer [4-5]. The final BC diagnosis, such as staging and grading, is performed by pathologists where visual inspection of histological samples is carried out. Histopathological evaluation is an extremely time-taking technical task, totally dependent on the pathologists and factors such as lack of attention and fatigue. There's a need for CAD (Computer-Assisted Identification) to reduce work stress on pathologists [6] from filtering clearly benign regions so that the specialists can concentrate on the other difficult to diagnose cases [7].

According to the latest report from the IARC, there are a number of new cases of cancer diagnosed globally, and many of them lost their lives from this disease [8]. One of them is Breast Cancer which causes death in women. It is the rarest cancer in the world, and it is increasing in incidence. The reason for this could be lack of awareness and poor initial stage diagnosis rate. Unfortunately, most breast cancers are diagnosed when they are in the advanced stages.

The data mining classification techniques i.e., Convolutional Neural Networks, Decision Tree, Logistic Regression, Random Forest, Linear Discriminant Analysis, and Support Vector Machine on the breast cancer data. The results are compared on the basis of parameters i.e., accuracy, precision, specificity, sensitivity, error rate, and confusion metric are also computed. The obtained results are presented and discussed in section V.

This paper contains 6 sections including this section. Section I describes the Introduction of research work carried out in this paper. Section II describes related work that has been done in relation to Data Mining Classification Techniques. Section III describes the implementation of the Methodology and a brief overview Experimental Setup and Breast Cancer Dataset used for the experiment. Section IV presents the results of experiments that were conducted. Section V presented the conclusion and future work.

2. Background

In this section, four major domains are going to be highlighted i.e., Machine Learning, Machine Learning models, Deep Learning, and Computer Vision.

In computer vision applications learning algorithms are used commonly. Let's look at the essentials of Machine Learning before considering image-related tasks. For modeling problems, Machine Learning is supposed to be a great tool for it. Such programs are programmed by hands to execute a task called Classical computer programs. But with machine learning, by a learning algorithm, a very small part of human involvement is being replaced [9]. As the data increases and accessibility of computational capacity, machine learning has become increasingly more functional through time, that's why it becomes more ubiquitous. A well-known and common method is supervised learning in machine learning [10]. Various samples have been marked by a human. E.g., like the object detection problem, we utilize pictures as training where the expert has to mark the interest points and area of related objects. Classifier learned from those points, the algorithm can predict the labels of previously unseen data. The most important types of these tasks are Classification and Regression. In Classification, the algorithm tries to predict which class is correct on the basis of training data. In regression, the algorithm attempts to predict from a constant output. On the other hand, an unsupervised learning algorithm tries to learn valuable features of the given data set on its own, without any expert system telling exactly what the proper output ought to be. Clustering is one the best example of unsupervised learning. Recently, Unsupervised pre-processing becomes a more famous tool in supervised learning, after the influx of deep learning [11].

Some type of pre-processing is nearly always required. Pre-processing the data to an easier and brand-new variable space is known as feature extraction. Most of the time, it's impossible or impractical to utilize the whole information from the training directly. Instead, detectors are designed to extract features of interest from the provided training data. Those features will be taken as input to the ML algorithm. For many years, humans were doing feature detection. The main problem behind this method is that we don't always know before, which attributes are needed. This also made a trend in machine learning feature detection.

Usually, the training data doesn't mention each promising feature of these input data, so it's must have the skill in learning algorithms to generalize to manage hidden data points. There is a high chance that a simple model could not capture important details of the real model. On the other hand, there is also a chance that a complicated approach can over-fit insignificant noise and details, which led to poor generalization. Usually, over-fitting occurs when a difficult approach is used on a very small training data set.

Neural networks are also known as artificial neural networks. For the reason that it develops to predict neural functioning and works like a human brain. Original research contains the brink logic unit from Walter Pitts and Warren McCulloch along with Frank Rosenblatt's Perceptron [12]. The human brain comprises around 100 billion neurons and works along parallel [13]. Basically, artificial neurons are mathematical functions that work on classical computers.

