International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 04, 2022
www.ijirse.com
ISSN: 2454-9665

Breast Cancer Prediction using SVM

Bhagyashree Dhawale[1], Soniya Ranaware[2], Divya Macharla[3], Dr.Nikita Kulkarni[4]

[1,2,3,4]*Student,Computer Engineering, Trinity Academy of Engineering, Pune, India*

ABSTRACT

Breast malignancy (cancer) is one of the most widely recognized malignancies among ladies around the world, addressing the larger part of new malignancy cases and cancer deaths around the world. It is additionally the subsequent cause of malignancy death among ladies keep on experiencing it. The early analysis of the sickness can work on the shot at endurance altogether as it can assist with convenient clinical treatment to patients. The utilization of measurable and AI calculations can be valuable for the prediction of breast cancer.The techniques used in proposed system for detection of breast cancer is present or not we are using Support Vector Machine (SVM) techniques. In Healthcare sectors machine learning has been proven important for early detection of any disease.

Keywords: - Cancer, Image processing, SVM, Breast

• INTRODUCTION

Huge number of ladies all throughout the planet succumb to malignant growth per annum. The actual body contains numerous cells each with its own remarkable capacity. Around 42000 ladies due to malignancy yearly, that is 1 lady every 13 minutes is dying from this disease a day. Malignant growth is normally brought about by a hereditary illness. In any case, just 5-10% of tumors are acquired from guardians. All things considered, 85-90% of breast cancer diseases are because of hereditary anomalies that occur because of the maturing system and along these lines the "mileage" of life for the most part. Growths could likewise be cancerous (harmful) or not cancerous (harmless).

Bosom(Breast) malignant growth is second most normal kind of disease in ladies, as per reports 14% ladies are enrolled because of Breast Cancer in India. In India analyze a bosom disease patient in at regular intervals in both rustic and metropolitan space of India. According to the reports of Breast malignancy 1, 62,468 are enrolled cases in India out of those 87,090 patients' demises. According to the reports 14.5 lakh cases are normal and it will increment to 17.3 lakh patients in 2020. AI calculation has assumed a significant part in the field of clinical, medical services areas to foresee illness with higher precision on huge measure of datasets, huge measure of information can be produces step by step in medical care areas. In the terms of patients a few illnesses reports and enormous measure of data set anyway medical services areas, clinical field these datasets can be generally circulated and utilized AI calculation is a course of that principally predicts the patient's dieses and report with higher precision The review of World Health Organization (WHO) reports that the bosom malignant growth is the most widely recognized disease among ladies. Around 5% of Indian ladies are have

ISSN: 2454-9665

hazard of bosom malignancy while Europe and in the U S, it is around 12.5%. For the most part, bosom disease can be handily recognized if explicit manifestations emerges. Nonetheless, a few ladies who are experiencing bosom disease have no indications. Thus, bosom disease acknowledgment at beginning phase is vital. Early discovery of bosom disease helps in early finding and treatment, on the grounds that the forecast is vital for long haul endurance. Early discovery, conclusion, and therapy of bosom malignant growth can save an existence of a patient.

- **RELATED WORK**

MuhammetFatihAket.al [1] used the dataset from Dr. William H. Walberg of the University of Wisconsin Hospital. Data visualization and machine learning techniques including logistic regression, k-nearest neighbors, support vector machine, naïve Bayes, decision tree, random forest, and rotation forest were applied to this dataset. R, Minitab, and Python were chosen to be applied to these machine learning techniques and visualization.. Results obtained with the logistic regression model with all features included showed the highest classification accuracy (98.1%), and the proposed approach revealed the enhancement in accuracy performances.

