

Breast Cancer Risk Prediction System

Project Idea

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.

Aim:

Develop a machine learning model that capable of detecting the breast cancer in early stages

Risk factors for breast cancer

A risk factor is anything that increases the risk of developing breast cancer. Many of the most important risk factors for breast cancer are beyond the control; such as age, family history, and medical history. however, there are some risk factors can control, such as weight, physical activity, and alcohol consumption.

Breast cancer risk assessment is very important for identifying women who may benefit from more intensive breast cancer surveillance; however, there is no standardized approach to office-based breast cancer risk assessment in the United States. This can lead to missed opportunities to identify women at high risk of breast cancer and may result in applying average-risk screening recommendations to high-risk women. Risk assessment and identification of women at high risk allow for referral to health care providers with expertise in cancer genetics counseling and testing for breast cancer-related germline mutations (eg, BRCA), patient counseling about risk-reduction options, and cascade testing to identify family members who also may be at increased risk

The main factors for breast cancer are female sex (more than 99% of cases of breast cancer occur in women) and advancing age. Although other characteristics have been associated with an increased risk of breast cancer, most women in whom invasive breast cancer is diagnosed do not have identifiable risk factors.

Objective

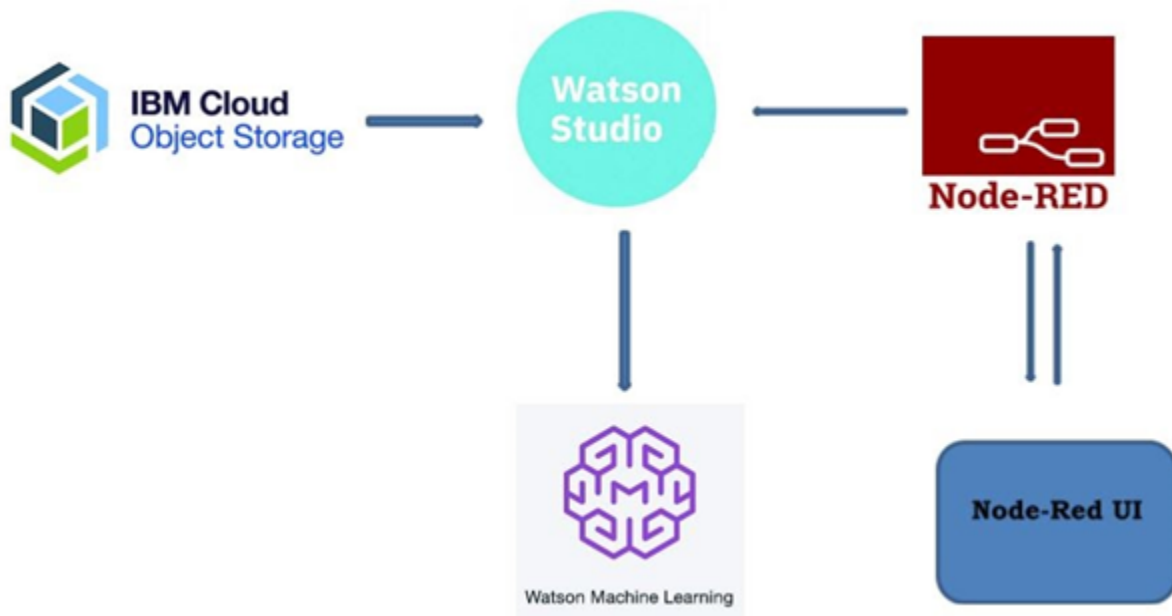
Use digital image analysis and machine-learning to:1) improve breast mass diagnosis based on fine needle aspirates(FNA), and 2) improve breast cancer prognostic estimations

Dataset

Five hundred sixty-nine consecutive patients (212 with cancer, 357 with benign masses) provided the data for the diagnostic algorithm, and an additional 118 (31 with malignant masses and 87 with benign masses) consecutive, new patients tested the algorithm. 190 of these patients with

invasive cancer and without distant metastases were used for prognosis

Architecture



IBM Cloud Object Storage is a service offered by IBM for storing and accessing unstructured data. The object storage service can be deployed on-premise, as part of IBM Cloud Platform offerings, or in hybrid form. The offering can store any type of object which allows for uses like data archiving and backup, web and mobile applications, and as scalable, persistent storage for analytics. Interaction with Cloud Object Storage is based on Rest APIs.

Watson Studio accelerates the machine and deep learning workflows required to infuse AI into your business to drive innovation. It provides a suite of tools for data scientists, application developers and subject matter experts, allowing them to collaboratively connect to data, wrangle that data and use it to build, train and deploy models at scale. Successful AI projects require a combination of algorithms + data + team, and a very powerful compute infrastructure.

Until today, there was a gap between data experts and domain experts. Only highly technical professionals in IT could organize and make sense of the vast amounts of data. Only domain experts could successfully convert data into the rich knowledge needed by AI. But domain experts and IT professionals worked in silos, with different tools and no visibility to each other work. The result was AI that fell short in its promise to augment people's expertise.

Watson Studio closes the gap with a unified experience to create new insights from knowledge contained in the data. Watson Studio enables multidisciplinary teams across the organization to collaborate. We are convinced, after working with clients around the world, that rich collaboration is key unlocking the full potential of AI

Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming. However, machine learning is not a simple process. As the algorithms ingest training data, it is then possible to produce more precise models based on that data. A machine-learning model is the output generated when you train your machine-learning algorithm with data. After training, when you provide a model with an input, you will be given an output. For example, a predictive algorithm will create a predictive model. Then, when you provide the predictive model with data, you will receive a prediction based on the data that trained the model

Node-RED is a rapidly growing easy-to-use modeling and prototyping tool that can help

someone with no development background to build sophisticated applications in minutes. Applications are developed by building data flows through a series of connected nodes.

The npm module node-red-node-watson exposes all Watson services as Node-RED nodes, enabling those services to be readily used in Node-RED application flows. Intricate applications can be built that allow Watson services to interact with the range of capabilities and services exposed as Node-RED nodes and developed by the Node-RED contributor community.

Interventions

Surgical biopsy specimens were taken from all cancers and some benign masses. The remaining cytologically benign masses were followed up for a year and surgical biopsy specimens were taken if they changed in size or character.

Outcome Measures

Cross validation was used to project the accuracy of the diagnostic algorithm and to determine the importance of prognostic features. In addition, the mean errors were calculated between the actual times of distant disease occurrence and the times predicted using various prognostic features. Statistical analyses were also done.

Results:

Using the dataset the proposed model was trained with 90% of data from the dataset. The model was tested with 10% of data from the dataset and the proposed model produced the output with the accuracy of 99%.