Project ID: SPS\_PRO\_300

Name: Adithya Narayanan

Internship title: RSIP Career Basic ML 248

**Breast Cancer Risk Prediction using IBM Auto AI**

**1: Introduction:**

**1.1: Overview:**

Breast cancer is a malignant growth/tumour beginning at the breast cells resulting from an uncontrolled division of cells which can invade into neighbouring tissues or spread (metastasize) to different parts of the body. The most common invasive cancer among women worldwide predominantly begins in the cells lining the ducts (ductal carcinoma). Sometimes breast cancer begins in the cells lining the lobules (lobular carcinoma), and in very few cases it starts in other tissues. There are different factors that contribute to the cause of this disease, such as genetics, obesity, exposure to radiation, dense breast tissue and estrogen exposure, to name a few.

**1.2: Purpose:**

Breast cancer is one of the main causes of cancer deaths worldwide. Early diagnosis 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 potential for improving the 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: Literature Survey:**

**2.1: Existing problem:**

Early detection can give patients more treatment options. In order to detect signs of cancer, breast tissue from biopsies is stained to enhance the nuclei and cytoplasm for microscopic examination. Then, pathologists evaluate the extent of any abnormal structural variation to determine whether there are tumours. Architectural Distortion (AD) is a very subtle contraction of the breast tissue and may represent the earliest sign of cancer. Since it is highly likely for it to be unnoticed by radiologists, several approaches have been proposed over the years but none using existing Machine Learning algorithms.

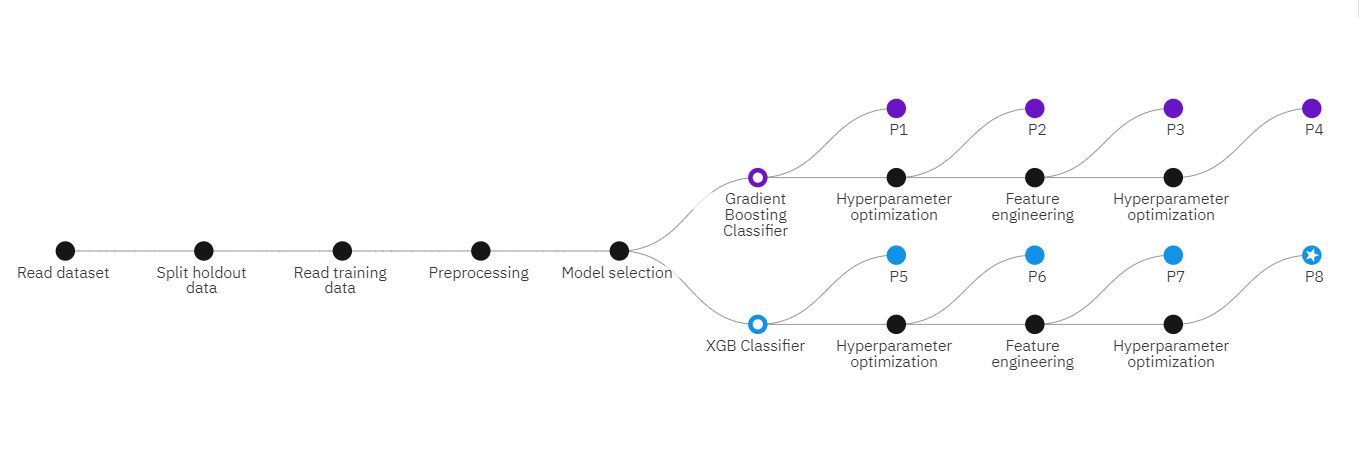
**2.2: Proposed solution:**

AI will become a transformational force in healthcare and soon, AI models will be able to get a higher accuracy when researchers have the access to more medical datasets. Here, we can develop an AI model to detect breast cancer using the dataset given.

**3: Theoretical analysis:**

**3.1: Block diagram:**

1: Auto AI:

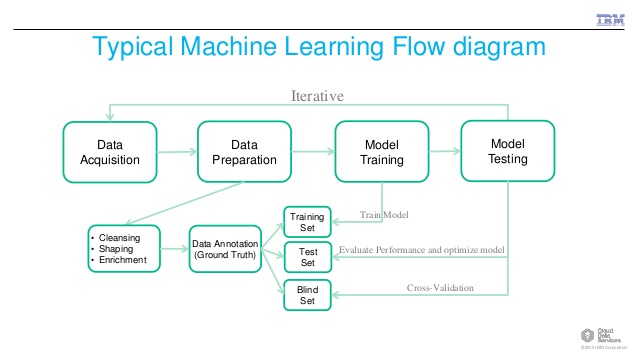


2: Nodered:

A close up of a map

Description automatically generated

3: Machine Learning Model:



**3.2: Hardware/Software designing:**

1. Run the Auto AI experiment on the given dataset. Select the Diagnosis column to be predicted and run the experiment. Select the Best Auto AI pipeline. Deploy the model and test it.
2. Create a node red service, install dashboard notes. Use various nodes such as form, function, http request node, debug, text nodes.