Neural networks are also known as artificial neural networks. For the reason that it develops to predict the neural functioning and works like a human brain. Original research contains the brink logic unit from Walter Pitts and Warren McCulloch along with Frank Rosenblatt's Perceptron. The human brain comprises around 100 billion neurons and works along parallel. Basically, artificial neurons are mathematical-functions work on classical computers.

When Artificial neurons are combined, it is said to be a neural network. The neurons are arranged in layers. In a fully-connected network, the output of each neuron layer is considered as an input for the very next neuron layer. Therefore, some layers are working on the input data whereas others working on the computation of the data obtained from different neurons. Each neuron carries some weight equivalent to the last layer of the neuron. Usually, a multi-layer network consists of 3 kinds of layers: a hidden layer (maybe one or more), an input layer, and an output layer. The work of the input layer is to just pass on the data without any modifications. Majorly computation takes place when data comes in a hidden layer. The last layer is the output layer performs the task of conversion as an outcome given by the hidden layer. A network that has at least one hidden layer may operate as a universal approximate, which means it is designed to calculate any function [14].

Modern neural networks are known as deep neural networks. Although multi-layer neural networks exist since the 1980s, many reasons prevented the effective practice of networks with multiple hidden layers. One of the primary issues is the curse of dimensionality. Since the number of variables increases, there is exponential growth in the number of different configurations. As the configurations increase, the amount of training data examples also increases in equivalent measures. Collecting a sufficient sized data set for training is time-taking and expensive. From the past years, Neural Networks had a rebirth, due to the accessibility of huge data sets and more powerful computers. In early 2000, it was discovered that NN could be trained more efficiently via graphic processing unit (GPU). It is more efficient for processing than conventional CPUs [15]. Today, researchers typically utilize high-end consumer GPUs, for example, NVIDIA Tesla K80 [16]. There is no need to manually set machine learning parameters that were used before with the help of deep learning. A standard face detection system e.g., needs a feature to hand-tuned detection stage before a machine learning stage. Neural networks extract the important features which are forwarded to the next layer for further processing.

It uses in the extraction of meaningful information from a video or a still image. It is different from the picture processing. It comprises the handling of visual details on the level of the pixel. Computer vision has various applications like image classification, augmented reality [17] etc. Today, machine learning is an essential part of several computer vision algorithms [18]. It is a combination of Machine learning and image processing. Effective solutions need algorithms that deals contained with the huge number of images data set, and for many other applications, can execute and gives result in real-time [19]. It is one of the conventional issues of computers. In most aspects, it's very alike to other computer vision tasks, since it develops a solution that changes in light and views perspective. Object detection contains both classifying and finding regions of an image which makes it a different problem.

3. Methodology

In this section, the algorithms used in this comparative study are described along with the proposed mythology followed in this research work. The algorithms that are used in this research for comparative analysis are Decision Tree, Keras, Random Forest, Linear Discriminant Analysis, Logistic Regression, Support Vector Machine.

3.1 Decision Trees (DT)

A decision tree is nothing but a tree-structured classifier that is used in supervised learning algorithms. It is used in both regression models and classification. It learns to decide from the sample data set which gives to the model. DT working on nodes, one is called decision nodes used to describe the test node and that test is performed on the basis of the feature. The second node is called the leaf node which describes the real classes. DT might not perform well with a huge number of data sets [20-21].

3.2 Keras (CNN)

Keras is an interface that uses the backend engine like Tensorflow, Theano, and CNTK. It is specifically used in deep learning neural network machine learning. It is a library written in the Python programming language by a Google engineer. Keras uses an API that allows creating its own models and can perform many other tasks like specifying layers, etc. Moreover, it focuses on being user-friendly. It has the support for experimentation with a convolutional neural network [22-23].

