1. INTRODUCTION

1.1. OVERVIEW

According to the world health organization (WHO) Breast cancer is the most frequent cancer among women, impacting 2.1 million women each year, and also causes the greatest number of cancer-related deaths among women. In 2018, it is estimated that 627,000 women died from breast cancer-that is approximately 15% of all cancer deaths among women. 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.

Breast cancer is one of the main causes of cancer death worldwide. 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. It is important to detect breast cancer as early as possible.

1.2 . PURPOSE

The goal is to increase the proportion of breast cancers identified at an early stage, allowing for more effective treatment to be used and reducing the risks of death from breast cancer. Since early detection of cancer is key to effective treatment of breast cancer I use machine learning algorithms and to predict if a tumor is benign or malignant, based on the features provided by the data.

I will be building a model in Watson Studio and deploying the model in IBM Watson Machine Learning. To interact with the model we will be using Node-Red and scoring Endpoint.

2. LITERARTURE SURVEY

2.1 EXISTING PROBLEM

Humans can always make mistakes. As 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

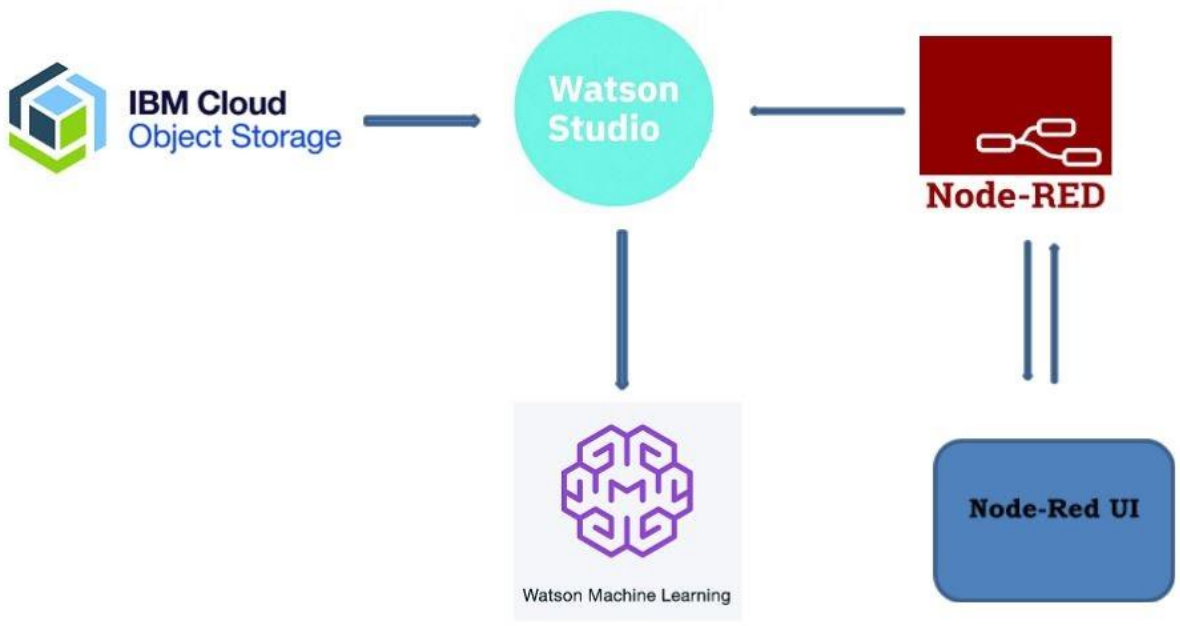
and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible.

2.2 PROPOSED SOLUTION

Develop a model that is capable of detecting the Breast Cancer in early stages. 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 is collected and predict whether the tumor is malignant or benign.

3. THEORETICAL ANLYSIS

3.1. BLOCK DIAGRAM



3.2. SOFTWARE DESIGNING

The complete project is done on IBM Watson studio on integration with IBM Machine learning, IBM cloud object storage and IBM Nodered.
IBM cloud object storage acts as a storage and using IBM Watson Machine learning we train the model and deploy them online. The UI is provided by IBM Nodered.