In [2] presented a novel method to detect breast cancer by employing techniques of Machine Learning such as Naïve Bayes classifier, SVM classifier, Bi-clustering Ada Boost techniques, RCNN classifier and Bidirectional Recurrent Neural Networks (HA-BiRNN). A comparative analysis was done between the Machine learning techniques and the proposed methodology (Deep Neural Network with Support Value) and the simulated results concluded that the DNN algorithm was advantageous in both performance, efficiency and quality of images are crucial in the latest medical systems whilst the other techniques didn't perform as expected

In [3] Proposed instinctive classification of mammogram images as Benign, Malignant and Normal using various machine learning algorithms. A comparative analysis is performed between Support Vector Machines, Convolutional Neural Network and Random Forest. The simulation results concluded that CNN is the best classifier as it results in instinctive classification of digital mammograms using filtering and morphological operations.

In[4] A comparative study on ANN and SVM and integrated various classifiers like CNN, KNN and Inception V3 for better processing of the dataset. The experimental results and performance analysis concluded that ANN was a better classifier than SVM as ANN proved to have a higher efficiency rate

In [5] performed a comparative analysis between SVM, Logistic Regression, Naïve Bayes and Random Forest. The Wisconsin Breast cancer dataset is used to perform the comparison. Based on the result of performed experiments, the Random Forest algorithm showed the highest accuracy (99.76%) with the least error rate. ANACONDA Data Science Platform was used to execute all the experiments in a simulated environment

- **MOTIVATION**

Breast Cancer is one of the leading cancer 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

ISSN: 2454-9665

early recognition. Hence, apart from medicinal solutions some Data Science solution needs to be integrated for resolving the death causing issue.

Breast cancer is the second most frequent cancer in women and men globally. In 2012, it factored about 12 percent of all latest cancer cases and 25 percent of women's total cancers.Breast cancer arises when cells in the breast start to develop out of control. These cells usually grow a tumor that can frequently be seen on an x-ray or considered a lump. The tumor is malignant (cancer) if the cells can expand into (invade) encompassing tissues or increase (metastasize) to different sections of the body.

- **SYSTEM ARCHITECTURE**

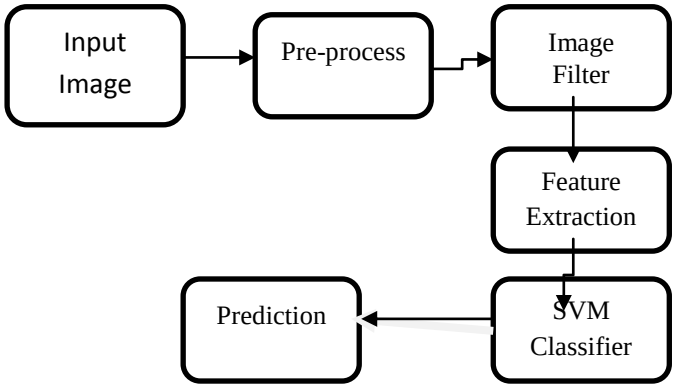


Fig: - System Architecture

- **METHODOLOGY**

The data used for the experiments was acquired from Kaggle. The dataset consists of around 2000-3000 which are divided into 80% for training and 20% for testing. Each magnification directory consists of two directories representing the tumors i.e. Benign and Malignant.Preprocess: - In this step we are processing the dataset where are removing the unwanted data and cleaning the dataset according .Later these dataset are further divided into 80% and 20% respectively. 80% is trained and remaining 20% is been used for testing the model.Feature Selection is the importance of feature selection in a machine learning model is inevitable. It turns the data to be free from ambiguity and reduces the complexity of the data. Also, it reduces the size of the data, so it is easy to train the model and reduces the training time. It avoids over fitting of data. Selecting the best feature subset from all the features increases the accuracy. Classification is this step the input given by user is been compared with the trained data set which has been trained using Support Vector Machine Techniques (SVM) Depending on that the prediction is be done by the machine.

Algorithm used: - Support Vector Machine:-

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. The objective is to find a plane that has the maximum

International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 04, 2022
www.ijirse.com
ISSN: 2454-9665

margin, i.e the maximum distance between data points of both classes. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. SVM clearly is the most effective classifier of all as it worksreally well with clear margin of separation and high dimensional data, but is not suitable for large data sets because the required training time is higher and also, underperforms when the data set has more noise

• **ACKNOWLEDGEMENT**

We wish to thank to Dr,Nikita Kulkarni Associate Professor at KJEI Trinity Acadamy of Engineering, Pune, Maharashtra, India for the constant support and encouragement in our work.