3.3 Random Forest (RF)

It is very similar to the Decision Trees and a great example of an ensemble learning model. It is also used in classification and regression tasks. RF is a kind of ensemble classifier that uses DT algorithm in a randomized way. This creates a forest with a number of DT. That's why it's best for robustness and accuracy. It randomly picks samples from the training data set and puts them into the bootstrap data set. For the tree construction, it considers a subset of variables from the features randomly. Having the most votes will be considered in wholes classification. It is very similar to the election; those who get the most votes will win. It handles the missing values and maintains the accuracy over it. That's why it performs well in a large number of data set and also it won't over-fit the model [24-25].

3.4 Logistic Regression (LR)

It is very much like linear regression but it's not linear. LR is very popular in machine learning because it is very efficient and does not require too many computational resources. Also, it uses in traditional statistics problems. Logistic regression can solve linear and non-linear problems. Its dependent variable is binary and the independent variable can be continuous or binary. It is also called logistic regression. LR basic goal is to find the best relationship between Independent and dependent variables. Also, we can say that it deals with probability to measure the relationship of variables. It has the ability to classify new models and provide profanities using discrete and continuous functions for measurement which makes it more popular in machine learning [26-27].

3.5 Support Vector Machine (SVM)

It is a supervised learning algorithm used for classification and regression analysis. It draws a hyper-plane between the input data points. In SVM data points are also known as vectors. When a test data set is given that hyper-plane decides which class it belongs to. The nearest points near to the hyper-plane or the points close to the opposite class are called support vectors. It considers the maximum width/margin level between data points to classify and produce the best results and reduce misclassification. That's why it is best for extreme conditions and it takes a lot of time for computation. A kernel is used when a multi-dimensional problem occurs and to convert non-linear space into linear space. It has different kinds of kernels for various decision functions [28-29].

3.6 Linear Discriminant Analysis (LDA)

LDA is a supervised learning or classification technique. It is very much used as a technique for dimensionality reduction. This is a pre-processing step for ML and patterns classification applications. Its aim is to protect the data points onto a lower-dimensional space with fine class reparability and also to reduce computational cost and time. The original technique was developed by Ronald A. Fisher in 1936 and has been named Fisher's Discriminant Analysis or Linear Discriminant. Basically, the Linear Discriminant was defined as a two-class technique. The multi-class edition was later shown as Multiple Discriminant Analysis by C.R Rao. They're all simply referred to as the Linear Discriminant Analysis [30-31].

3.7 Proposed Technique

The proposed approach consists of three steps Database Management, Feature Extraction, and Model Creation and Testing Dataset on Classifier(s) as shown in fig. 1.

(1) *Dataset Management*: In this step, we download the data set and pre-process it. Although in our case images were already pre-processed (cropped into 50x50 px).

(2) *Feature Extraction & Model Creation*: In this step, we extract the features (Hu Moments, Haralick Texture, Color Histogram) from the pre-processed image data set. After then, applied normalization to standardize data and saved the training model into the h5py package. Hu Moments are used to describe, characterize, and quantify the shape of an object in an image whereas Haralick Feature it used to compute Haralick texture features from a gray level co-occurrence matrix (GLCM) is a common method to represent image texture [32]. The Color Histogram feature is also used for the representation of the distribution of colors in an image.

(3) *Testing Dataset on Classifier(s)*: In this step, we need to load the trained model and apply split and test to classify using ML classifiers (Decision Trees, Random Forest, Logistic Regression, Support Vector Machine, Linear Discriminant Analysis) and get results.

After getting the classification report we compare all models according to the resulting parameters that which classifier performed well. The implementation on Keras is originally done by Adrian Rosebrock [33]. Their configuration file using data set from Kaggle. To build the data set it stores images after creating training, testing, and validation split. Eighty percent (80%) of the data set is used for training and the remaining 20% for testing whereas the validation uses 10% of the training data. The next step is for the implementation of CNN using Keras deep learning library named "CancerNet" using Convolution filters, max-pooling, etc. The last part is to train the data set. It uses multiple libraries which include: matplotlib: To save training plots to disk, sklearn: Helps in implementation of classification report and confusion matrix and imutils and NumPy: Typically used for numerical processing.

