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 effective for tasks such as tumour segmentation and classification (Pellicer-Valero et al., 2022).

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 possibility 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.

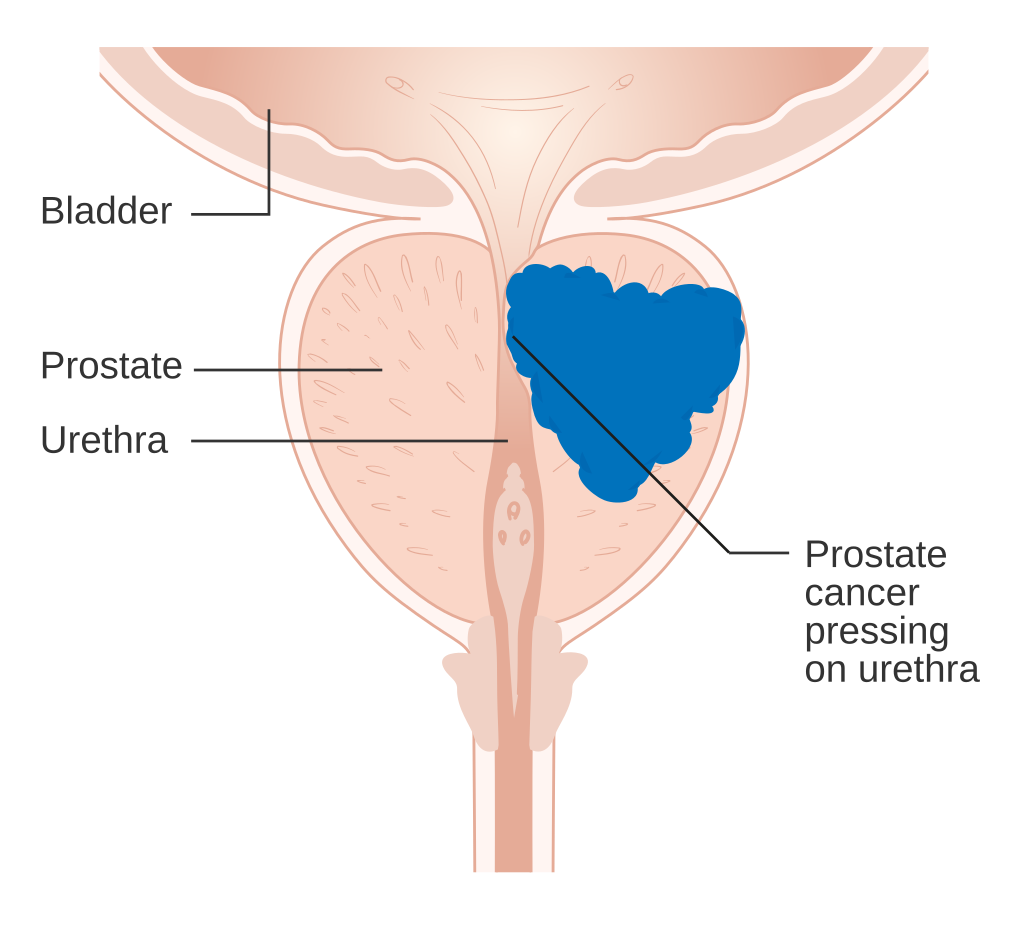
1. Investigate the machine learning methods that are currently being utilised to categorise and predict prostate cancer.
2. Load and prepare the Prostatex dataset images (T2 weighted images) specified in the detailed description from the Prostatex challenge.
3. Using the prepared data, create a convolutional 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

## Biology

Prostate cancer is a major health concern worldwide, being the second most common cancer diagnosed in men and the second leading cause of cancer related deaths (Rawla, 2019). The prostate is a small gland that sits in front of the rectum and beneath the bladder in the male reproductive system. It encircles the urethra, the tube that exits the body with the urine. The production of seminal fluid, which feeds and moves sperm, is the prostates main job.