• **RESULT AND DISCUSSION**

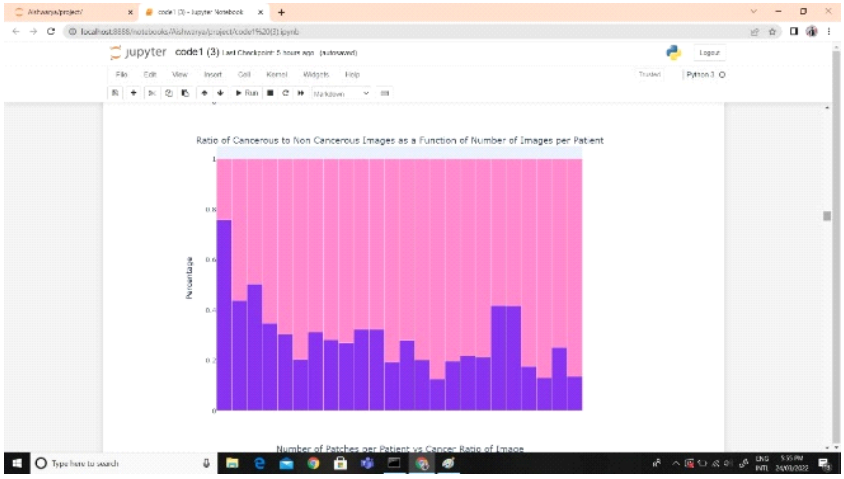


Fig a:-

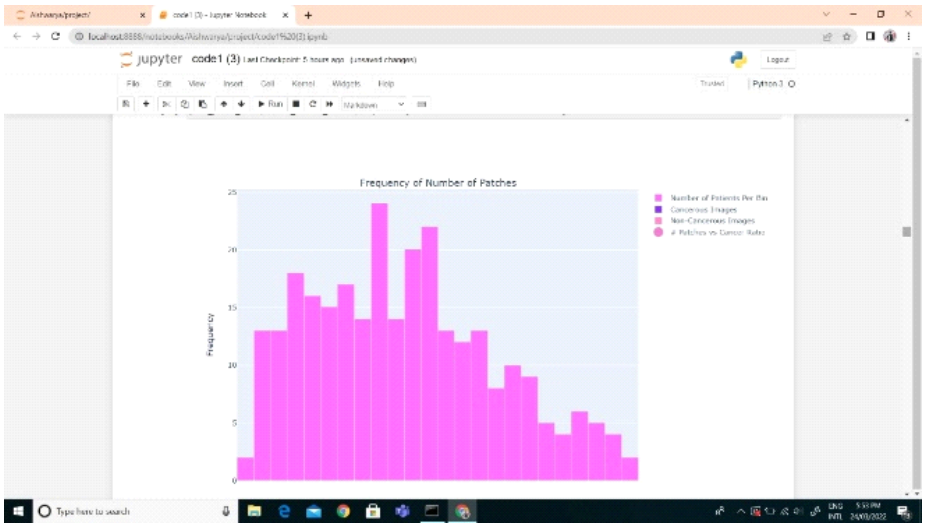


Fig b :-

International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 04, 2022
www.ijirse.com
ISSN: 2454-9665