3.7.1 Dataset

The data set used for this research is acquired from an online resource [34] and is consisting of 277,524 images in Portable Network Graphics (.png) format which are in the equal size of 50x50 px. Total of 198,738 negative image patches (i.e., IDC (-)) and 78,786 positive image patches (i.e., IDC (+)) were used. Each image file in the folder has a particular naming convention. For example, 15515 idx5 x601 y1351 class0 along with interpreting the naming structure of file; Case ID: 15515 idx5, x-axis of image: 601, y-axis of image: 1351and Class: 0 (0 means no cancer while 1 means cancerous).

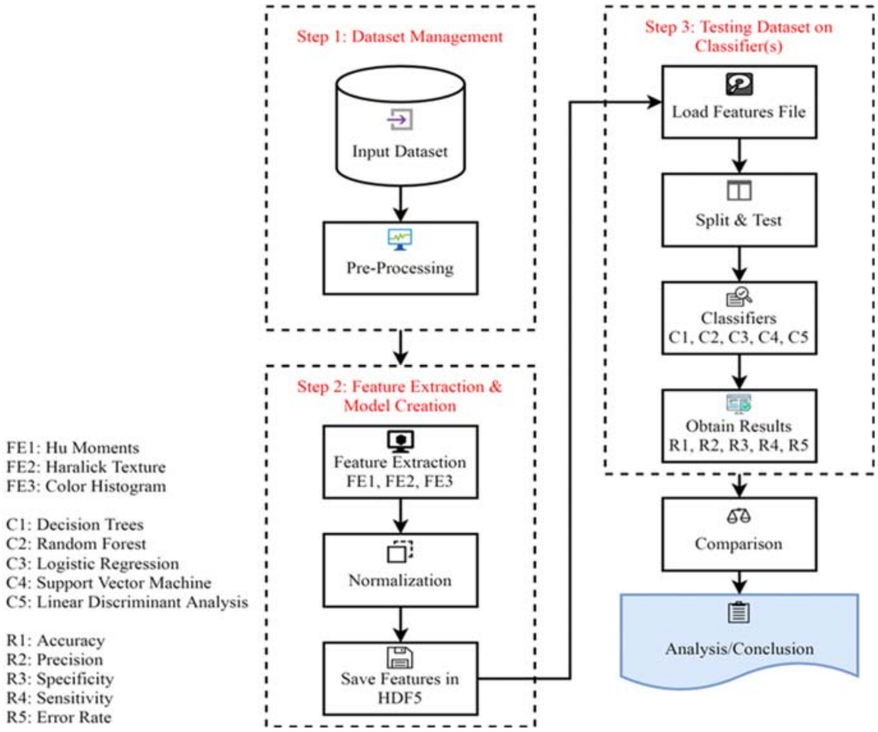

Figure 1: Proposed Approach shows all major steps.

3.7.2 Tools

The whole research study is done on MS Windows 7. This is one of the most successful variants of the Windows operating system line. Google Colab was used for computing. It is a very popular cloud service for machine learning which gives free access to GPU and TPU computing. Deepnote was used for python programming. It is free data science notebook, Jupyter-compatible, and easy deployment. A desktop PC with system specifications; Operating System: Microsoft Windows 7, CPU Type: Intel i5-2400 @ 3.10 GHz, GPU: NVIDIA GeForce GTX 950, Memory: 8GB DDR3 1333MHz and Disk Space: 500GB SSD and 3TB Hard Drive was used to conduct the experiment.