Prostate cancer develops when cells in the prostate gland grow uncontrolled. The majority of prostate cancers are adenocarcinomas, which arise from the gland cells responsible for producing prostate fluid. Prostate cancer can grow slowly or aggressively. Slow growing prostate cancer may not produce noticeable signs or difficulties for many years. Aggressive prostate cancer, on the other hand, can spread rapidly to other regions of the body, including the bones and lymph nodes, resulting in serious health concerns.

Prostate cancer develops and progress as a result of numerous genetic and molecular alterations. The key biological process is:

**Genetic mutations**: mutations in specific genes, such as PTEN, TP53, and BRCA1/2, can contribute to the development and progression of prostate cancer. These mutations can activate oncogenes or deactivate tumour suppressor genes, affecting normal cell growth regulation (Robinson et al., 2015).

**Androgen Receptor Signalling**: Prostate cancer cells often rely on androgens to proliferate. The androgen receptor (AR) pathway is essential for the information and maintenance of prostate tissues. In prostate cancer, this pathway is frequently dysregulated, resulting in enhanced cell proliferation and survival. Therapies targeting the AR pathway, such as androgen deprivation therapy (ADT), are routinely used to treat prostate cancer (Attard et al., 2016).

**Inflammation**: Chronic inflammation in the prostate has been related to the development of prostate cancer. Inflammatory activities can cause DNA damage and promote malignant alterations in prostate cells (De Marzo et al., 2007).

**Microenvironment**: The tumour microenvironment, which include interactions with stromal cells, immune cells, and the extracellular matrix, is crucial to prostate cancer growth and metastasis. Understanding these relationships can help create new therapeutic tactics (Baron and Rowley, 2012).

Prostate cancers molecular complexity emphasises the necessity of early identification and efficient therapy. Advances in molecular biology and genetics continue to increase our understanding of prostate cancer, leading to more effective diagnostic tools and therapies.

## Current detection methods

Methods for detection and diagnosis of prostate cancer

**Digital Rectal Exam**:

During a DRE, a healthcare provider inserts a gloved, lubricated finger into the rectum to inspect the prostate for any abnormalities or tumours. While this method can detect anomalies, it is subjective and has low sensitivity and specificity (Heidenreich et al., 2014).

**Prostate specific Antigen (PSA) test:**

The PSA test measures the amount of prostate specific antigen in the blood. Elevated PSA values may suggest prostate cancer, benign prostatic hyperplasia or prostate inflammation. Although commonly used, the PSA test can produce false positives and negatives, resulting in overdiagnosis or missed diagnosis (Mottet et al., 2017)

**Transrectal Ultrasound (TRUS):**

TRUS involves inserting an ultrasonic probe into the rectum to obtain pictures of the prostate. It is routinely used to guide prostate biopsies however it has little ability to identify prostate cancer on it own (Heidenreich et al., 2014).

**Prostate Biopsy**:

If initial results indicate the existence of prostate cancer, a biopsy is performed to collect tissue samples from the prostate. These samples are analysed under a microscope to confirm the existence of cancer cells and calculate the Gleason score, which represents the cancers aggressiveness (Epstein et al., 2016).

**Multiparametric Magnetic Resonance Imaging (mpMRI):**

MpMRI uses both anatomical and functional imaging to provide extensive information about the prostate and any worrisome abnormalities. It comprises T2-weighted imaging, diffusion-weighted imaging and dynamic contrast-enhanced imaging. MpMRI has enhanced accuracy and is now widely used to guide biopsies and treatment planning (Rosenkrantz et al., 2016).

Despite advances in detecting methods, getting accurate and consistent diagnoses remains a challenge. The incorporation of ML approaches into prostate cancer detection, particularly with mpMRI shows potential for increasing diagnostic accuracy and decrease variability. ML models can analyse complex imaging data discover subtle trends and help radiologists and oncologists make more exact diagnoses (Litjens et al., 2017).

## 2.2 Machine learning in medical imaging

Machine learning (ML), particularly deep learning has transformed the area of medical imaging the area of medical imaging by allowing computers to automatically analyse and interpret complicated medical pictures. These algorithms learn from enormous datasets of annotated photos, detecting patterns and details that humans may not see. The use of machine learning in medical imaging encompasses a wide range of modalities and clinical illness detection, classification, segmentation, and patient outcome prediction.

