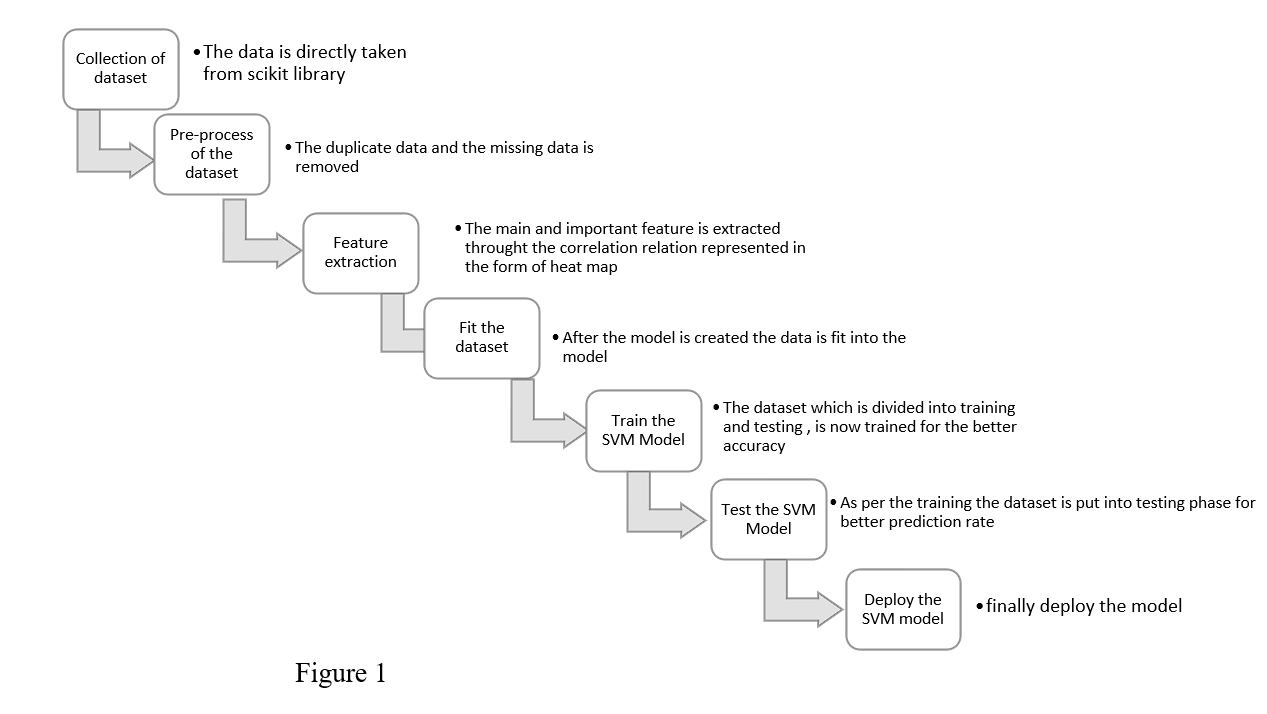
***Early stage of breast cancer detection using machine learning algorithms***

|  |  |
| --- | --- |
| Tarun Pandey  School of Computer Science and Engineering  Vellore Institute of Technology, Chennai  Tarun.pandey2022@vitstudent.ac.in | Prof. Abhirami S  School of Computer Science and Engineering  Vellore Institute of Technology, Chennai  abirami.s@vit.ac.in |

**Abstract:**

Since breast cancer is a common cancer in women, early detection is crucial to increase survival rates and reduce intrusive treatments. Support vector machines (SVMs) and logistic regression, in particular, have demonstrated promise in the early-stage identification of breast cancer, frequently in conjunction with thermal imaging. Prognostic Breast Cancer datasets study explores the application of SVM for early-stage breast cancer identification. The SVM system will be trained and tested using a collection of patient data, categorising patients as normal or abnormal (cancerous). Metrics like recall, accuracy, and precision will be taken into account during performance evaluation. This study also analyses the use of Logistic Regression for breast cancer screening and investigates feature extraction techniques, including seven essential features. To evaluate the relative advantages and disadvantages of SVM and Logistic Regression, a comparison between the two is made. The goal of the results analysis is to evaluate the performance, dependability, and non-invasiveness of SVM and Logistic Regression in the early-stage identification of breast cancer. The voting classifier is also used to get a combined accuracy of SVM and Logistic regression. At last, the comparison of all the algorithm is shown.

**Keywords— breast cancer detection, machine learning, Support Vector Machine, Prognostic Breast Cancer datasets, Logistic Regression, Voting classifier, comparison graph, feature extraction, result analysis**



**4.1 SVM ALGORITHM:**

SVM algorithms can be used to classify data based on a set of pre-defined features, and have been shown to be effective in detecting early-stage breast cancer in mammography images. Prognostic datasets can provide additional information that can help to improve the diagnostic accuracy of the SVM algorithm. By incorporating information about patient demographics, medical history, and imaging findings, it may be possible to identify early-stage breast cancer with greater accuracy.