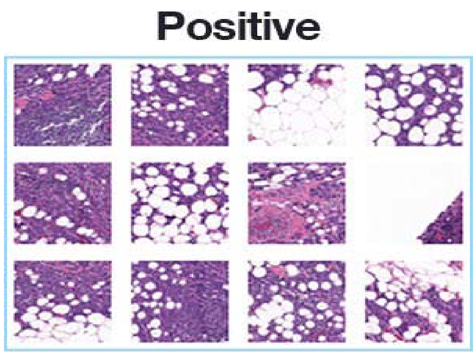

Figure 2: Example of positive image samples

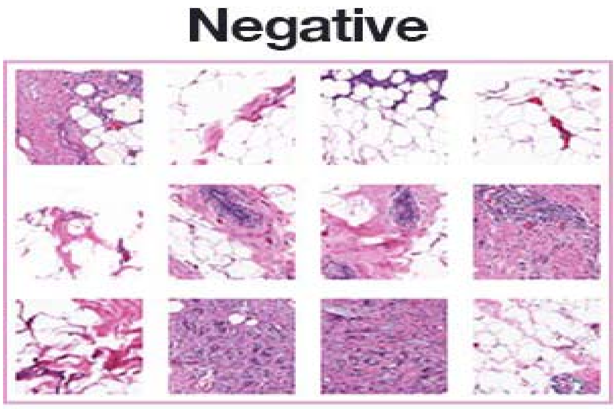

Figure 3: Example of negative image samples

4. Results and discussions

Results have been obtained by applying various algorithms including Convolutional Neural Networks, Decision Tree, Logistic Regression, Random Forest, Linear Discriminant Analysis, and Support Vector Machine on the breast cancer data. In this section details regarding obtained results are presented. The results are compared on the basis of parameters i.e., accuracy, precision, specificity, sensitivity, error rate, and confusion metric are also computed. In the following tables row 0, column 0 denotes the number of True Negative images, and row 1, column 1 denotes the number of True Positive images.

Table 1: Confusion Matrix for Keras

Confusion Matrix for CNN (Keras)	**0**	**1**
0	34153	5583
1	2811	12958

The value 34153 represents the True Negative images which mean CNN algorithm detects as Negative correctly, 5583 value represent the False Positive which means it detects as Positive but it actually is not, 2811 represent the False Negative which means it detects as Negative but it actually is not and 12958 represent True Positive which means they are detected Positive correctly as shown in Tab. 1. This Model's Accuracy is 84.87%. Precision is 82.17%, Specificity is 85.94%, Sensitivity is 82.17% and Error Rate is 15.12%.

Table 2: Confusion Matrix for Decision Trees

Confusion Matrix for *Decision Tree*	**0**	**1**
0	35152	4698
1	4458	11197

The value 35152 represents the True Negative images which mean Decision Tree algorithm detects as Negative correctly, 4698 value represent the False Positive which means it detects as Positive but it actually is not, 4458 represent the False Negative which means it detects as Negative but it actually is not and 11197 represent True Positive which means they are detected Positive correctly as shown in Tab. 2. This Model Accuracy is 83.50%. Precision is 70.44%, Specificity is 71.52%, Sensitivity is 88.21% and Error Rate is 16.49%.

Table 3: Confusion Matrix for Random Forest

Confusion Matrix for *Random Forest*	**0**	**1**
0	37571	2279
1	3936	11719

The value 37571 represents the True Negative images which mean Random Forest algorithm detects as Negative correctly, 2297 value represent the False Positive which means it detects as Positive but it actually is not, 3936 represent the False Negative which means it detects as Negative but it actually is not and 11719 represent True Positive which means they are detected Positive correctly as shown in Tab. 3. This Model Accuracy is 88.80%. Precision is 83.71%, Specificity is 74.85%, Sensitivity is 94.28% and Error Rate is 11.19%.

Table 4: Confusion Matrix for Logistic Regression

Confusion Matrix for *Logistic Regression*	**0**	**1**
0	36907	2943
1	4904	10751

The value 36907 represents the True Negative images which mean Logistic Regression algorithm detects as Negative correctly, 2943 value represent the False Positive which means it detects as Positive but it actually is not, 4904 represent the False Negative which means it detects as Negative but it actually is not and 10751 represent True Positive which means they are detected Positive correctly as shown in Tab. 4. This Model Accuracy is 85.86%. Precision is 78.50%, Specificity is 68.67%, Sensitivity is 92.61% and Error Rate is 14.13%.

