A Study Into The Applications of Machine Learning in The Diagnosis & The Prognostication of Breast Cancer

Department of Computing

Letterkenny Institute of Technology

Machine Learning

Ultan Kearns & Liam Millar

Abstract—**Machine Learning has been used in a variety of healthcare fields and has proven to be very successful in regards to automating / helping with certain tasks many tasks ranging from identifying similarities in different strains of DNA to diagnosing varies types of diseases and predicting the prognosis. Machine Learning is particularly useful when it comes to analyzing and comparing data, this data can then be used to train a model to notice certain similarities in malignant and benign tumours. This paper aims to analyze which methods would be best suited to identifying factors which contribute to a cancer being malignant or benign and to predict the prognosis of breast cancer based on those factors.**
keywords - Machine Learning, Breast Cancer, Prognosis, Diagnosis

I. INTRODUCTION

Machine Learning is a field which aims to train machines to think and learn like humans. This is done by feeding data into the machine, the machine then analyzes the data and begins to start noticing patterns within the data and starts to create a model, this model can be used to train the machine so that when new data is feed in the machine can easily detect certain characteristics about the data and classify it. Machine Learning is a subset of AI which is comprised of many fields, and has been proven to perform very well in predicting outcomes in a large variety of areas such as STEM fields, Economics, Social Sciences and various others. The ultimate goal of machine learning is to train a machine to perform a task without any human intervention and to learn to adapt to new data and new patterns which appear in the data. There has been an increased reliance on machine learning in the past few years and it has become ubiquitous in the modern world.

There are many ways which a machine may choose to learn we will explain the types of learning we used when trying this model in the upcoming sections.

A. Supervised Learning

Supervised learning is a type of machine learning that involves training a machine with labelled training data to make inferences about new data. The machine reads in the data and begins to notice certain patterns from it and uses the learned patterns to make inferences about new data. There are many algorithms which can be used in supervised learning, I will list a few we used below.

1) Linear Regression: This algorithm involves finding a line to best fit the data, this is achieved by modifying parameters to find the best fitting line for a given dataset. This method works by learning to predict a pattern within the data and drawing a line through a series of data points so that the line

fits is close to as many of the points as possible, in this way we can make predictions by seeing how closely a given point would fit our line.

2) Naive Bayes: Naive Bayes is a technique which makes features of an object have the same weight when trying to classify the data. This algorithm works well on small data sets and it can improve the results of a model when working with a limited dataset. An example of Naive Bayes would be to train an image classfier for identifying cars, using Naive Bayes the property of a cars wheels would have just as much impact as the number of doors of the car.

3) K Nearest Neighbour: K Nearest Neighbour is an algorithm which measures the distance between data points by using a distance metric eg: Cosine distance, Euclidean distance, Jaccardian distance etc and groups data points which are near each other into clusters. In K Nearest Neighbour we make the assumption that data points within a certain distance must have some common properties or features, in this way we can separate the data into groups which share these features.

B. Breast Cancer

Breast cancer is a type of cancer which forms in the breast tissue, it has many signs and symptoms such as:

- New lump in the breast or underarm (armpit).
- Thickening or swelling of part of the breast.
- Irritation or dimpling of breast skin.
- Redness or flaky skin in the nipple area or the breast.
- Pulling in of the nipple or pain in the nipple area.
- Nipple discharge other than breast milk, including blood.
- Any change in the size or the shape of the breast.
- Pain in any area of the breast.

[1]

Breast cancer is also one of the most common occurring cancers in women and the average risk of a woman developing cancer in her lifetime is about 13% or 1/8 in the United States alone[3] and in 2020 there were 2.3 million women diagnosed with the disease of which 685,000 succumbed to the disease[2]. Breast cancer treatment is has a very high survival rate if caught early, the survival rate for those in high income countries was 90% [2].

C. Discuss various uses of machine learning in medicine today

D. Aims of the project

The aim of this project is to train a model using machine learning to analyze data from both benign and malignant tumours to detect patterns in the data and to predict whether tumours will be malignant or benign given certain parameters with reasonable accuracy.

II. METHODOLOGY

A. Algorithms

B. Machine Learning Techniques

III. CHALLENGES

A. Over fitting

Over fitting occurs when the model is performing very well on the training data but performs poorly on new data, this is due to the fact that the model has been trained to fit the training data so well that it cannot adapt to different data.

B. Under fitting

C. Outliers

IV. APPLICATIONS

A. Diagnosis

B. Prognosis of Breast Cancer

V. CONCLUSION

A. Accuracy of Model

B. What could be done differently

REFERENCES

[1] CDC. *What Are the Symptoms of Breast Cancer?* URL: https://www.cdc.gov/cancer/breast/basic_info/symptoms.htm.

[2] World Health Organization. *Breast Cancer.* URL: https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[3] American Cancer Society. *How Common Is Breast Cancer?* URL: https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html.