**INTRODUCTION**

Breast cancer (BC) is one of the most common cancers among women worldwide, representing the majority of new cancer cases and cancer-related deaths according to global statistics, making it a significant public health problem in today’s society.

The early diagnosis of BC can improve the prognosis and chance of survival significantly, as it can promote timely clinical treatment to patients. Further accurate classification of benign tumors can prevent patients undergoing unnecessary treatments. Thus, the correct diagnosis of BC and classification of patients into malignant or benign groups is the subject of much research. Because of its unique advantages in critical features detection from complex BC datasets, machine learning (ML) is widely recognized as the methodology of choice in BC pattern classification and forecast modelling.

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.

**LITERATURE SURVEY**

The following are some of the known risk factors for breast cancer. However, most cases of breast cancer cannot be linked to a specific cause. Talk to your doctor about your specific risk.

**Age.** The chance of getting breast cancer increases as women age. Nearly 80 percent of breast cancers are found in women over the age of 50.

**Personal history of breast cancer.** A woman who has had breast cancer in one breast is at an increased risk of developing cancer in her other breast.

**Family history of breast cancer.** A woman has a higher risk of breast cancer if her mother, sister or daughter had breast cancer, especially at a young age (before 40). Having other relatives with breast cancer may also raise the risk.

**Genetic factors.** Women with certain genetic mutations, including changes to the BRCA1 and BRCA2 genes, are at higher risk of developing breast cancer during their lifetime. Other gene changes may raise breast cancer risk as well.

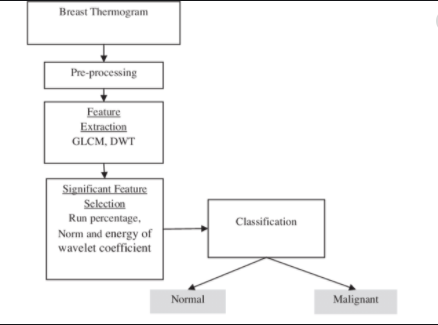
PROPOSED SOLUTION

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.

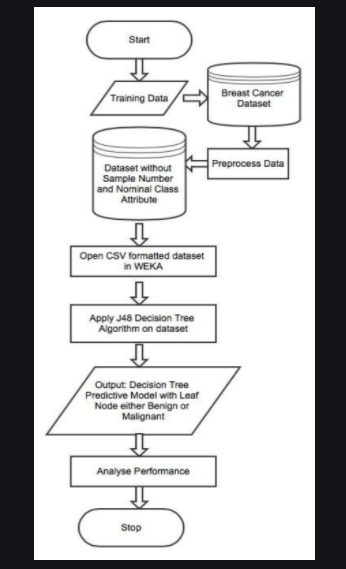
We will use machine learning with IBM Watson AI to come up with a solution.

**THEORITICAL ANALYSIS**

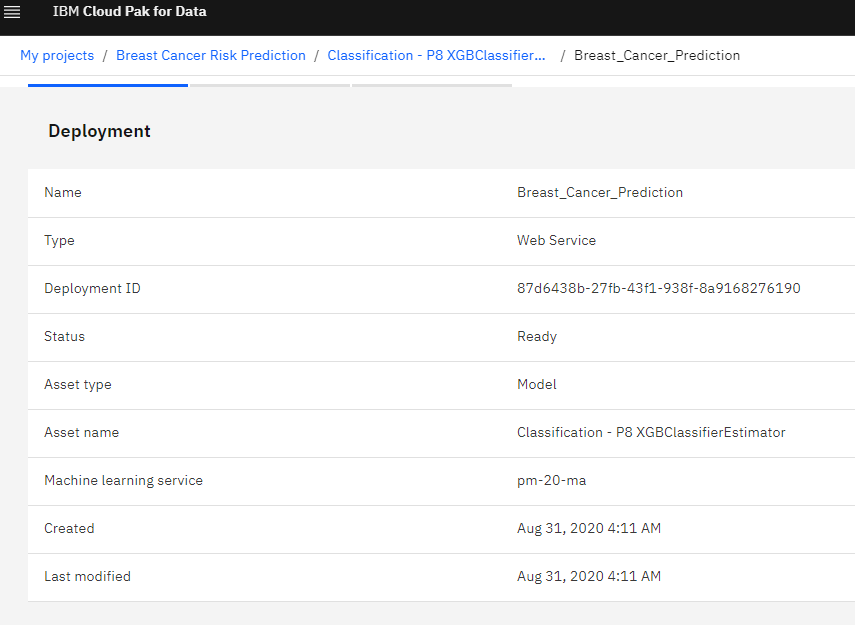
**BLOCK DIAGRAM**



**FLOW CHART**



**RESULT**



**ADVANTAGES AND DISADVANTAGES**

**1) Available 24x7:**

An Average human will work for 4–6 hours a day excluding the breaks. Humans are built in such a way to get some time out for refreshing themselves and get ready for a new day of work and they even have weekly offed to stay intact with their work-life and personal life. But using AI we can make machines work 24x7 without any breaks and they don’t even get bored, unlike humans.