In radiology, machine learning algorithms are used to detect and classify anomalies in images such as X-rays, CT scans, and MRIs. CNN have been used to identify lung nodules in chest CT images, breast cancer in mammograms and brain tumours in MRI scans (Litjens et al., 2017). In digital pathology machine learning algorithms analyse tissue samples to discover cancer cells, classify tumour kinds and forecast illness prognosis. CNNs and other deep learning models have demonstrated great accuracy in detecting various malignancies from histopathological pictures (Komura and Ishikawa, 2018).

Another important application of machine learning in medical imaging is image segmentation, which involves splitting a picture into relevant sections. ML models, particularly U-Net and its derivatives are commonly utilised for segmentation tasks in imaging modalities such as MRI and CT. this precision is required for treatment planning and disease monitoring (Ronneberger et al., 2015).

ML models can also predict patient outcomes using imaging data and clinical information. For examples, machine learning algorithms may analyse brain MRI scans to predict the course of neurological illness such as Alzheimer’s or the chance of recurrence in cancer patients following therapy (Esteva et al., 2019). ML approaches are used to improve the image quality and rebuild images from the raw datasets. These methods are particularly beneficial for decreasing noise in low dose CT scans, increasing MRI image resolution and creating synthetic images from incomplete datasets (Wang et al., 2018).

ML in medical imaging provides various benefits such as ML algorithms can recognise minute patterns in medical images that humans may miss, resulting in more accurate and timely diagnosis. This is particularly useful for detecting tiny or unclear lesions (Litjens et al., 2017). Unlike human radiologists, ML models produce consistent results lowering inter and intra observer variability and increasing the readability of diagnoses and treatment regimens (Shen, Wu, and Suk, 2017). Furthermore, algorithms can rapidly analyse large amounts of imaging data considerably lowering the time necessary for picture interpretation. This efficiency is critical settings with large patient populations and limited radiology resources (Wang et al., 2016). ML models can combine imaging data with other patient information such as genetic and clinical data to deliver personalised therapy recommendations improving the precision of treatments customised to specific patient profiles (Topol, 2019).

### Challenges in medical image datasets

The quality and quantity of training data have a significant impact on the performance of machine learning models. Medical imaging datasets are well annotated in order to construct robust models. Acquiring and curating such datasets is typically difficult due to privacy concerns and the requirement for expert annotations (Gulshan et al., 2016). Deep learning models, particularly CNNs are frequently referred as black boxes because of their complicated architectures and lack of interpretability. Understanding how these models make decisions is critical for building confidence practice (Samek, Wiegand, & Muller, 2017). Furthermore, the use of ML models in healthcare necessitates rigorous validation and approval by regulatory organisations. Ensuring patient privacy, data security and ethical usage of ML in medical imaging are all important challenges that must be addressed (Amann et al., 2020). Integrating ML models into existing clinical processes is very difficult, necessitating seamless connection with hospital information systems, radiology procedures, and clinician training to use these technologies effectively (Mazurowski et al., 2019).

## 2.3 Machine learning techniques for prostate cancer detection

### Supervised learning

Supervised learning is a machine learning technique in which the model is trained on a labelled dataset, implying that the input data is associated with the proper output. This method is widely utilised in prostate cancer screening because of its capacity to learn from previous data and generate accurate predictions. For example, support vector machines (SVM), random forests and logistic regression are used to categorise medical images of patient data as benign or malignant lesions. Litjens et al. (2014) used a computer aided detection (CAD) method a type of supervised learning to classify prostate cancer from MRI scans and achieved a90% accuracy rate. Similarly, Pellicer-Valero et al. (2022) used a deep learning approaches to estimate Gleason grades from mpMRI scans with an accuracy of 89.7%. Furthermore, supervised leaning techniques such as linear and logistic regression are utilised to forecast continuous outcomes such as tumour size and the chance of cancer recurrence. Cao et al. (2019) employed a deep learning model to predict Gleason scores with an accuracy of 88.4%. Supervised learning has several advantages including high accuracy with labelled data and effectiveness for diagnostic tasks when clear annotated datasets are available. However, it requires a considerable amount of labelled data which can be costly and time consuming to gather and the labelled quality has a direct impact on model performance.