Table 5: Confusion Matrix for Support Vector Machine

Confusion Matrix for *Support Vector Machine*	**0**	**1**
0	37156	2694
1	4470	11185

The value 37156 represents the True Negative images which mean Logistic Regression algorithm detects as Negative correctly, 2694 value represent the False Positive which means it detects as Positive but it actually is not, 4470 represent the False Negative which means it detects as Negative but it actually is not and 11185 represent True Positive which means they are detected Positive correctly as shown in Tab. 5. This Model Accuracy is 87.09%. Precision is 80.58%, Specificity is 71.44%, Sensitivity is 93.23% and Error Rate is 12.90%.

Table 6: Confusion Matrix for Linear Discriminant Analysis

Confusion Matrix for *Linear Discriminant Analysis*	**0**	**1**
0	36842	3008
1	4987	10668

The value 36842 represents the True Negative images which mean Logistic Regression algorithm detects as Negative correctly, 3008 value represent the False Positive which means it detects as Positive but it actually is not, 4987 represent the False Negative which means it detects as Negative but it actually is not and 10668 represent True Positive which means they are detected Positive correctly as shown in Tab. 6. This Model Accuracy is 85.59%. Precision is 78.00%, Specificity is 68.14%, Sensitivity is 92.45% and Error Rate is 14.40%.

4.1 Comparison for Accuracy, Precision, Specificity, Sensitivity, and Error Rate

It is evident from Tab. 7 and Fig. 4 that the Random Forest scores are significantly better than other classifiers on the same data set. Its Accuracy is 88.80%, Precision is 83.71% and Recall is 94.28% whereas Table 8 clearly depicts that with respect to time Logistic Regression executes quite fast (1 hour, 48 minutes) than all other algorithms.

Figure 4: Graphical representation obtained results of all algorithms

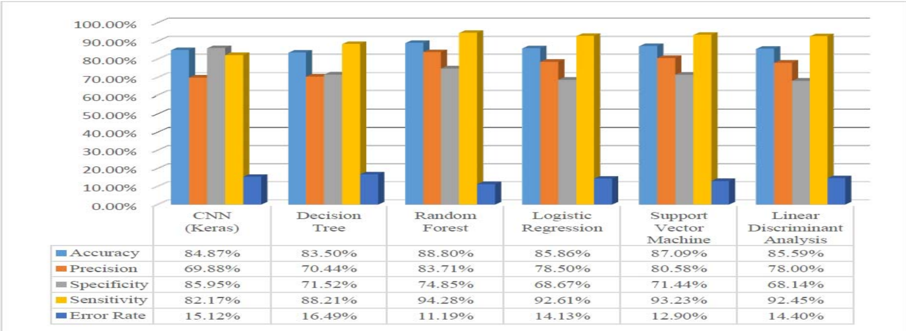

Table 7: Classification Comparison

Technique	Accuracy	Precision	Specificity	Sensitivity	Error Rate
CNN (Keras)	84.87 %	69.88 %	85.95 %	82.17 %	15.12 %
Decision Tree	83.50 %	70.44 %	71.52 %	88.21 %	16.49 %
Random Forest	88.80 %	83.71 %	74.85 %	94.28 %	11.19 %
Logistic Regression	85.86 %	78.50 %	68.67 %	92.61 %	14.13 %
Support Vector Machine	87.09 %	80.58 %	71.44 %	93.23 %	12.90 %
Linear Discriminant Analysis	85.59 %	78.00 %	68.14 %	92.45 %	14.40

4.2 Comparison by Execution Time

Another comparison is performed in the research that is a comparison of algorithms in terms of execution time. Tab. 8 shows the results of execution time taken by each algorithm. The results show that the execution time of the Logistic Regression algorithm is lesser than all other algorithms on the same data set.