**2) High Costs of Creation:**

As AI is updating every day the hardware and software need to get updated with time to meet the latest requirements. Machines need repairing and maintenance which need plenty of costs. It’ s creation requires huge costs as they are very complex machines.

**APPLICATION**

Breast cancer risk assessment provides an estimation of disease risk that can be used to guide management for women at all levels of risk. In addition, the likelihood that breast cancer risk is due to specific genetic susceptibility (such as BRCA1 or BRCA2 mutations) can be determined. Recent developments have reinforced the clinical importance of breast cancer risk assessment. Tamoxifen chemoprevention as well as prevention studies such as the Study of Tamoxifen and Raloxifene are available to women at increased risk of developing breast cancer. In addition, specific management strategies are now defined for BRCA1 and BRCA2 mutation carriers. Risk may be assessed as the likelihood of developing breast cancer (using risk assessment models) or as the likelihood of detecting a BRCA1 or BRCA2 mutation (using prior probability models). Each of the models has advantages and disadvantages, and all need to be interpreted in context. We review available risk assessment tools and discuss their application. As illustrated by clinical examples, optimal counseling may require the use of several models, as well as clinical judgment, to provide the most accurate and useful information to women and their families.

**CONCLUSION**

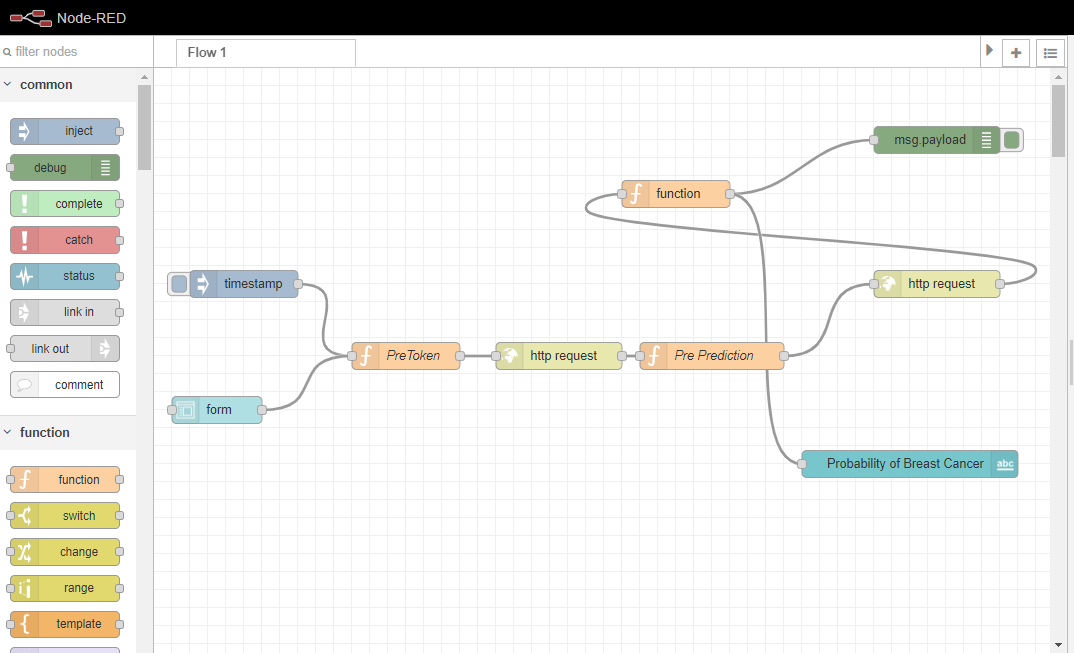
So finally we have built our classification model and we can see that XGB Classifier algorithm gives the best results for our dataset. Well its not always applicable to every dataset. To choose our model we always need to analyze our dataset and then apply our machine learning model.

**FUTURE SCOPE**

Further research may be conduct in group clustering into different sub-groups or hierarchy groups to optimize the research study such as WAM k-means cluster make more optimism.

Big data analysis method can be applied which performs data preparation, modeling and explore even if the cancer patient large data sets is an unstructured.

**SOURCE CODE**



**UI OUTPUT**

