BREAST TUMOR CLASSIFICATION

Team Members:
Rahul Koli (31)
Sahil Lotya (33)
Pawan Lulla (34)

ABOUT THE DISEASE 01

02 LITERATURE SURVEY

DATASET USED 03

04 METHODOLOGY
& ALGORITHMS USED

RESULT OBTAINED 05

06 CONCLUSIONS

INTRODUCTION: ABOUT THE DISEASE

Cancer begins when healthy cells in the breast change and grow out of control, forming a mass or sheet of cells called a tumor.

Breast cancer spreads when the cancer grows into adjacent organs or other parts of the body or when breast cancer cells move to other parts of the body through the blood vessels and/or lymph vessels.

There are 4 stages to a Breast Cancer. Stage 0 being early-stage and stage 3 locally advanced invasive breast cancer.

Types of breast cancer:

- Ductal carcinoma
- Invasive lobular carcinoma

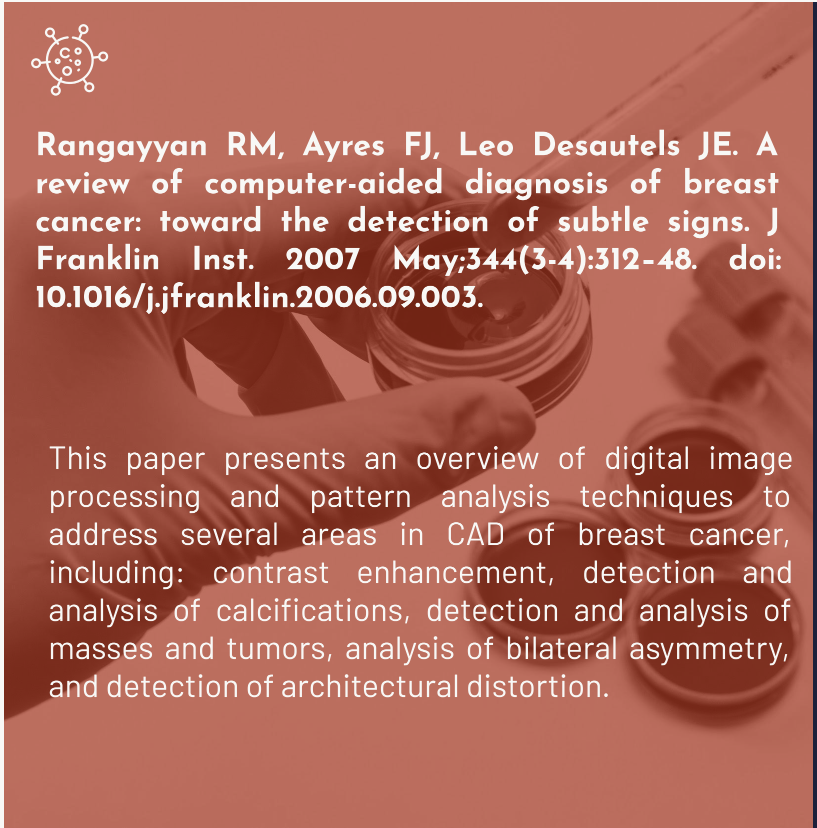
SYMPTOMS OF THE DISEASE

Lump in your breast or underarm

Pain and tenderness

Skin sores

LITERATURE SURVEY



Rangayyan RM, Ayres FJ, Leo Desautels JE. A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. J Franklin Inst. 2007 May;344(3-4):312-48. doi: 10.1016/j.jfranklin.2006.09.003.

