Report

# Chapter 1 introduction

## Background

Prostate cancer is one of the most common types of cancer found in men worldwide. Its significant impact on public health is demonstrated by the fact that it ranks as the second most common cause of cancer related deaths in men (Siegel, Miller and Jemal, 2020). Male reproductive health depends on the prostate gland, which is situated in front of the rectum and behind the bladder. A cancerous change of the prostate gland will have serious health effects.

The early identification of prostate cancer is critical to improving the lives of the patients. Prostate-specific antigen (PSA) testing and digital rectal exams (DRE) are examples of diagnostic techniques. Although these techniques are popular, their sensitivity and specificity are limited, which frequently leads to false positives and false negatives (Heidenreich et al., 2014). These errors can result in an under diagnosis, which postpones the treatment, or an overdiagnosis and overtreatment, which cause patients unnecessary worry about potential problems.

Recent advances in the imaging technology, specifically multiparametric magnetic resonance imaging (mpMRI), have improved the ability to identify and characterise prostate cancer. Combining functional and anatomical imaging, MRI offers a thorough perspective that improves the ability to distinguish between healthy and cancerous tissues (Rosenkrantz et al., 2016). Nonetheless, mpMRI image interpretation requires a high level of knowledge and is highly variable among the observers.

The integration of machine learning (ML) into medical imaging offers a promising solution for these kinds of problems. Deep learning-based machine learning algorithms, in particular, are capable of processing enormous volumes of imaging data and spotting intricate patterns that could be invisible to human observers. According to Litjens et al. (2017), these algorithms may decrease variability, increase diagnostic accuracy, and support clinical decision making.

Machine learning comprises a wide range of approaches that are divided into two categories supervised and unsupervised learning. In supervised learning, models are trained using labelled data to predict or categorise the data. On the other hand, unsupervised learning entails identifying structures or hidden patterns in unlabelled data. The medical industry can benefit greatly from both forms of learning, particularly in the areas of cancer detection and diagnosis.

In the context of prostate cancer, machine learning models have been developed for various tasks such as tumour detection, Gleason grade prediction, and treatment response monitoring. Recurrent neural networks (RNN), convolutional neural network (CNN), and support vector machines (SVM) are notable machine learning approaches. To varying extents these models have improved prostate cancer diagnosis efficiency and accuracy.

For instance, since CNNs can learn spatial hierarchies from input images, they are particularly useful for image analysis tasks like tumour segmentation and classification (Pellicer-Valero et al., 2022). Similar to this, RNNs which are capable at preprocessing sequential data have been used to forecast treatment results by taking into account the history and advancements of patients (Mirsamadi et al., 2017).

Even with the improvements, there are still a number of difficulties in applying ML to clinical practice. These include obtaining clinical validation to verify the model’s efficiency in real world scenarios, guaranteeing the interpretability and transparency of ML decisions, and training robust models on big annotated datasets. To tackle these obstacles, data scientists, physicians, and regulatory agencies must continue their research and work together.

The development of sophisticated imaging technologies and machine learning presents a possible alternative to the poor performance of existing methods for the identification of prostate cancer. Utilising these technologies can lead to better patient outcomes by enhancing diagnostic accuracy and personalising treatment approaches. To overcome current obstacles and fully realise the potential of these state-of-the-art instruments in the treatment of prostate cancer, more research and innovation in this area are imperative.

## Problem statement

Prostate cancer is a major worldwide health concern, as it is one of the most often diagnosed cancers among the men and the leading cause of cancer related death. For successful therapy and better patient outcomes, clinically significant prostate cancer lesions must be identified early and accurate. Due to the poor sensitivity and specificity of traditional diagnostic techniques such as digital rectal exams (DRE) and prostate specific antigen (PSA) testing, there is a risk of overdiagnosis, overtreatment, or missing diagnoses.

Multiparametric magnetic resonance imaging (mpMRI) improves the capacity to detect and characterise prostate cancer by providing both anatomical and functional imaging. The assessment of prostate lesions clinical significance (ClinSig) using mpMRI picture interpretation is still difficult and heavily reliant on radiologists’ skill, which often results in inter observer variability.

Machine learning offers a promising answer to these problems by automating the interpretation of mpMRI images and predicting the ClinSig score of prostate lesions. Large amounts of imaging data can be processed by ML models, especially deep learning approaches, which can then be used to spot subtle patterns that human observers might miss, enhancing diagnostic consistency and accuracy.

This project focuses on creating and deploying a machine learning model for predicting the ClinSig score of prostate lesions based on T2 weighted mpMRI images.

## justification of the study

This study is primarily justified by the possiblility that it may greatly enhance the diagnosis of prostate cancer by offering a more precise and reliable way to predict the ClinSig score of prostate lesions. By lowering diagnostic mistakes and inter observer variability, this automated method can assist radiologists in making more informed treatment decisions. Furthermore, the effective diagnostic process can be improved by integrating machine learning models into clinical processes. This guarantees prompt and suitable interventions, which are critical for improving patient outcomes. This project aims to lessen the burden of prostate cancer on healthcare systems and contribute to personalised treatment regimens by increasing regimens by increasing diagnosis accuracy and consistency.

## Research questions

1. How accurately can a convolutional neural network (CNN) model predict the ClinSig score of prostate lesions form the T2 weighted mpMRI images?

## Aims and objectives

This project’s main goal is to create and verify a machine learning model that may be used to reliably predict prostate lesion clinical significance (ClinSig) scores using multiparametric magnetic resonance images (mpMRI) data. This goal will be met through the following specific objectives.

1. Investigate the machine learning methods that are currently being utilised to categorise and predict prostate cancer.
2. Load and prepare the Prostatex dataset images (T2 weighted images) specified in the detailed description from the Prostatex challenge.
3. Using the prepared data, create a convolutional neural network model to predict the ClinSig score of prostate lesions
4. Evaluate the created CNN models performance in comparison to current classifier models and conventional diagnostic techniques

# Chapter 2 Literature review

## 2.1 overview of the prostate cancer

* About prostate cancer and its biology
* Current detection methods

## 2.2 Machine learning in medical imaging

* Application of machine learning in medical imaging
* Advantages and challenges

## 2.3 Machine learning techniques for cancer detection

* Supervised learning
* Unsupervised learning
* Deep learning

## 2.4 Related work

This section is about the previous studies about the prostate cancer detection using the machine learning.

Table showing all the papers for the literature review (try to include 10 papers)

|  |  |  |  |
| --- | --- | --- | --- |
| Paper | Classifiers | Dataset Used | Results |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## 2.5 summary of findings

* Comparison of results
* Discussion on the best performing techniques

# Chapter 3 Research methodology

## 3.1 Overview

Outline of the research approach

## 3.2 Data collection

Description of the dataset and source of the dataset

## 3.3 Data Preprocessing

Explaining the pre-processing steps

## 3.4 Feature Selection

Explaining what are the features consider for the model training and any methods done for choosing those features

## 3.5 Model Development

Machine learning models used in the project. Their parameters, architecture and training process

## 3.6 Model evaluation

Accuracy, precision, recall, F1-Score

# Chapter 4 results and analysis

## 4.1 Data analysis

Viewing of the dataset and understanding about the dataset, preprocessing, feature selection

## 4.2 Model performance

detailed results of the model training and evaluation, including the confusion matrix etc

## 4.3 Comparison with the existing methods

Comparing the model’s performance with the literature reviewed models’ performance

# Chapter 5 Conclusion and the future work

## 5.1 conclusion

## 5.2 limitations

## 5.3 recommendations for the future work