Introduction:

We as humans in order to keep on living we constantly replace our cells by a process called mitosis, in which when a cell grows it splits into 2 cells in order to reproduce. This process happens millions of times every second around our body that is composed of trillions of cells. But sometimes this process is disrupted when corrupted or damaged cells split when they are not supposed too and create tumors. Depending on the nature of this tumors they can be benign (they do not spread all over your body) or malignant (they spread invasively in the rest of the individual’s body). So, it is of great importance to have methods to be able to distinguish if a tumor is benign or malignant.

So, with the new advances in technology and the development of AI, models have been trained in order to give prognosis of patience that could have cancer. AI is a powerful tool that can give an edge to the health industry to improve the efficacy of their diagnostics.

Breast Cancer

One of the most common cancers that women develop is breast cancer coming second to skin cancer. Breast cancer happens when malignant tumors form in the breasts of a women. But it has different areas where it can develop. The most common area is the duct that delivers the flow of breast milk to the nipples, and it is named ductal cancer. Second to that would be in the glands called lobules that make the breast milk called lobular cancer. [1] Figure 1 shows the anatomy of the breasts, and we can appreciate the areas of origin were the cancer can develop

Diagram

Description automatically generated

Figure 1

The way the cancer spreads is that the tumors extend to the lymph nodes in order to give the cancer cells access to the rest of your body.

Dataset

the data retrieved is of the distinct measurements of the tumors of patients retrieved by the University of Wisconsin Hospitals from 1989 to 1991. It is composed of 699 instances of patients with their tumor’s measurements and diagnosis (benign or malignant). The data has a class distribution of 358 benign tumors (65.5%) and 241 malignant tumors (34.5%). It was necessary to remove 16 instances because they had some unknown values in either one or multiple attributes.

The instances had a total of 9 attributes that embodied different measurements of the tumors in the patients. it is necessary to give a brief understanding on these measurements and in order to get a better grasp of the findings. Some of the measurements that are self-explanatory are clump thickness (tumor size), uniformity of cells size, uniformity of cell shape, bare nuclei (how exposed are the tumor cell’s nucleuses), normal nucleoli (how abnormal the nucleolus [the housing of the nucleus] is) and Mitosis (how abnormal the splitting of cells is). Others need to be explained further and so we will start with marginal adhesion which describes how much cells stick to each other. Because normally cancer cells tend to not be as sticky as normal cells. Another measurement is single epithelial cell sized which just loos at how abnormal your surface cells are on size. Then you have bland chromatin, chromatin is the source material of chromosomes that make the DNA living in the nucleus of the cells, and in this case, it means how uniform this is.

All these attributes have a value in range of 1-10 1 being the most normal and 10 the most abnormal. One can conclude that if all the attributes are in an abnormal range, we can assume that the tumor is going to be malign, and in some sense that could be correct. But what matters most is what is the relation of these attributes and which of them have the most impact.

Models

To be able to analyze the data in this project I came to the conclusion that giving the user an option of using multiple models in order to get to a conclusion was necessary to give not only more input in the analysis itself but also the models.

Logistic Regression

One of this models is Logistic Regression

References:

Admin. (2020, October 22). Nucleolus - Function, Difference Between Nucleus &amp; Nucleolus. BYJUS. <https://byjus.com/biology/nucleolus/>.

Chromatin. Genome.gov. (n.d.). <https://www.genome.gov/genetics-glossary/Chromatin>.

Farlex. (n.d.). Bare Nucleus. The Free Dictionary. <https://medical-dictionary.thefreedictionary.com/Bare+Nucleus#:~:text=Naked%20Nucleus,typically%20seen%20in%20cell%20degeneration>.

UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set. (n.d.). <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>.

What Is Breast Cancer?: Breast Cancer Definition. American Cancer Society. (n.d.). <https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html>.

What Is Cancer? National Cancer Institute. (n.d.). <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>.