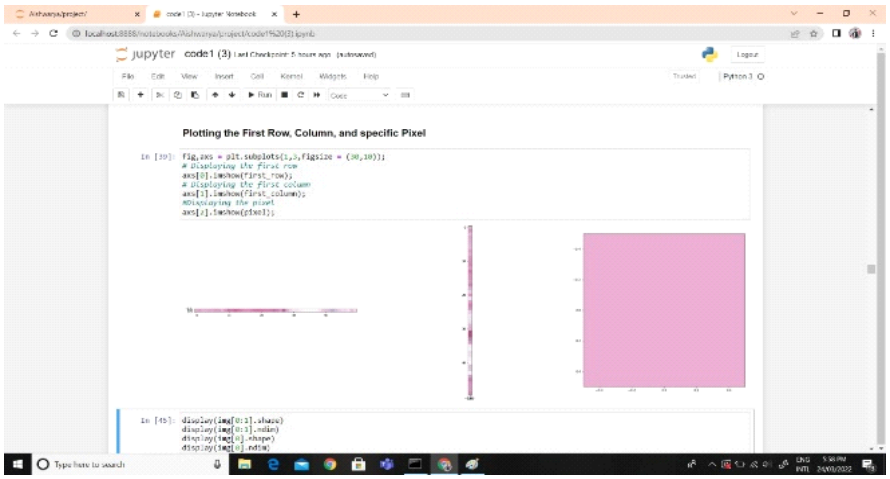


Fig c:-

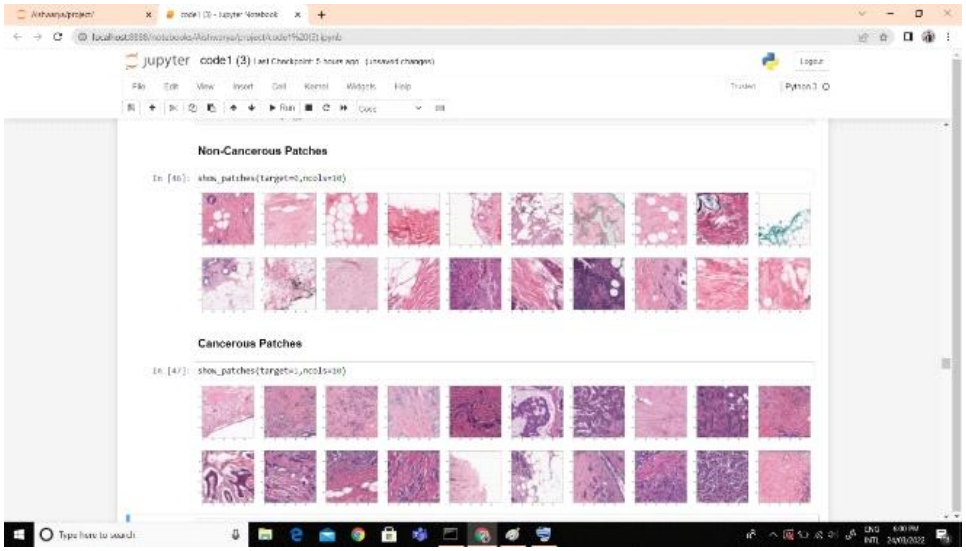


Fig d:-

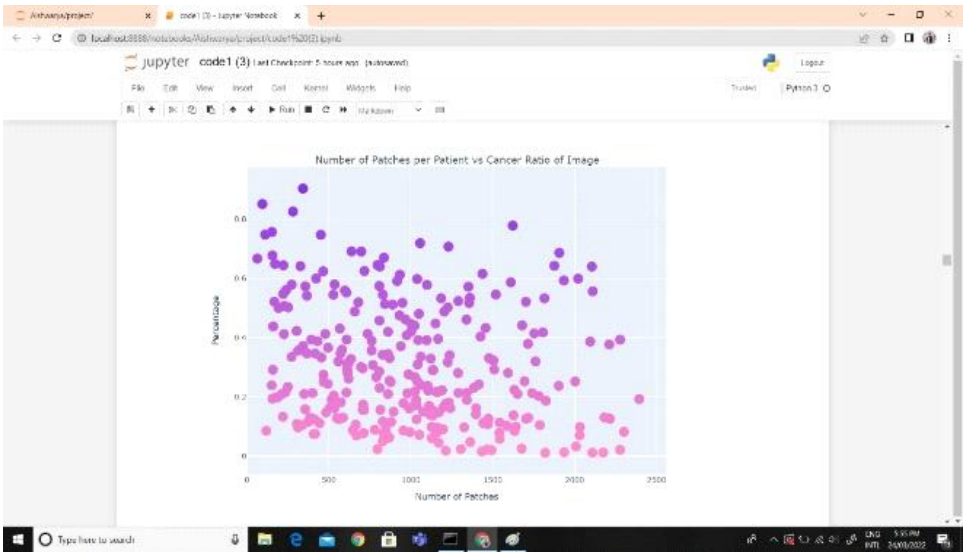


Fig e:-

International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 04, 2022
www.ijirse.com
ISSN: 2454-9665

