1. INTRODUCTION

1.1 OVERVIEW

Breast cancer is cancer that develops from breast tissue. Most types of Breast cancer are easy to diagnose by microscopic analysis of a sample of the affected area of breast. Also there are types of breast cancer that require specialized lab examinations.

More than 3.8 million US women with a history of breast cancer were alive on January 1, 2019. Approximately 1 in 8 women (13%) will be diagnosed with invasive breast cancer in their lifetime and 1 in 39 (3%) will die from breast cancer. Life time risk is an average of risk for all women and accounts for death that may preempt a breast cancer diagnosis. In order to improve breast cancer outcomes and survival, early detection is critical. There are two early detection strategies for breast cancer: early diagnosis and screening. Early diagnostics significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death, saving the lives of many. Hence, it is very important to detect this disease as early as possible.

1.2 PURPOSE

The aim of the project is to increase the number of breast cancers identified at an early stage, thus helps to provide proper treatment as early as possible. Here, I use the machine learning algorithms to predict if a tumor is benign or malignant, based on the features provided by the data. Here, I am building a model in Watson Studio and deploying the model in IBM Watson Machine Learning. To interact with the model I use Node-Red and scoring Endpoint.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

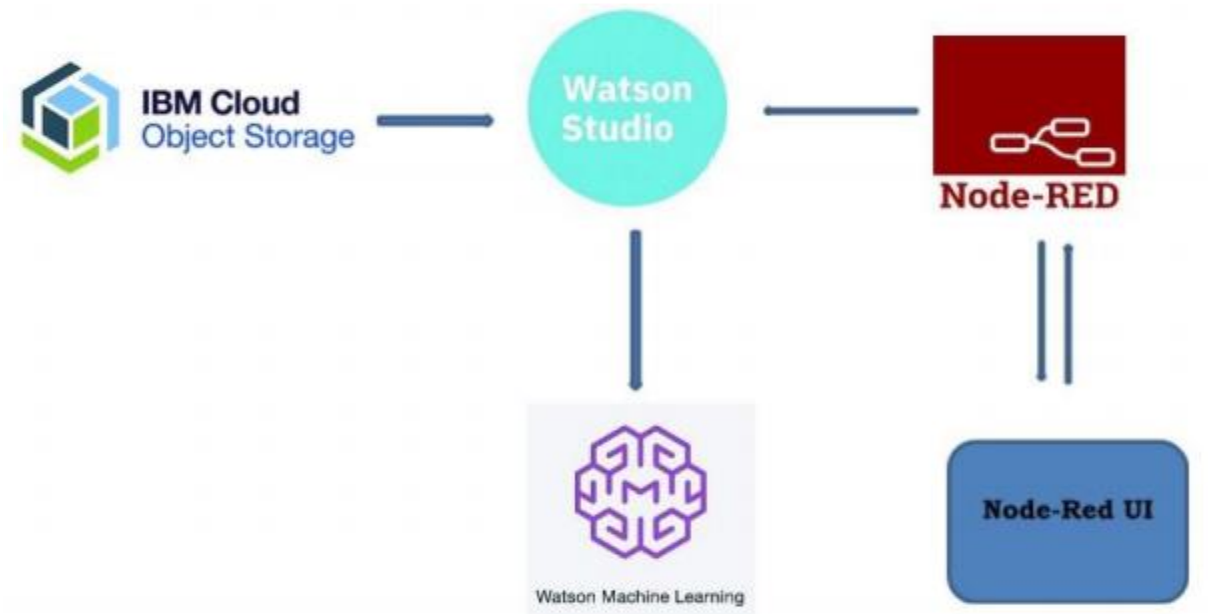
Early diagnostics significantly increases the chances of proper treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. Even thought early identification and diagnosis can significantly reduce the chances of death. This is a major existing problem in our present scenario.

2.2 PROPOSED SOLUTION

The proposed solution is to develop a model that is capable of detecting the Breast Cancer at its early stage. The Machine learning model is trained and deployed on IBM Watson Studio and an endpoint is created. The web application is built using IBM Node-Red. UI is provided and the details about tumor are collected and result is the prediction if the tumor is malignant or benign.

3. THEORETICAL ANLYSIS

3.1. BLOCK DIAGRAM



3.2. SOFTWARE DESIGNING

The project is done on IBM Watson studio on integration with IBM cloud object storage, IBM Machine learning, and IBM Node Red. IBM cloud object storage acts as storage. We train the model in IBM Watson Machine Learning and deploy them online.
IBM Node Red provides the required UI.

