Breast Cancer Diagnosis Using Machine Learning Algorithms

Aroofa Mohammad(aam170007)

Machine Learning Final Project Proposal

March 3, 2022

Introduction

Cancer is the second most common cause of death in the US, after heart disease. One of the most common and deadly types of cancer is breast cancer. Which has caused around 43,250 women and 530 men to lose their life in 2022. Therefore, the ability to predict and diagnose cases of breast cancer is crucial for the importance of someone's life. With life and death playing a part, accuracy is very important in the survival rate of the patient and the best course of treatment. Machine learning models/algorithms have shown a significant impact on the process of breast cancer prediction and early diagnosis.

Motivation and Objective of This Research

The objective is to decrease the number of lives lost to breast cancer. Since there is no real cure that just targets the cancer cells, being able to detect breast cancer early can save a lot of lives. Using machine learning to detect cancer cells will also allow doctors to predict the stage of breast cancer to determine the best course of action/treatment.

Research Aim

The objective of this research work is to develop and implement the most accurate machine learning techniques to find the most effective and accurate detection of early stages of breast cancer. The algorithms that will be used in this research are going be Logistic Regression, Support Vector Machine (SVM), Decision tree, and K-Nearest Neighbors (KNN) on database provided by Wisconsin Breast Cancer.

Background of the study

Deep learning, a class of machine learning algorithms, has sparked a lot of interest in its application to medical imaging problems. Detecting breast cancer early can be done by implementing techniques taught in machine language. Researchers have exploited numerous databases provided by hospitals to determine various features. Research provided by such researchers has great significance in providing the strength and weaknesses of each machine learning technique. Such techniques are Logistic Regression, Support Vector Machine (SVM), Decision tree, and K-Nearest Neighbors (KNN). Using the Wisconsin Breast Cancer Database, this research project will formulate and implement machine learning models. All work of this research will be documented regardless of where the research worked or did not.

Proposal method

The research will start with a systematic analysis of each individual machine learning model/algorithm (named above). Then each technique will be formula and implementation to help determine breast cancer. Finally, the end report will contain an analysis of the weaknesses and strengths of the different approaches. The report will also contain a list of all work done in the research, including what did not work.

Conclusion

In conclusion breast cancer is big problem that has yet to solved but with the help of machine learning algorithm like Logistic Regression, Support Vector Machine (SVM), Decision tree, and K-Nearest Neighbors (KNN) there is still hope for those who have cancer.

References

https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

<https://www.cancer.org/latest-news/facts-and-figures-2022.html>

ACS Medical Content and News Staff. “2022 Cancer Facts & Figures Cancer: Cancer Death Rate Drops.” *American Cancer Society*, American Cancer Society, 12 Jan. 2022.

<https://www.nature.com/articles/s41598-019-48995-4>

Shen, Li, et al. “Deep Learning to Improve Breast Cancer Detection on Screening Mammography.” *Nature News*, Nature Publishing Group, 29 Aug. 2019.