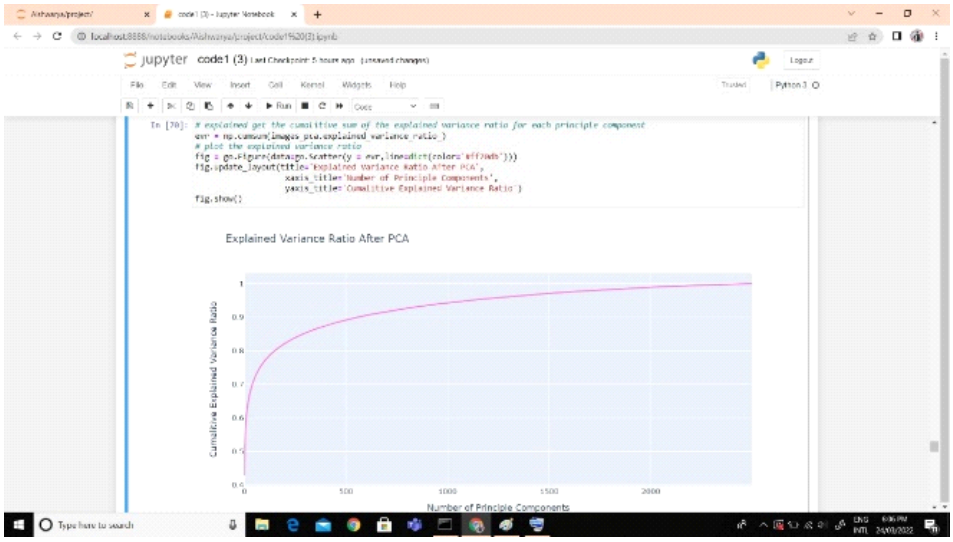


Fig f:-

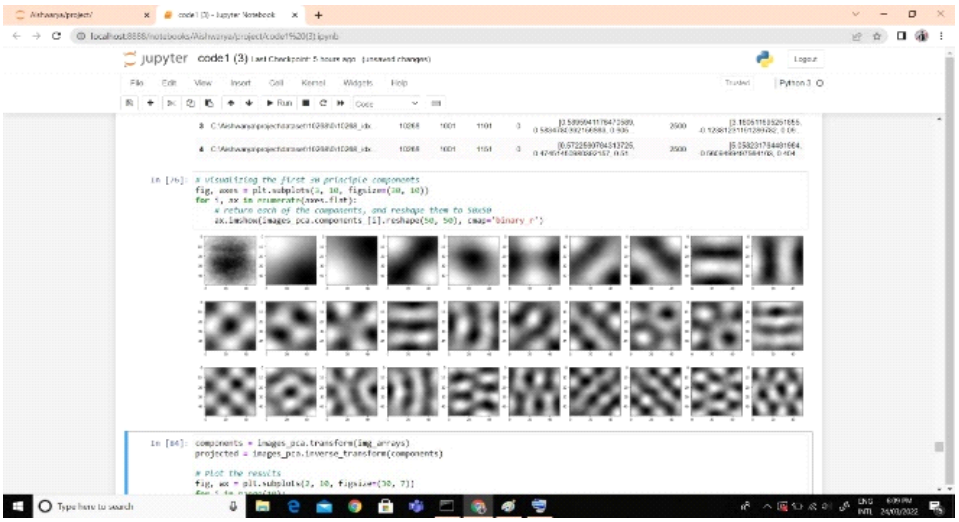
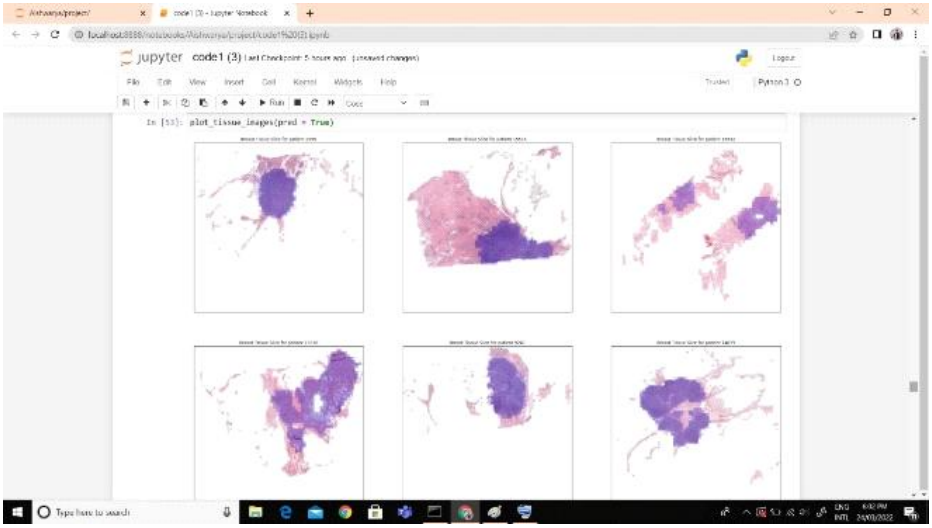