4. EXPERIMENTAL INVESTIGATIONS

4.1 DATA ANALYSIS AND PREPROCESSING

Our dataset contain 569 samples in which 357 are benign and 212 are malignant breast cancer. The data columns are given by data.columns ().The shape of dataset is (569, 33).

```
In [2]: data=pd.read_csv(r"C:/Users/User/Desktop/data.csv")
        data.head()
```

Out[2]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	...
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	...
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	...
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	...
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	...

5 rows × 33 columns

```
In [3]: data.shape

Out[3]: (569, 33)
```

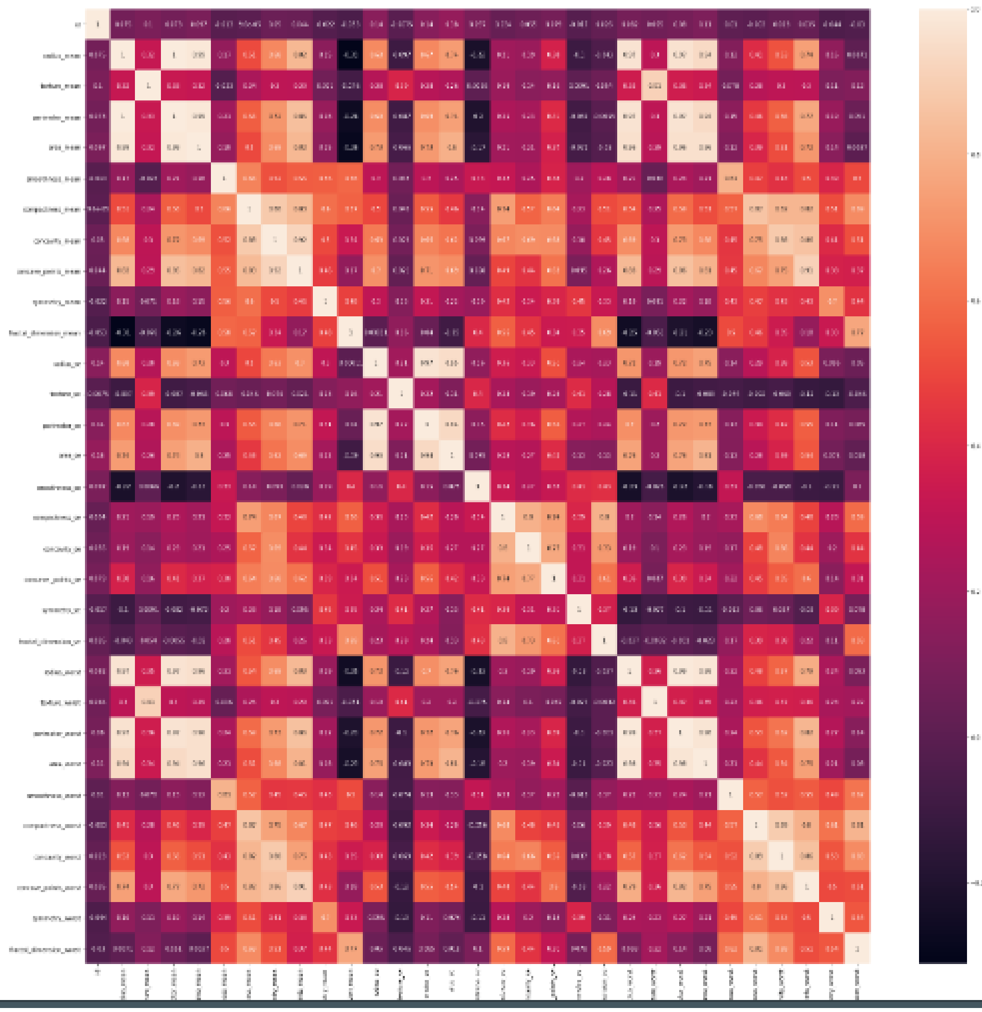
```
In [6]: data.columns

Out[6]: Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
               'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
               'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
               'radius_se', 'texture_se', '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', 'Unnamed: 32'],
              dtype='object')
```