Table 8: Comparison by Execution Time

Technique	**Execution Time**
CNN (Keras)	1 day, 7 hours, 12 minutes
Decision Tree	1 hour, 52 minutes
Random Forest	2 hours, 1 minute
Logistic Regression	1 hour, 48 minutes
Support Vector Machine	4 hours, 33 minutes
Linear Discriminant Analysis	2 hours, 6 minutes

5. Conclusion

The data set that has been used in this study has a variety of shapes of tumor which means the image samples in the data set are non-linear as shown in Fig. 2 and Fig. 3. Although Support Vector Machine (Accuracy: 87.09%, Precision: 80.58%, Recall: 93.23%) and Logistic Regression (Accuracy: 85.86%, Precision: 78.50%, Recall: 92.61%) are used for non-linear data sets but in this case both classifiers are not performing well. While if we compare Decision Trees (Accuracy: 83.50%, Precision: 70.44%, Recall: 88.21%) and Random Forest (Accuracy: 88.80%, Precision: 83.71%, Recall: 94.28%) we came to the decision that Random Forest is outperforming all the other algorithms used in this research study. This comparative study of state-of-the-art classifiers (Keras, Decision Trees, Logistic Regression, Random Forest, Linear Discriminant Analysis, and Support Vector Machine) is presented by performing experiments using feature extraction (Haralick Texture, Hu Moments, Color Histogram). The RF scores are significantly better than other classifiers on the same dataset. Its Accuracy is 88.80%, Precision is 83.71%, and Recall is

94.28%. Whereas with respect to time Logistic Regression executes quite fast (1 hour, 48 minutes). But we compromise on time to get more accurate results and to save a life. By considering all of the above facts it has been concluded that Random Forest is best among all (Keras, Decision Trees, Logistic Regression, Linear Discriminant Analysis, and Support Vector Machine) suitable for Breast Cancer disease prediction. In fact, this study indirectly might help in increasing the survival rate and save from dying because of this disease. It is believed that this research will be helpful for pathological analysis.

This research could be extended further as Mammographic, Ultrasound, and other types of biopsies and screening are used to predict cancer. To save a life and prevent unnecessary surgeries. So, it could make a huge contribution to a healthcare system if an application would be developed based on the outcome of this research study on disease prediction. Through this application using computer vision and machine learning human effort can be saved and human error can be reduced which is a critical factor in any healthcare system.