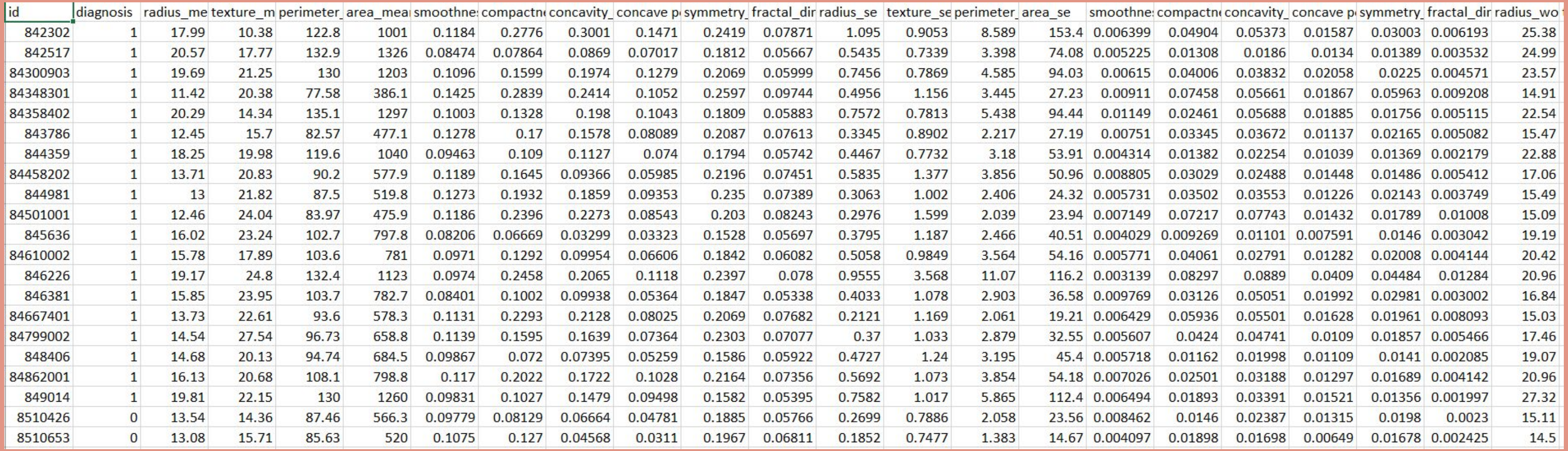
This paper presents an overview of digital image processing and pattern analysis techniques to address several areas in CAD of breast cancer, including: contrast enhancement, detection and analysis of calcifications, detection and analysis of masses and tumors, analysis of bilateral asymmetry, and detection of architectural distortion.

Mohammed S.A., Darrab S., Noaman S.A., Saake G. (2020) Analysis of Breast Cancer Detection Using Different Machine Learning Techniques. In: Tan Y., Shi Y., Tuba M. (eds) Data Mining and Big Data. DMBD 2020. Communications in Computer and Information Science, vol 1234. Springer, Singapore. https://doi.org/10.1007/978-981-15-7205-0_10

In this paper, we propose an approach that improves the accuracy and enhances the performance of three different classifiers: Decision Tree (J48), Naïve Bayes (NB), and Sequential Minimal Optimization (SMO). We also validate and compare the classifiers on two benchmark datasets: Wisconsin Breast Cancer (WBC) and Breast Cancer dataset.

DATASET USED

- The Dataset was taken from https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
- This dataset consists of 570 people with 10 different parameters. We have divided the dataset into 470 for training and the remaining 100 for testing.
- Snapshot of the Dataset:

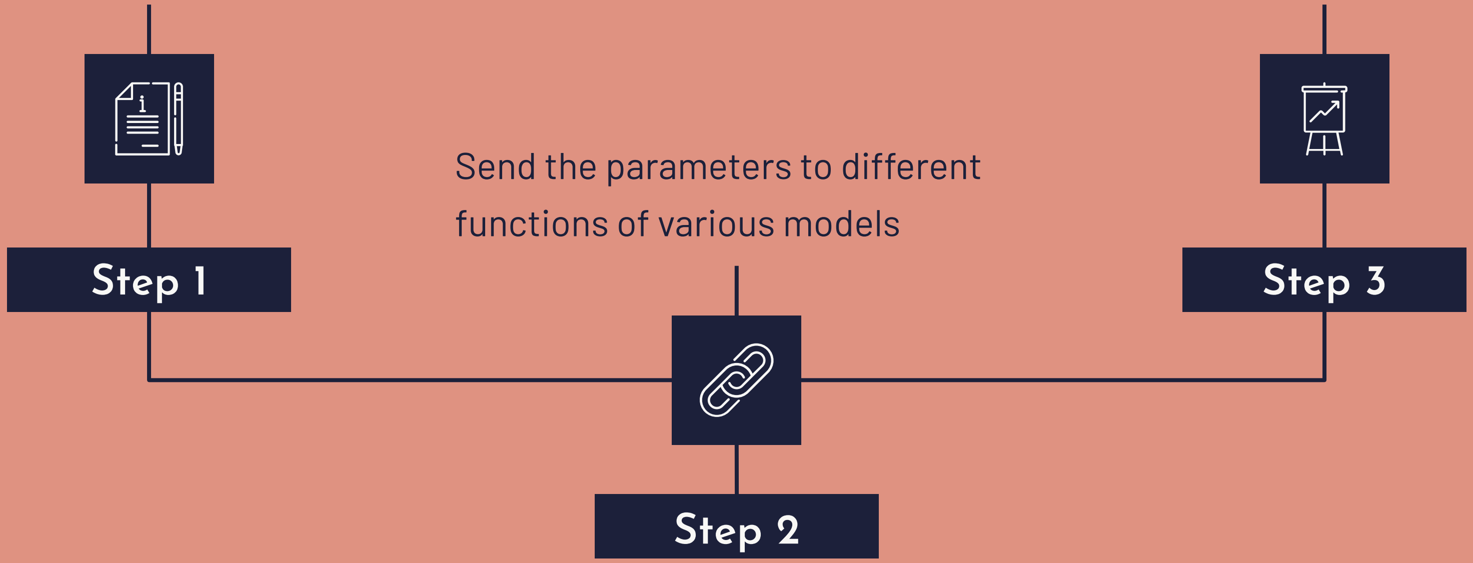


id	diagnosis	radius_me	texture_m	perimeter_	area_mea	smoothne:	compactn	concavity_	concave p	symmetry_	fractal_dir	radius_se	texture_se	perimeter_	area_se	smoothne:	compactn	concavity_	concave p	symmetry_	fractal_dir	radius_wo
842302	1	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871	1.095	0.9053	8.589	153.4	0.006399	0.04904	0.05373	0.01587	0.03003	0.006193	25.38
842517	1	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.005225	0.01308	0.0186	0.0134	0.01389	0.003532	24.99
84300903	1	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.7869	4.585	94.03	0.00615	0.04006	0.03832	0.02058	0.0225	0.004571	23.57
84348301	1	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.156	3.445	27.23	0.00911	0.07458	0.05661	0.01867	0.05963	0.009208	14.91
84358402	1	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.7813	5.438	94.44	0.01149	0.02461	0.05688	0.01885	0.01756	0.005115	22.54
843786	1	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.8902	2.217	27.19	0.00751	0.03345	0.03672	0.01137	0.02165	0.005082	15.47
844359	1	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.7732	3.18	53.91	0.004314	0.01382	0.02254	0.01039	0.01369	0.002179	22.88
84458202	1	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451	0.5835	1.377	3.856	50.96	0.008805	0.03029	0.02488	0.01448	0.01486	0.005412	17.06
844981	1	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.002	2.406	24.32	0.005731	0.03502	0.03553	0.01226	0.02143	0.003749	15.49
84501001	1	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243	0.2976	1.599	2.039	23.94	0.007149	0.07217	0.07743	0.01432	0.01789	0.01008	15.09
845636	1	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697	0.3795	1.187	2.466	40.51	0.004029	0.009269	0.01101	0.007591	0.0146	0.003042	19.19
84610002	1	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842	0.06082	0.5058	0.9849	3.564	54.16	0.005771	0.04061	0.02791	0.01282	0.02008	0.004144	20.42
846226	1	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078	0.9555	3.568	11.07	116.2	0.003139	0.08297	0.0889	0.0409	0.04484	0.01284	20.96
846381	1	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338	0.4033	1.078	2.903	36.58	0.009769	0.03126	0.05051	0.01992	0.02981	0.003002	16.84
84667401	1	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069	0.07682	0.2121	1.169	2.061	19.21	0.006429	0.05936	0.05501	0.01628	0.01961	0.008093	15.03
84799002	1	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077	0.37	1.033	2.879	32.55	0.005607	0.0424	0.04741	0.0109	0.01857	0.005466	17.46
848406	1	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586	0.05922	0.4727	1.24	3.195	45.4	0.005718	0.01162	0.01998	0.01109	0.0141	0.002085	19.07
84862001	1	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	0.5692	1.073	3.854	54.18	0.007026	0.02501	0.03188	0.01297	0.01689	0.004142	20.96
849014	1	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498	0.1582	0.05395	0.7582	1.017	5.865	112.4	0.006494	0.01893	0.03391	0.01521	0.01356	0.001997	27.32
8510426	0	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05766	0.2699	0.7886	2.058	23.56	0.008462	0.0146	0.02387	0.01315	0.0198	0.0023	15.11
8510653	0	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967	0.06811	0.1852	0.7477	1.383	14.67	0.004097	0.01898	0.01698	0.00649	0.01678	0.002425	14.5