4. EXPERIMENTAL INVESTIGATIONS

DATA ANALYSIS AND PREPROCESSING

We have a dataset containing 569 samples in which 357 benign and 212 malignant breast cancer. The shape of dataset is (569,33)

The data columns is given by data.columns().

```
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

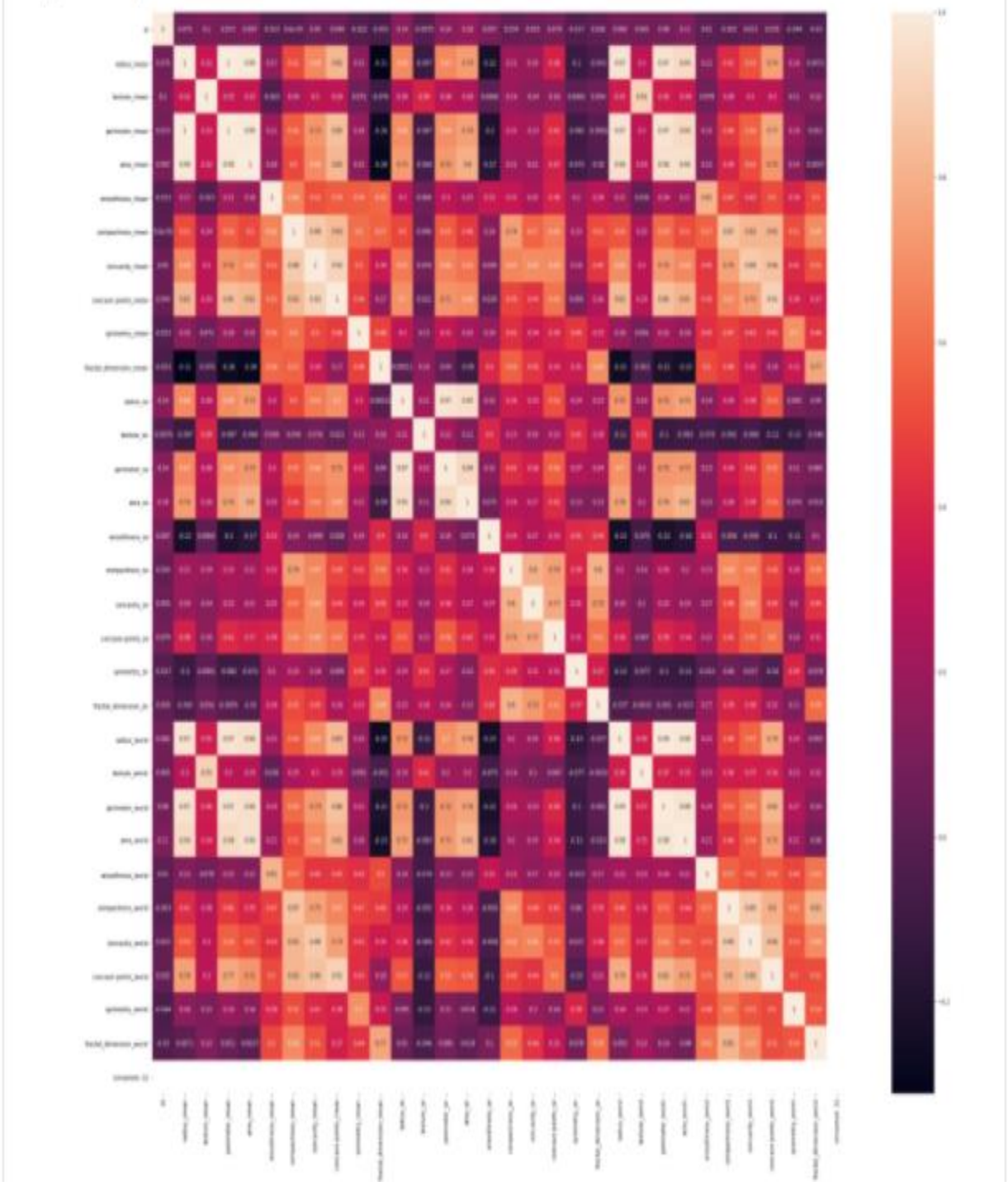
```
In [3]: data.shape
Out[3]: (569, 33)
```

```
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 between the features is given by the correlation heatmap.

```
In [9]: plt.figure(figsize=(30,30))
        sns.heatmap(data.corr(), annot=True)
```

```
Out[9]: <AxesSubplot:>
```



