1   INTRODUCTION

1.1 overview

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

Breast cancer represents one of the diseases that make a high number of deaths every year. It is the most common type of all cancers and the main cause of women's deaths worldwide. Classification and data mining methods are an effective way to classify data. Especially in medical field, where those methods are widely used in diagnosis and analysis to make decisions. In this paper, a performance comparison between different machine learning algorithms: Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB) and k Nearest Neighbors (k-NN) on the Wisconsin Breast Cancer (original) datasets is conducted. The main objective is to assess the correctness in classifying data with respect to efficiency and effectiveness of each algorithm in terms of accuracy, precision, sensitivity and specificity. Experimental results show that SVM gives the highest accuracy (97.13%) with lowest error rate. All experiments are executed within a simulation environment and conducted in WEKA data mining tool.

2   LITERATURE SURVEY
2.1 Existing problem

The second major cause of women's death is breast cancer (after lung cancer) 1. 246,660 of women's new cases of invasive breast cancer are expected to be diagnosed in the US during 2016 and 40,450 of women's death is estimated2 .

Breast cancer represents about 12% of all new cancer cases and 25% of all cancers in women3 . Information and Communication Technologies (ICT) can play potential roles in cancer care. In fact, Big data has advanced not only the size of data but also creating value from it; Big data, that becomes a synonymous of data mining, business analytics and business intelligence, has made a big change in BI from reporting and decision to prediction results4 . Data mining approaches, for instance, applied to medical science topics rise rapidly due to their high performance in predicting outcomes, reducing costs of medicine, promoting patients' health, improving healthcare value and quality and in making real time decision to save people's lives.

2.2 Proposed solution

In this project I will 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.

3 THEORITICAL ANALYSIS
3.1 Block diagram
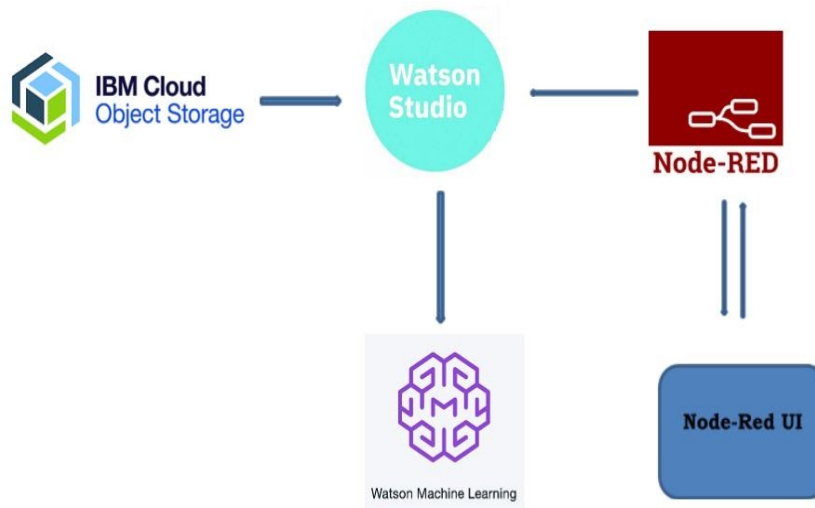
3.2 Hardware / Software designing:

in this project the skills required are basicaly software among these skills wecan cite: IBM Nodered,IBM Watson Studio,IBM Machine Learning,IBM Cloud Object Storage

4   EXPERIMENTAL INVESTIGATIONS

in this part of the report we will talk about building a machine learning models with and without autoAI in fact the role of a data scientist is one that many people aspire to, but it can come with demanding responsibilities and challenges. Some of these challenges include selecting the best algorithm for your use case and modifying it over several iterations, identifying and tuning the right set of hyperparameters several times before settling back to the defaults, and waiting for that last epoch of your model to finish.From a specifique investigation ,we can conclued that using AutoAI to build machine learning models can save you significant time and effort over manually building the same models. Steps that are repeatable, time-consuming, and prone to human error, make ideal candidates for automation.But it's so basic and made for bigenners to learn more.

5   FLOWCHART
1. collecting dataset from kaggle
2. creating IBM academic initiative account
3. login to IBM cloud
4. create cloud object storage service
5. download watson studio desktop
6. create watson studio platform
7. create machine learning service
8. create a project in watson studio
9. upload the dataset
10. Import Libraries
11. Load & Explore the Dataset

12. Perform LabelEncoding

13. Split the data into Independent and Dependent sets and perform Feature Scaling

14. Building Logistic Regression Classifier

15. Evaluate the performance of the model

16. deploy the model in watson machine learning
17. generate scoring end point
18. create node-red application
19. build UI with node-red