METHODOLOGY

- In this problem statement, we need to detect the breast cancer using the parameters given like radius of tumor etc.
- For performing this, we have used various algorithms
    - KNN
    - Logistic regression
    - Linear and Polynomial Support Vector Machines
    - Decision Tree, etc.

METHODOLOGY



Step 1

Upload the CSV file
Separate training and testing data

Step 2

Send the parameters to different functions of various models

Step 3

Obtain the accuracy and plot the graph for the same

ALGORITHMS USED

KNN

Stores all available data and classifies a new data point based on similarity

Logistic Regression

Can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

SVM's

Can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

Decision Tree

Each internal node denotes a test, each branch represents an outcome of the test, and each leaf node holds a class label.

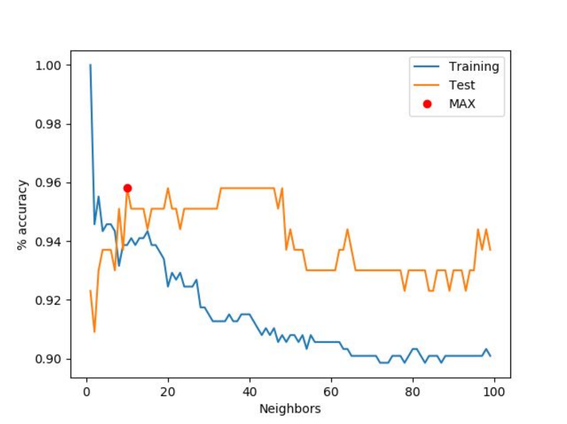
Gradient Boosting

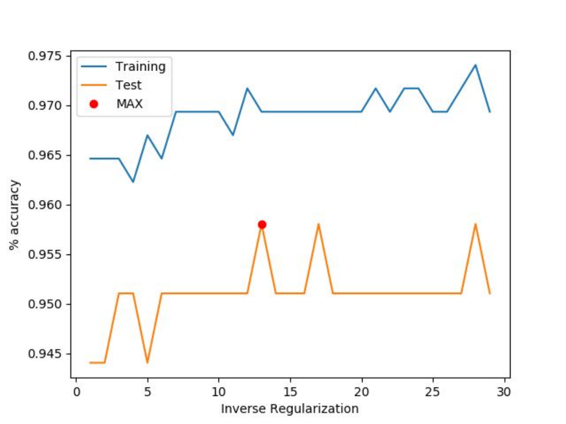
Ensemble of shallow trees in sequence with each tree learning and improving on the previous one.

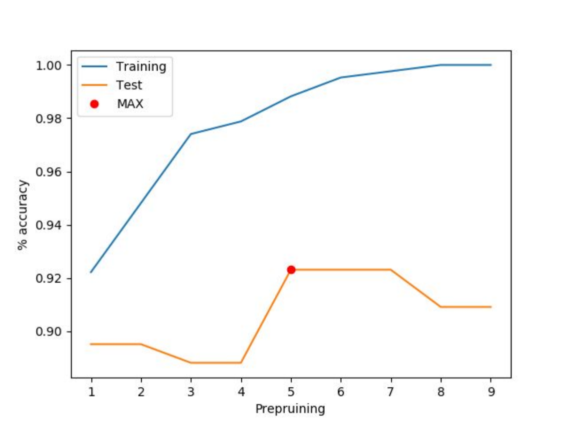
Naive-Bayes

Models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set.