Fig g:-



International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 04, 2022
www.ijirse.com
ISSN: 2454-9665

Fig h:-

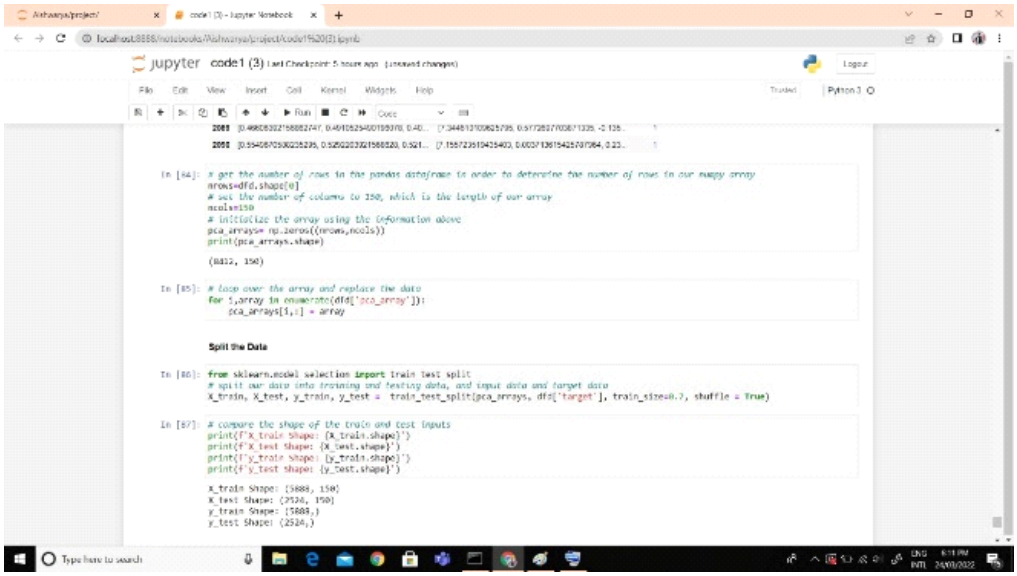


Fig I :-

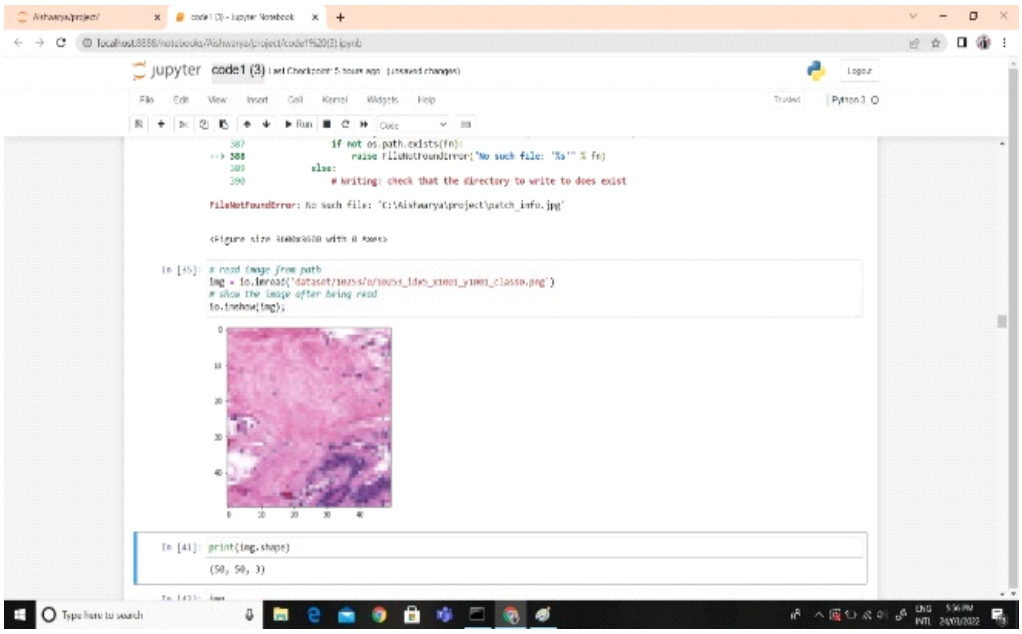


Fig j:-

- **CONCLUSION**

Breast cancer if found at an early stage will help save lives of thousands of women or even men. These projects help the real world patients and doctors to gather as much information as they can. By using machine learning algorithms we will be able to classify and predict the cancer into being or malignant. Machine learning algorithms can be used for medical oriented research, it advances the system, reduces human errors and lowers manual mistakes

ISSN: 2454-9665

REFERENCES

[1] MuhammetFatihAk, "A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications", 2020

[2] Anji Reddy Vaka, BadalSoni and SudheerReddy K, "Breast Cancer Detection by Leveraging Machine Learning" , 2020

[3] S.Vasundhara, B.V. Kiranmayee and Chalumuru Suresh, "Machine Learning Approach for Breast Cancer Prediction", 2019

[4] KalyaniWadkar, Prashant Pathak and Nikhil Wagh, "Breast Cancer Detection Using ANN Network and Performance Analysis with SVM", 2019

[5] Sivapriya J, Aravind Kumar V, Siddarth Sai S and Sriram S , "Breast Cancer Prediction using Machine Learning",2019

[6] Sultana, Jabeen, Abdul KhaderJilani, &"Predicting Breast Cancer Using Logistic Regression and Multi-Class Classifiers." International Journal of Engineering & Technology

[7] Gupta, P., and P. S.. "Analysis of Machine Learning Techniques for Breast Cancer Prediction". InternationalJournal of Engineering and Computer Science, Vol. 7, no.05, May 2018

[8] Syakur, M. A., et al. "Integration k-means clustering method and elbow method for identification of the best customer profile cluster." IOP Conference Series: Materials Science and Engineering. Vol. 336. No. 1. IOP Publishing, 2018.

[9] Chaurasia, Vikas, Saurabh Pal, and B. B. Tiwari. "Prediction of benign and malignant breast cancer using data mining techniques." Journal of Algorithms & Computational Technology, 2018

[10] Puneet Yadav et al. "Diagnosis of Breast Cancer using Decision Tree Models and SVM", International Research Journal of Engineering and Technology, Vol. 5, Issue 3, Mar 2018