References

[1] P. Boyle and B. Levin, World cancer report 2008. IARC Press, International Agency for Research on Cancer, 2008.
[2] N. R. Council, "Saving women's lives: strategies for improving breast cancer detection and diagnosis," 2005.
[3] B. Stenkvist et al., "Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations," vol. 38, no. 12, pp. 4688-4697, 1978.
[4] R. Rubin, D. S. Strayer, and E. Rubin, Rubin's Pathology: Clinicopathologic Foundations of Medicine. Lippincott Williams & Wilkins, 2011.
[5] P. H. Tan et al., "The 2019 World Health Organization classification of tumours of the breast," 2020.
[6] P. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. J. E. S. w. A. Honeine, "Multiple instance learning for histopathological breast cancer image classification," vol. 117, pp. 103-111, 2019.
[7] C. Désir, C. Petitjean, L. Heutte, M. Salaun, and L. J. I. t. o. b. e. Thiberville, "Classification of endomicroscopic images of the lung based on random subwindows and extra-trees," vol. 59, no. 9, pp. 2677-2683, 2012.
[8] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. J. C. a. c. j. f. c. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," vol. 68, no. 6, pp. 394-424, 2018.
[9] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.
[10] C. M. J. R. S. Bishop, "Pattern recognition and machine learning-springer 2006," 2006.
[11] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in Proceedings of ICML workshop on unsupervised and transfer learning, 2012, pp. 17-36: JMLR Workshop and Conference Proceedings.
[12] G. R. Bowman, V. S. Pande, and F. Noé, An introduction to Markov state models and their application to long timescale molecular simulation. Springer Science & Business Media, 2013.
[13] L. N. Long, A. J. J. o. A. C. Gupta, Information,, and Communication, "Scalable massively parallel artificial neural networks," vol. 5, no. 1, pp. 3-15, 2008.
[14] K. J. N. n. Hornik, "Approximation capabilities of multilayer feedforward networks," vol. 4, no. 2, pp. 251-257, 1991.
[15] D. Steinkraus, I. Buck, and P. Simard, "Using GPUs for machine learning algorithms," in Eighth International Conference on Document Analysis and Recognition (ICDAR'05), 2005, pp. 1115-1120: IEEE.
[16] R. J. P. Girshick, NJ: IEEE., "Fast r-cnn In Proceedings of the IEEE International Conference on Computer Vision (pp. 1440–1448)," 2015.
[17] R. Szeliski, Computer vision: algorithms and applications. Springer Science & Business Media, 2010.
[18] N. Sebe, I. Cohen, A. Garg, and T. S. Huang, Machine learning in computer vision. Springer Science & Business Media, 2005.
[19] M. Lürig, S. Donoughe, E. I. Svensson, A. Porto, and M. Tsuboi, "Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology," 2020.
[20] T. Chen et al., "A decision tree-initialised neuro-fuzzy approach for clinical decision support," vol. 111, p. 101986, 2021.
[21] A. Schidler and S. Szeider, "SAT-based Decision Tree Learning for Large Data Sets," in Proceedings of AAAI, 2021, vol. 21.
[22] D. Grattarola and C. J. I. C. I. M. Alippi, "Graph neural networks in tensorflow and keras with spektral application notes.," vol. 16, no. 1, pp. 99-106, 2021.
[23] M. Ashraf et al., "Prediction of Cardiovascular Disease Through Cutting-Edge Deep Learning Technologies: An Empirical Study Based on TENSORFLOW, PYTORCH and KERAS," in International Conference on Innovative Computing and Communications, 2021, pp. 239-255: Springer.
[24] N. M. Abdulkareem, A. M. J. I. J. o. S. Abdulazeez, and Business, "Machine learning classification based on Radom Forest Algorithm: A review," vol. 5, no. 2, pp. 128-142, 2021.
[25] J. Dhaliwal et al., "Accurate Classification of Pediatric Colonic Inflammatory Bowel Disease Subtype Using a Random Forest Machine Learning Classifier," vol. 72, no. 2, pp. 262-269, 2021.
[26] R. Khera et al., "Use of machine learning models to predict death after acute myocardial infarction," 2021.
[27] I. Ibrahim, A. J. J. o. A. S. Abdulazeez, and T. Trends, "The role of machine learning algorithms for diagnosing diseases," vol. 2, no. 01, pp. 10-19, 2021.
[28] K. Dhasaradhan, R. Jaichandran, K. Shunmuganathan, S. U. Kiruthika, and S. Rajaprakash, "Hybrid Machine Learning Model Using Decision Tree and Support Vector Machine for Diabetes Identification," in Data Engineering and Intelligent Computing: Springer, 2021, pp. 293-305.
[29] M. M. Abdelsalam and M. J. I. A. Zahran, "A Novel Approach of Diabetic Retinopathy Early Detection Based on Multifractal Geometry Analysis for OCTA Macular Images Using Support Vector Machine," vol. 9, pp. 22844-22858, 2021.
[30] Z. Rustam, Y. Amalia, S. Hartini, and G. S. J. I. I. J. o. A. I. Saragih, "Linear discriminant analysis and support vector machines for classifying breast cancer," vol. 10, no. 1, p. 253, 2021.
[31] H. Li et al., "Raman Microspectral Study and Classification of the Pathological Evolution of Breast Cancer Using Both Principal Component Analysis-Linear Discriminant Analysis and Principal Component Analysis-Support Vector Machine," vol. 2021, 2021.
[32] R. M. Haralick, K. Shanmugam, I. H. J. I. T. o. s. Dinstein, man,, and cybernetics, "Textural features for image classification," no. 6, pp. 610-621, 1973.
[33] A. Rosebrock, "Breast cancer classification with keras and deep learning," ed, 2019.
[34] B. Hamner. (2010). Breast Histopathology Images Online. Available: https://www.kaggle.com/paultimothymooney//breast-histopathology-images.