* Form Node: Adds a form to user interface.
* Function Node: It is a JavaScript function to run against the messages being received by the node.
* http request Node: Sends http request and returns the response.
* Debug Node: Used to display the output coming out of each node.
* Text Node: Used to display text on user interface.

**4: Experimental investigations:**

We are using a Breast cancer dataset. This breast cancer dataset was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. The dataset contains various values of following details:

1. mean\_radius

2. mean\_texture

3. mean\_perimeter

4. mean\_area

5. mean\_smoothness

6. diagnosis

**5: Flow of Events:**

1. Download the dataset.
2. Create cloud object storage, Watson studio, Machine learning service.
3. Create project in Watson studio and set up Auto AI environment.
4. Import Dataset and run the model.
5. Select the Best Auto AI pipeline.
6. Deploy and test the model in Watson Studio.
7. Create Nodered service.
8. Install dashboard notes.
9. Build an UI with Nodered.

**6: Result:**

The auto AI pipeline model was sucessfully deployed and came out with expected and accurate results during testing. A UI, namely Nodered was also successfully built for user interface. The diagnosis was successfully predicted.

**7: Advantages and Disadvantages:**

Advantages:

* Time taken to build models significantly reduces since ‘AutoAI’ prepares data, identifies features, performs optimizations, and generates models much faster than humans doing the work by themselves.
* Uncovers more use cases because exploring models is quicker, giving more time for data scientists to experiment.
* Identifying key predictors that make a difference by using the auto-feature engineering option, which makes it simpler to extract predictions from a data set.
* Ranking and exploring models by comparing candidate pipelines to determine the best model for the particular task.
* Deploying models easily through AutoAI-generated pipelines. The deployed models can then be accessed and predictions made through REST APIs.

Disadvantages:

* Limited in the number of models it can choose from.
* Data must be refined before being fed in.
* Model can’t be edited yet in a more granular way.

**8: Applications:**

AI excels at recognizing patterns in large volumes of data, extracting relationships between complex features in the data, and identifying characteristics in data that cannot be perceived by the human brain. It has already produced results in radiology, where clinicians use computers to process images rapidly, thus allowing radiologists to focus their time on aspects for which their technical judgment is critical. For example, last year, the Food and Drug Administration approved the first AI-based software to process images rapidly and assist radiologists in detecting breast cancer in screening mammograms. Integration of AI technology in cancer care could improve the accuracy and speed of diagnosis, aid clinical decision-making, and lead to better health outcomes. AI-guided clinical care has the potential to play an important role in reducing health disparities, particularly in low-resource settings. NCI will invest in supporting research, developing infrastructure, and training the workforce to help achieve these goals and more.

**9: Conclusions:**

High-accuracy prediction techniques are important in personalized medicine because they facilitate stratification of prevention strategies and individualized clinical management. This model could serve as the basis of new cost-effective and non-invasive tools to inform and prompt screening and immediate and long-term preventative actions with the potential to increase early detection and reduce the incidence of breast cancer.

**10: Future scope:**

With AI being used for cancers as complex as breast and prostate, we will hopefully see AI aid in diagnoses of even rarer cancers. By detecting cancer earlier and more accurately, prognoses of other complicated cancers, such as mesothelioma and ovarian cancer, could greatly improve. Like any other new technology, there have been criticisms as to how accurate it is and an expectation that it should work immediately. However, machine learning takes time and is constantly improving in order to provide better recommendations. As the technology becomes more advanced and researchers provide better data for it to learn from, the possibilities will be endless. Hopefully, we will see AI utilized in more industries, and help improve more lives.

**11: Bibilography:**

1. www.ibm.com
2. www.cancer.gov
3. zerocancer.org

**12: Appendix:**

UI Output Screenshot:

A screenshot of a cell phone

Description automatically generated

**Link:** <https://node-red-guwqq-2020-08-29.eu-gb.mybluemix.net/red/#flow/5cd54f9e.e2992>