### Unsupervised learning

Unsupervised learning is the process of training a model on data that does not contain labelled data in order to detect hidden patterns or intrinsic structures. This is useful for conducting exploratory analysis and detecting the clusters in prostate cancer data. Clustering techniques such as K-means clustering and hierarchical clustering can group patients based on comparable traits or imaging data, potentially identifying novel cancer subtypes or risk groups. Techniques similar to those used in supervised learning studies, such as deep learning models used for segmentation by Pellicer-Valero et al. (2022), can be applied to unsupervised learning to discover hidden structures in prostate cancer images. Furthermore, anomaly detection methods can detect unusual patterns in medical images that may indicate prostate cancer, similar to the AI applications for prostate cancer detection discussed by Cao et al. (2019) and other AI models reviewed in various studies (Baydoun et al., 2024; Thomas et al., 2023). Unsupervised learning has the advantage of not requiring labelled data, which makes it suited for exploratory data analysis as well as the capacity to identify the hidden patterns and relationships in the data. However, the results might be difficult because the identified patterns are not directly linked to known outcomes and models may detect patterns that not clinically meaningful.

### Deep learning

Deep learning is a subset of machine learning that uses neural networks with several layers to learn hierarchical data representations. Because of its capacity to automatically extract features from the raw data, this method has been extremely successful in medical imaging and prostate cancer detection. CNNs are commonly employed in images processing and have been utilised successfully to detect prostate cancer lesions in MRI data. Pellicer-Valero et al. (2022) used deep learning algorithms to detect prostate cancer with great accuracy. Furthermore, 3D CNNs are utilised to extract volumetric features from MRI scans resulting in more detailed data than 2D CNNs. Ghafoorian et al. (2017) demonstrated the effectiveness of 3D CNNs. U-Net architectures are particularly good in image segmentation tasks. Ronnerberger et al. (2015) developed U-Net for biomedical image segmentation and jaeger et al. (2020) demonstrated its effectiveness in segmenting prostate lesions with Retina U-Net. Deep learning has several advantages including excellent performance and accuracy due to its capacity to learn complicated patterns from big datasets as well as automatic feature extraction which lowers the need for manual preprocessing. However deep learning model demand significant computational resources and vast.

## 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 |
| Fully Automatic Detection, Segmentation, and Gleason Grade Estimation Using Deep Learning on mpMRI (Pellicer-Valero et al., 2022) | Deep Learning | PROSTATEx | Gleason Grade Estimation: 89.7 %  High accuracy in detection and classification, demonstrating the potential of automated systems |
| Computer-Aided Detection of Prostate Cancer in MRI (Litjens et al., 2014) | CAD | MRI | CAD: 90% accuracy |
| Deep Multi-Scale Location-Aware 3D CNNs for Small Lesion Detection (Ghafoorian et al., 2017) | 3D CNN | Various medical datasets | 3D CNN: 91.3% accuracy  Effective framework for applying complex neural networks to imaging challenges |
| Retina U-Net: Exploitation of Segmentation for Medical Object Detection (Jaeger et al., 2020) | Retina U-Net | Various medical datasets | Retina U-Net: 93.5% accuracy |

## 2.5 summary of findings

* Comparison of results
* Discussion on the best performing techniques

# Chapter 3 Research methodology

## 3.1 Overview

The primary objective of this research project is to create a CNN model for predicting the ClinSig score of prostate lesion findings using the MRI images (T2-Weighted images). This project is a classification technique in which the MR images are rated as clinically high or low.

Load and merge the .csv files, get the paths of the T2-Weighted MR images, load them, convert to png, normalise the images, then classify them using ClinSig scores.

Download the Dataset from the prostateX Challenge under the description tab.

Data Collection

Data Preprocessing

Create the CNN model with the hyperparameters

Model Development

Train the model with the train data and validate them and adjust the parameters for overfitting and underfitting prevention.