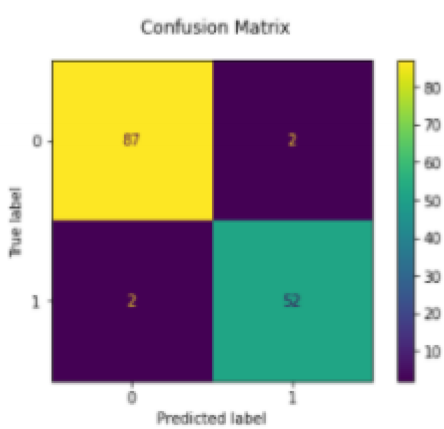
The correlation heatmap shows the correlation between various features.

```
In [82]: plt.figure(figsize=(30,30))
         sns.heatmap(data.corr(), annot=True)
```

Out[82]: <matplotlib.axes._subplots.AxesSubplot at 0x7fcd6eb83c90>



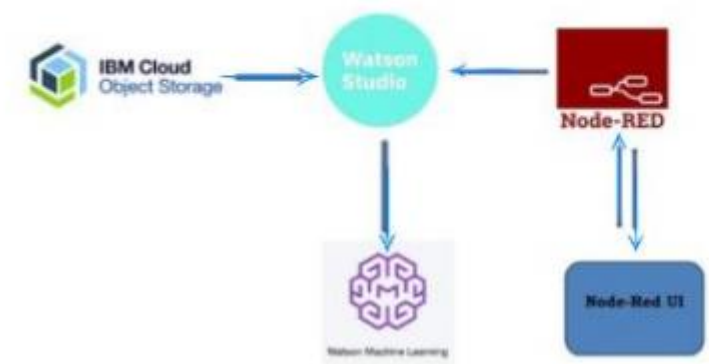
4.2 LINEAR SVM



accuracy is 97.2027972027972

A support vector machine (SVM) with a linear kernel is employed here and an accuracy of about 97.20 percent is obtained. The confusion matrix is given in the above figure.

5. FLOWCHART



This complete project is done on IBM Watson studio on integration with IBM cloud object storage, IBM Machine learning and IBM Nodered.IBM Watson Machine Learning is used to train the model deploy them online. IBM cloud object storage is used for storage purpose. IBM Node red provides the required UI.

6. RESULT

Here, we train a model using SVM classifier and it it's deployed online. The UI is provided by Node Red. The test accuracy is approximately 97.20

percent. The User Interface in which we need to enter the details to predict whether the cancer is malignant or benign is shown in the figure above.

7. ADVANTAGES & DISADVANTAGES

The first advantage is that machines work faster than humans, so faster results would be obtained. Also, chances of error occurrence can be avoided because a machine learning model will always give correct prediction while sometimes human doctors may end up with wrong conclusion.

The disadvantage is that we need a dataset to train, which should contain many data values but the person may not be having sufficient data for the model to predict. It need about 30 features to predict whether the cancer is malignant or benign.

8. APPLICATIONS

The application of this project is in Medical field. It can be employed to predict whether the breast cancer is benign or malignant and hence treatment can be given accordingly.

9. CONCLUSION

Breast Cancer (BC) is the most frequent cause of a cancer death for women around the world. Hence, if this cancer id detected at its early stage, it would be greatly helpful to improve prognosis and therefore to avoid chances of death.

10. FUTURE SCOPE

This project and the model deployed have a great scope in future. Also, more features can be added to this model to make its more accurate and faster

11. BIBILOGRAPHY

https://dataplatform.cloud.ibm.com/
https://smartinternz.com/Student/badge_workspace
https://cloud.ibm.com/docs
https://www.ibm.com/support/knowledgecenter/
https://nodered.org/docs/
https://cognitiveclass.ai/
https://towardsdatascience.com/
https://www.who.int/cancer/prevention/diagnosisscreening/breast-cancer/en/

APPENDIX

a. Source code

```
model = svm.SVC (kernel='linear')
model.fit (X_train, y_train)
predicted = model.predict (X_test)
disp = metrics.plot_confusion_matrix (model, X_test, y_test)
disp.figure_.suptitle ("Confusion Matrix")
```

```
print("Confusion matrix:\n%s" % disp.confusion_matrix)
accuracy=(disp.confusion_matrix[0,0]+disp.confusion_matrix[1,1])*100/X_
test.shape[0] plt.show()
print ("accuracy is ", accuracy)
```

The video URL is https://drive.google.com/file/d/1A8ATvKBJ-b_AzzOSh0GjhL1--NBOWMP4/view?usp=sharing