6   RESULT
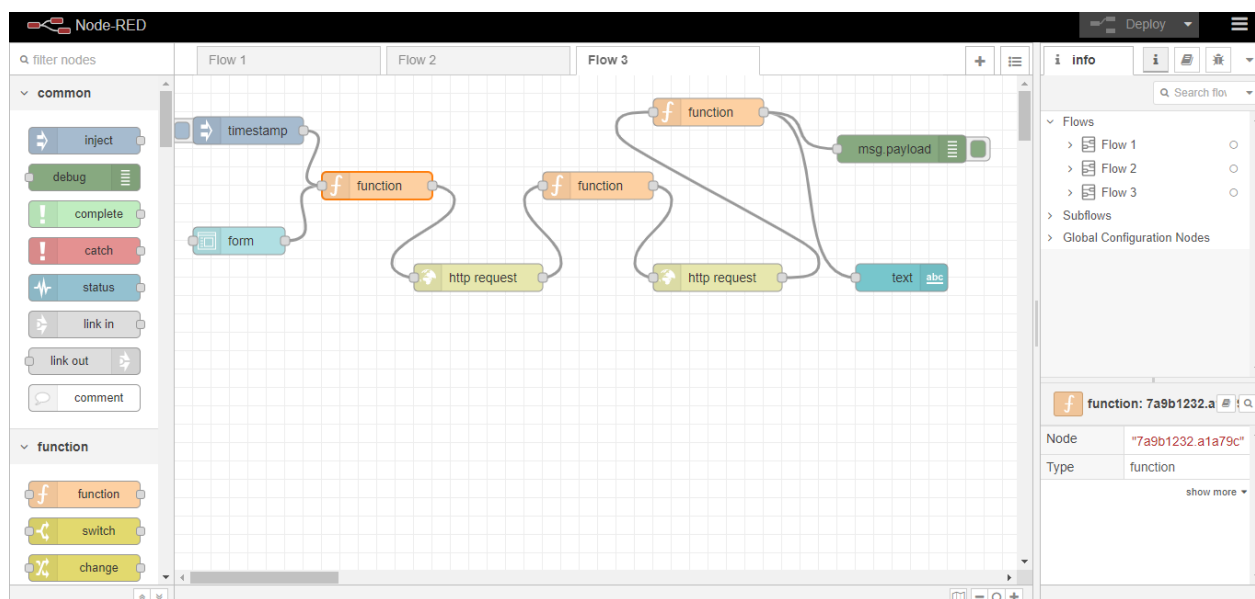
7   ADVANTAGES & DISADVANTAGES

ADVANTAGES

- Easily identifies trends and patterns
- Continuous Improvement
- Wide Applications

DISADVANAGES

- Data Acquisition:it requires massive data sets to train on
- Time and Resources: it needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy.
- Interpretation of Results: Another major challenge is the ability to accurately interpret results generated by the algorithms. You must also carefully choose the

algorithms for your purpose.
- High error-susceptibility

8   APPLICATIONS

This kind of projects is used especially in medical field, where machine learning methods are widely used in diagnosis and analysis to make decisions.
More importantly, scientists and researchers are using machine learning to churn out variety of smart solutions which will ultimately help in diagnosing and treating an illness.

Patients are set to profit the foremost because the technology can improve their outcome by analyzing the simplest sorts of treatment for them.

This application seems to remain a hot topic for the last three years.

Several promising start-ups of this industry as they are gearing up their effort with a focus on healthcare.

9   CONCLUSION

In this project i developed a model that is capable to detect the Breast Cancer in early stages and the Machine learning model is trained and deployed on IBM Watson Studio and an endpoint is created.

10   FUTURE SCOPE

The future of technology is sustainability. With hospitals and health systems looking for long-term virtual health solutions, they are asking themselves: How long will this product last and will it evolve over the next few years? Can I trust that my patients will be better served through the use of AI and ML?

The goal for introducing AI and ML technologies in healthcare is to enhance the provider and patient experience. Without AI integrated in technology, we see healthcare teams adopt multiple products to complete workflows that involve virtual health enablement, virtual team collaboration, virtual community collaboration and virtual rounding.Working in the ML technology space, our future roadmaps will be focused on building a more comprehensive set of virtual team and patient collaboration capabilities. The current

AI-integrated virtual health platforms are building sustainability for healthcare organizations and their patients, while also ensuring that regulatory compliance, including the most recent ADT Event Notification Conditions of Participation for CMS, is met each year.

11   BIBILOGRAPHY

https://www.kaggle.com/uciml/breast-cancer-wisconsin-data
https://cloud.ibm.com/login
https://cloud.ibm.com/catalog?search=object
https://www.ibm.com/in-en/products/watson-studio-desktop
https://cloud.ibm.com/catalog/services/watson-studio
https://cloud.ibm.com/catalog/services/watson-studio

https://youtu.be/Lhokhq7Hk_0

https://towardsdatascience.com/data-visualization-for-machine-learning-and-data-science-a45178970be7
https://thesmartbridge.com/documents/spsaimldocs/Datapreprocessing.pdf
https://wml-api-pyclient.mybluemix.net/

https://www.youtube.com/watch?v=ge3nYkMu1hY

https://wml-api-pyclient.mybluemix.net/

https://developer.ibm.com/components/node-red/tutorials/how-to-create-a-node-red-starter-application/

https://cloud.ibm.com/catalog

https://youtu.be/apFbFikesjA

APPENDIX

https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/940cb21f-678e-4ccd-9ffa-39ff3b905857/view?access_token=697dd8905a09c0d3c4138c51f79143881e56995158d3a885b2d8ea694efeb142