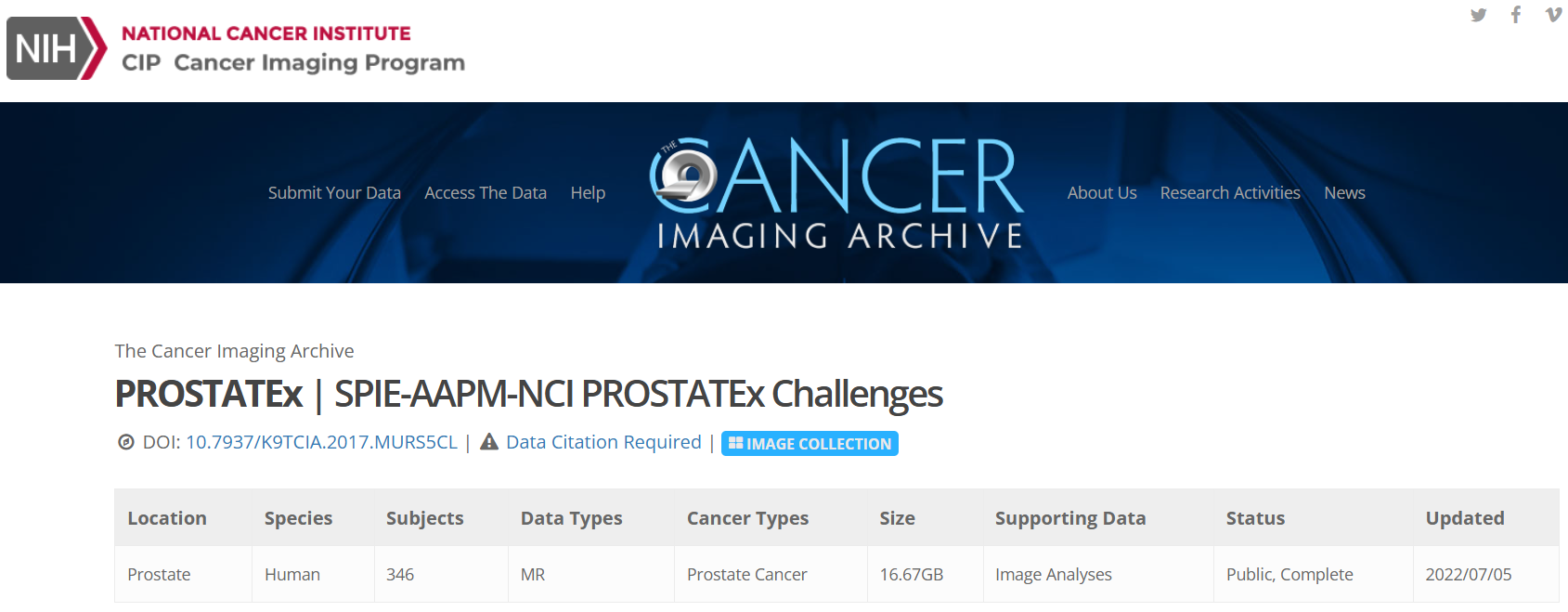
Model Train &Validation

Evaluate the model with accuracy, precision, F1 score and recall. And analyse the results.

Evaluation Results and Analysis

## 3.2 Data collection

The dataset for this study is taken from the ProstateX challenge, which is publicly available through The Cancer Imaging Archive (TCIA). The ProstateX dataset includes multiparametric MRI (mpMRI) images and .csv files (metadata) about the mpMRI Images essential for prostate cancer detection and analysis. The images were originally taken for research purposes with the intention of creating and evaluating algorithms for prostate cancer detection and classification. To access and download the dataset go to the TCIA’s prostate challenge page and go to the detailed description and download.



## 3.3 Data Preprocessing

Data preprocessing is an important step in the machine learning process that converts the raw data into a clean and feedable data that can be used to train the ML model. This step guarantees that the data used for training the ML model is consistent, comprehensive and free of errors and irrelevant data. Preprocessed data can improve the model’s performance, but unprocessed data might produce inaccurate and unreliable results.