RESULT ANALYSIS







KNN

Accuracy: 95.80% at 10 neighbours

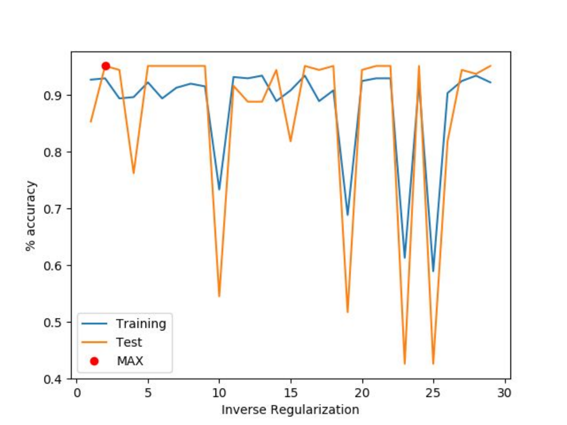
Logistic Regression

Accuracy: 95.80% at Inverse Regularization: 16

Decision Tree

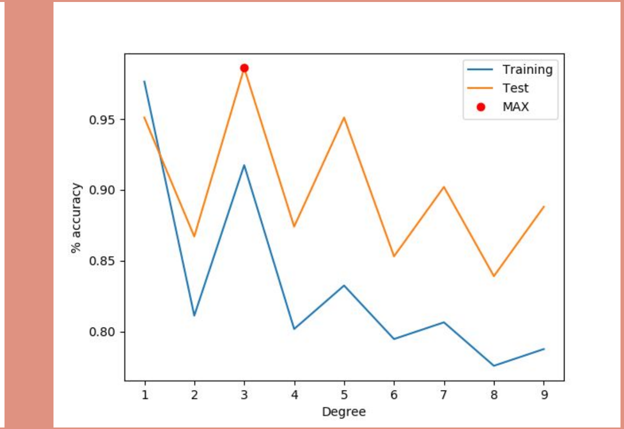
Accuracy: 92.30% At Pre-pruning = 5

RESULT ANALYSIS



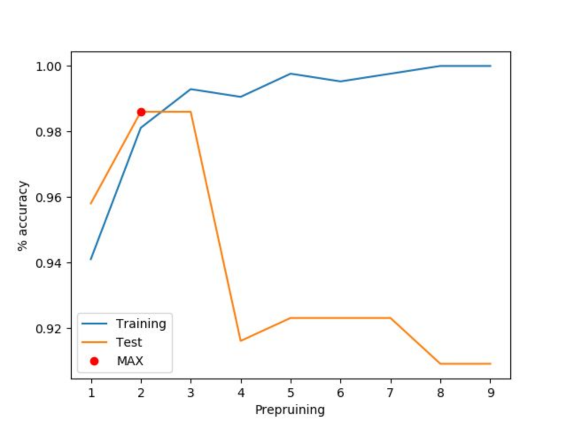
Linear SVM

Accuracy: 95.16% at Inverse Regularization: 20



Polynomial SVM

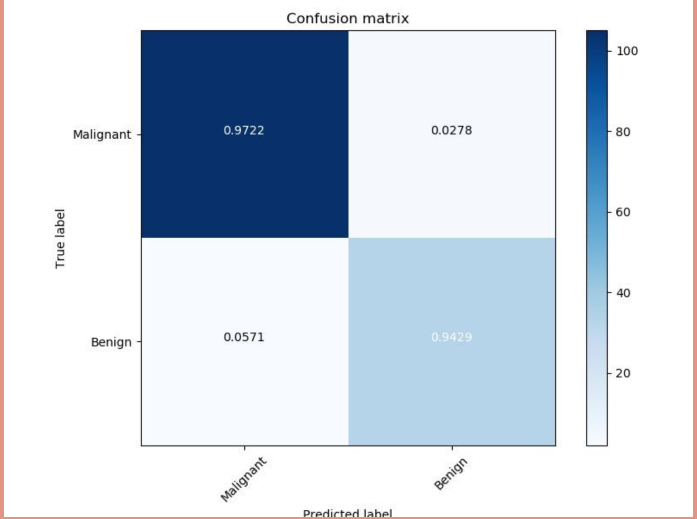
Accuracy: 98.60% at Degrees: 3



Gradient Boosting