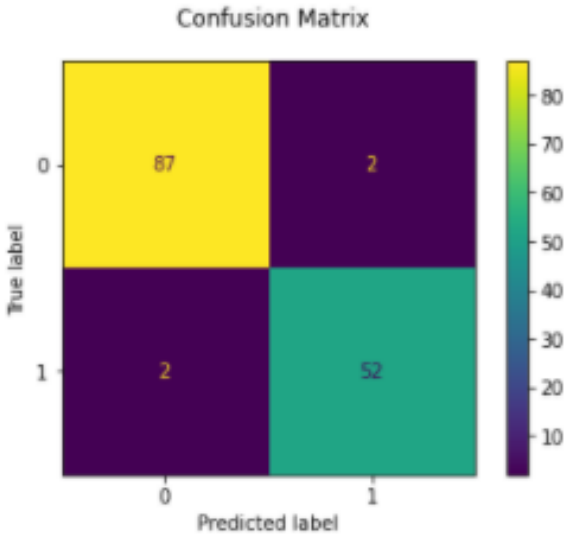
LINEAR SVM

```
model = svm.SVC(kernel='linear')
model.fit(X_train, y_train)
```

```
SVC(kernel='linear')
```

```
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)
```

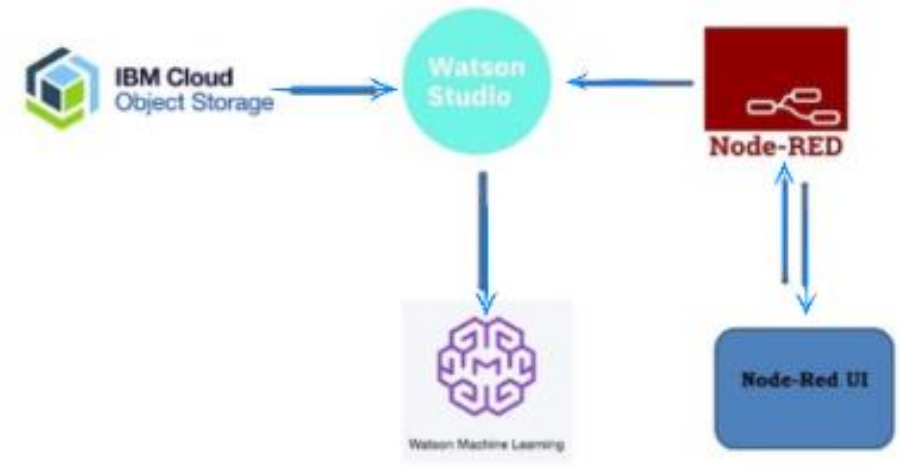
```
Confusion matrix:
[[87  2]
 [ 2 52]]
```




```
accuracy is  97.2027972027972
```

Here we use a support vector machine(SVM) with a linear kernel and got 97.2027 percent accuracy.the confusion matrix is given above.

5. FLOWCHART



The complete project is done on IBM Watson studio on integration with IBM Machine learning, IBM cloud object storage and IBM Nodered. IBM cloud object storage acts as a storage and using IBM Watson Machine learning we train the model and deploy them online. The UI is provided by IBM Nodered

6.RESULT

Trained a model with svm classifier and deploy it online . The test accuracy is 97.20 .Deployment is done online.The UI is made with nodered .The user interface in

which the details must be entered for the model to predict whether the cancer is benign or malignant is shown above.

7. ADVANTAGES & DISADVANTAGES

The advantage is that sometimes human docters may end up with wrong conclusion. But a machine learning model will always give correct prediction.

User friendly

The disadvantage is that the person may not be having sufficient data for the model to predict. It need about 30 features to predict. Precise data is need to predict correctly.

8. APPLICATIONS

The application is in MEDICAL FIELD to predict whether the breast cancer is benign or malignant.

9. CONCLUSION

In this project by using Machine Learning Algorithms to detect breast cancer, based on a dataset. Breast Cancer (BC) is a common cancer for women around the world. Early detection of BC can greatly improve prognosis and survival chances by promoting clinical treatment to patients.

10. FUTURE SCOPE

In future, more features can be added to it so can increase its understanding and prediction.

11. BIBILOGRAPHY

https://smartinternz.com/Student/badge_workspace

https://dataplatform.cloud.ibm.com/

https://www.ibm.com/support/knowledgecenter/

https://cloud.ibm.com/docs

https://cognitiveclass.ai/

https://nodered.org/docs/

https://towardsdatascience.com/

https://seaborn.pydata.org/generated/seaborn.heatmap.html

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 Github link for the breast cancer predictor is:

https://github.com/SmartPracticeschool/SPS-7640-Breast-Cancer-Risk-Prediction-System/blob/main/BREAST%20CANCER.ipynb

The video url is :

https://drive.google.com/file/d/15q9R4_OzUivcAxxpgi9IuoUyS93Chfm5/view?usp=sharing