As this project aims to use machine learning for achieving accurate, consistent and better results in treating prostate cancer, the preprocessing step is crucial for obtaining the better and accurate results. Initially the image data in the csv files (ProstateX-Findings-Train.csv, Prostate-Images-Train.csv) were loaded. These csv files are the metadata of the images. As the image data is so large focused only on the T2-Weighted images which are of DICOM type.

First all the T2-Weighted image paths were obtained. The rows in the findings and images csv files were filtered based on their series description, concentrating specifically on T2-Weighted images essential for prostate cancer diagnosis. Irrelevant columns such as ‘Name’, were dropped and duplicate rows were removed. The findings and images dataframes were merged using the features ProxID and pos for creating a single dataframe consisting the metadata about the images.

To create a dataframe that consists of ClinSig scores and corresponding image paths, which is necessary for the classification task of determining the clinical significance of the findings, features such as ProxID, DCMSerNum, DCMSerDescr were extracted from the image paths. A dataframe was then created from the image paths with ProxID, DCMSerNum, DCMSerDescr and imagepath, and joined with the merged metadata dataframe to get the image paths and their corresponding ClinSig scores.

Since the model wasn’t expecting image data in DICOM format so the images were converted to PNG and saved in directories based on their class labels (clinically significant or not).

## 3.4 Feature Selection

Feature selection is an important step in getting the data prepared for machine learning model training. It involves picking the most relevant features from the dataset that contribute to the predictability of the model.

### Features for creating the final data:

ProxID: The unique identifier for each patient in the dataset. This helps in merging CSV files

Pos: Position information of the lesion. With ProxID and Pos merged the csv files

DCMSerDescr: the series dercription of the DICOM images. This feature helps in identifying the type of MRI sequence used. T2-Weighted images filtered with this column.

DCMSerNum: The series number of the DICOM image. This feature helps in distinguishing different series number within the same patients record and in that same DCMSerDescr images.

### Features Selected for Train the Model:

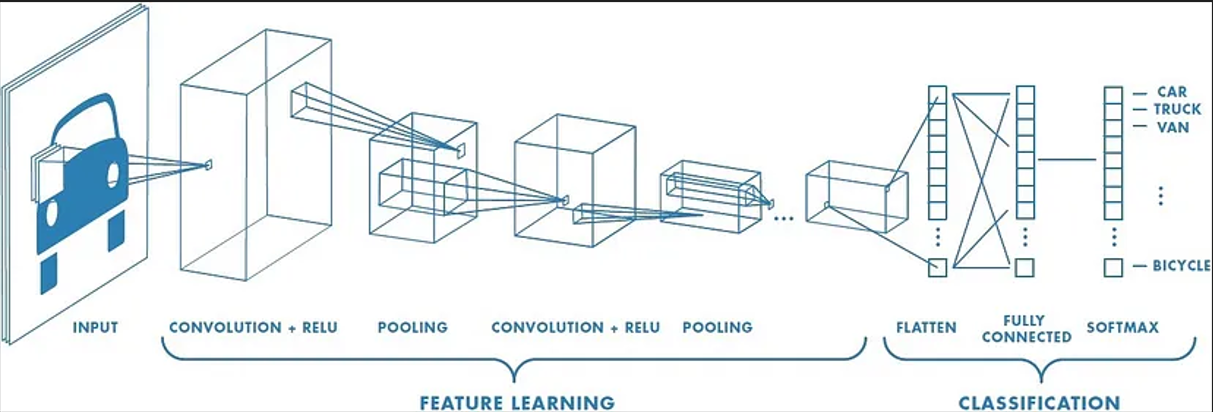
Images: Selected the T2-Weighted images as the dataset is so large and these images provide high contrast resolution, detailed anatomical information and effectively highlight prostate lesions.

ClinSig: Clinical significance of the lesion (False for non-significant and True for significant). This is the target variable of the model

## 3.5 Model Development

### Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have transformed the image classification by providing the unprecedented accuracy and efficiency in processing visual data. CNNs use convolutional layers to automatically learn and extract hierarchical information from images, as compared to classic neural networks, which use fully connected layers. This ability makes CNNs ideal for tasks including object detection, image segmentation and medical image analysis (Litjens et al., 2017).