Accuracy: 98.60% at Estimator: 2

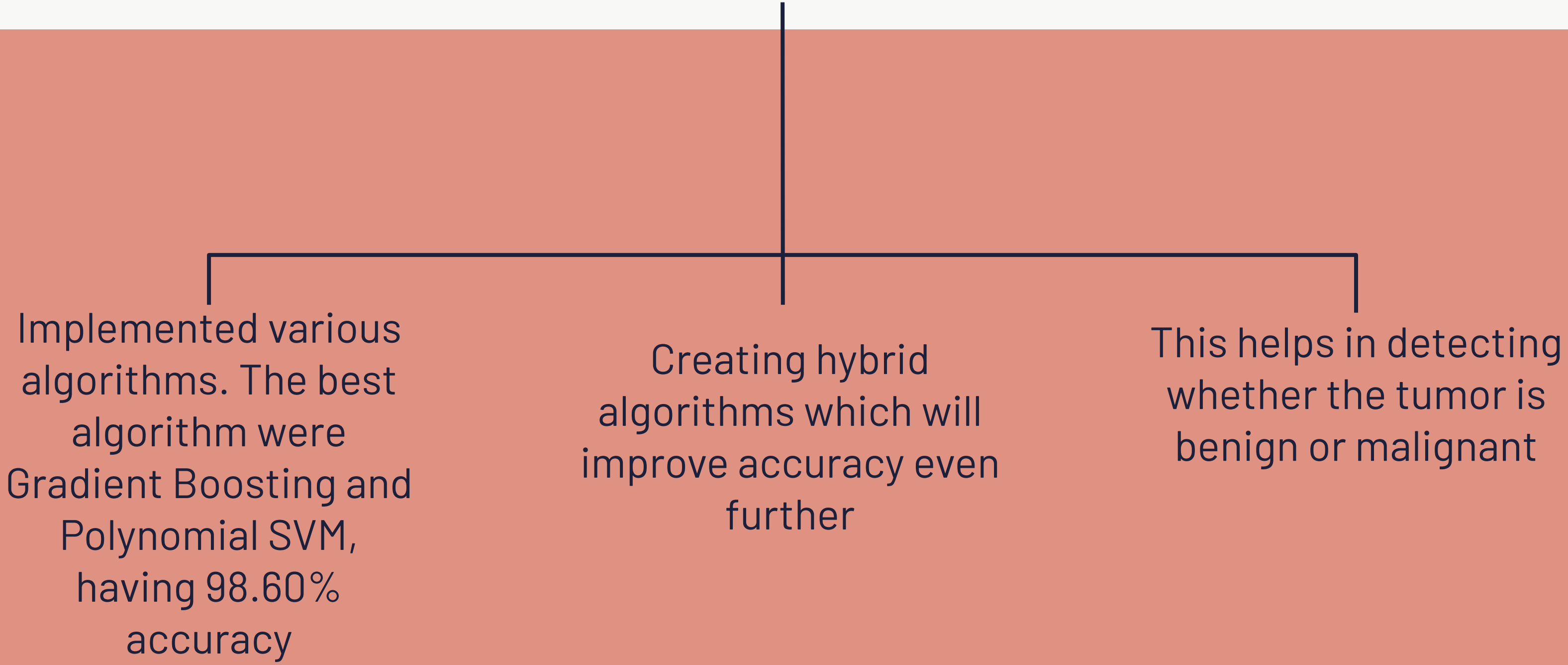
RESULT ANALYSIS



Naive-Bayes
Accuracy: 96.50%

From the above results obtained, we can see that Gradient Boosting and Polynomial SVM have the most accuracy of 98.60% .

CONCLUSION



Implemented various algorithms. The best algorithm were Gradient Boosting and Polynomial SVM, having 98.60% accuracy

Creating hybrid algorithms which will improve accuracy even further

This helps in detecting whether the tumor is benign or malignant