**4.2 LOGISTIC REGRESSION ALGORITHM:**

For the purpose of early breast cancer detection, the proposed approach will make use of the logistic regression technique. The system will train and test the logistic regression model to categorise patients into normal or abnormal (cancerous) instances using a large dataset of breast cancer patient information. Accuracy, precision, and recall measurements will be used to assess the system's performance and reveal how well the algorithm detects breast cancer. A dependable, non-invasive, and accurate method for the early diagnosis of breast cancer is what the suggested system wants to achieve.

**4.3 VOTING CLASSIFIER:**

A final prediction is made by the voting classifier, a machine learning ensemble learning technique that integrates the predictions of various separate models. It is a type of model aggregation in which every model is individually trained on the same dataset, and then its predictions are added together to reach a judgement.

Both classification and regression tasks can be performed with the voting classifier

**COMPARISON BETWEEN SVM AND LOGISTIC REGRESSION:**

Comparison bar graph between SVM and Logistic regression, with 97% accuracy of SVM algorithm and 75% with logistic regression.

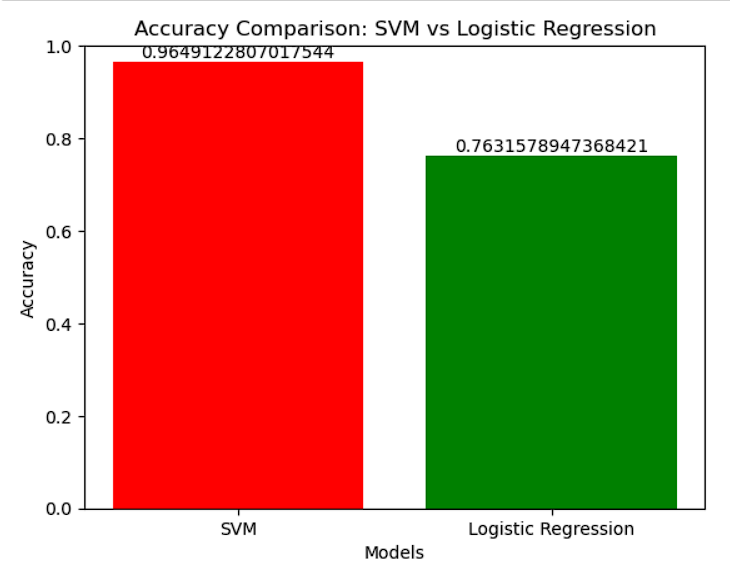


Figure 9.

**COMPARISON AMONG SVM,** **LOGISTIC REGRESSION AND VOTING CLASSIFIER:**

Comparison bar graph between SVM Logistic Regression and Voting Classifier, with 97% accuracy of SVM algorithm, 75% with logistic regression and 95% with voting classifier.

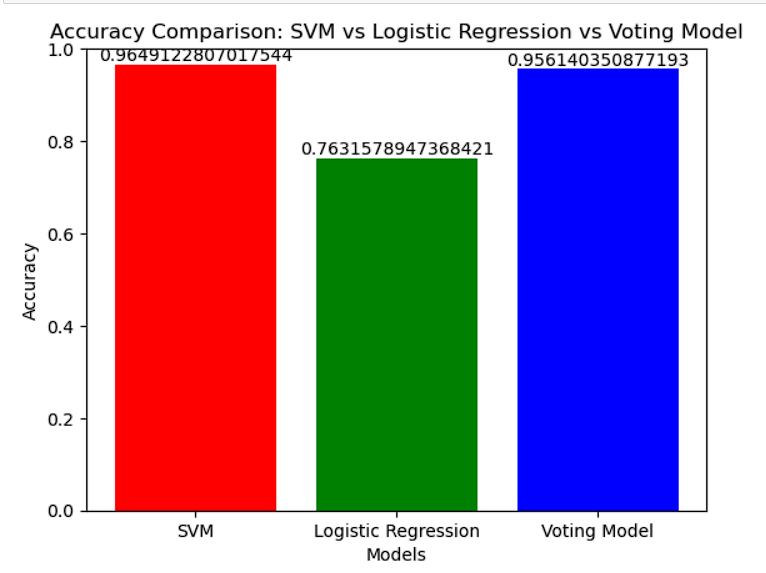


Figure 10.

All the algorithm with their accuracy:

|  |  |
| --- | --- |
| Algorithms | Accuracy |
| SVM | 96.49% |
| LR | 76.31% |
| Voting classifier  (SVM + LR) | 95.61% |