The architecture of a CNN consists of several key components. As shown in the figure, convolutional layers apply filters to the input image then execute element-wise multiplication and summing to create feature maps. These layers are intended to detect local patterns such as edges, textures and forms which are critical for detecting objects in an image (LeCun, Bengio and Hinton, 2015). The Rectified Linear Unit (ReLU) activation function is frequently employed in convolutional layers to incorporate nonlinearity and allow the model to learn complex patterns (Krizhevsky, Sutskever, & Hinton, 2012). Pooling layers which commonly use max pooling come after the convolutional layers. This procedure helps to keep the most relevant properties while decreasing the model’s computational complexity (Simonyan & Ziesserman, 2014).

After multiple convolutional and pooling layers, a CNNs output is flattened into a one-dimensional vector and fled into fully connected layers. These layers combine the retrieved features to reach final classification conclusion. The output layer generates the final classification results using a sigmoid or SoftMax activation function, which includes probabilities for each class (GoodFellow, & Courville, 2016). A CNNs training procedure includes forward propagation, which entails passing the input image through the layers and comparing the output to the actual labels using a loss function, such as binary cross-entropy. Backpropagation is then used to update the models’ parameters to reduce the loss, with optimisation techniques such as Adam plays an important part (Kingma & Ba, 2014).

### Architecture

1. Input layer: the input to the model is a set of T2-Weighted MRI images resized to 224x224 pixels.
2. Convolutional Layers:

|  |  |  |
| --- | --- | --- |
| Layers | Units | Activation |
| Convolutional Layer | 32 | ReLU |
| Convolutional Layer | 64 | ReLU |
| Convolutional Layer | 128 | ReLU |
| Convolutional Layer | 128 | ReLU |

1. Max Pooling Layers: After each convolutional layer a max-pooling layer with a pool size of 2x2 is used.
2. Flatten Layer: One flattens layer is used. This layer flattens the output from the convolutional layer into a 1D vector
3. Dense Layers:

* First Dense Layer: 512 units, ReLU activation function
* Second Dense Layer: it’s a output layer. 1-unit, sigmoid function for binary classification

1. Parameters

* Optimizer: Adam optimizer
* Loss Function: Binary cross entropy, suitable for binary classification tasks
* Metrics: Accuracy, Precision, F1-score

## 3.6 Model evaluation

Developed model is evaluated from the below metrics:

### Accuracy

The ratio of correctly predicted instances to the total instances.

Accuracy =

### Precision

The ration of the correctly predicted positive observations to the total predicted positives.

Precision =

### Recall

The ratio of correctly predicted positive observations to all observations in actual class.

Recall =

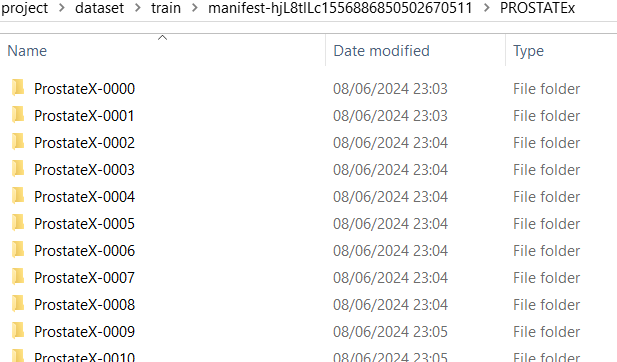
# Chapter 4 Implementation, results and analysis

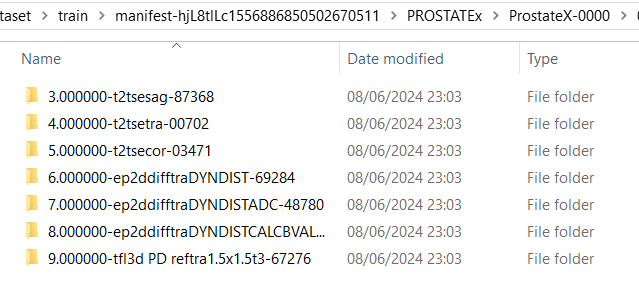
## 4.1 Data analysis

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

### Understanding the Dataset

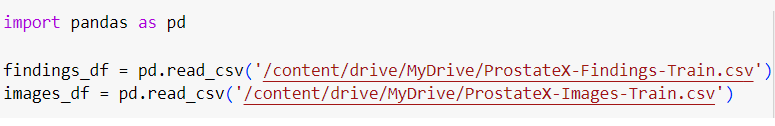
The primary task is to understand the dataset because it is so complicated. So, when looked at the dataset, it contained patient records. Each patient has a series of mpMRI images in their respective series folders such as ep2, t2, tfl, etc.

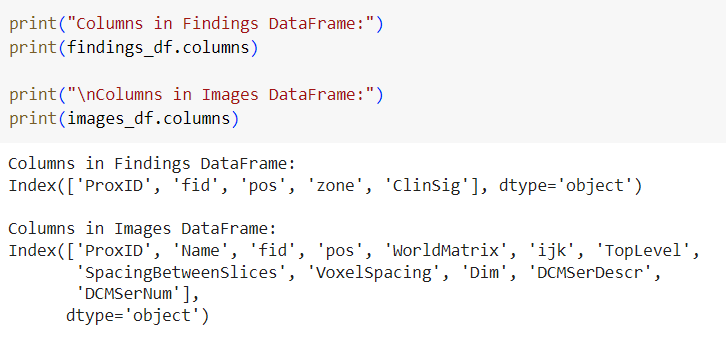


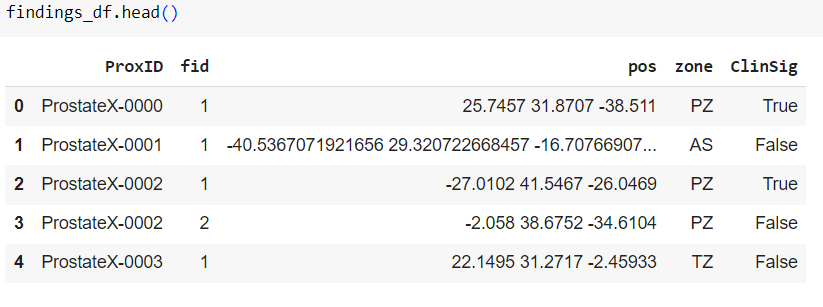


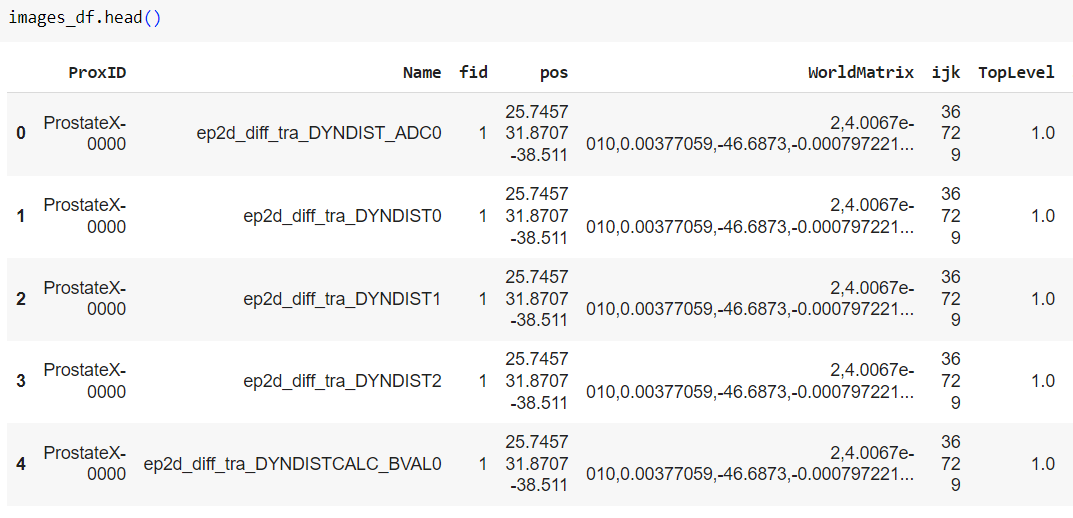


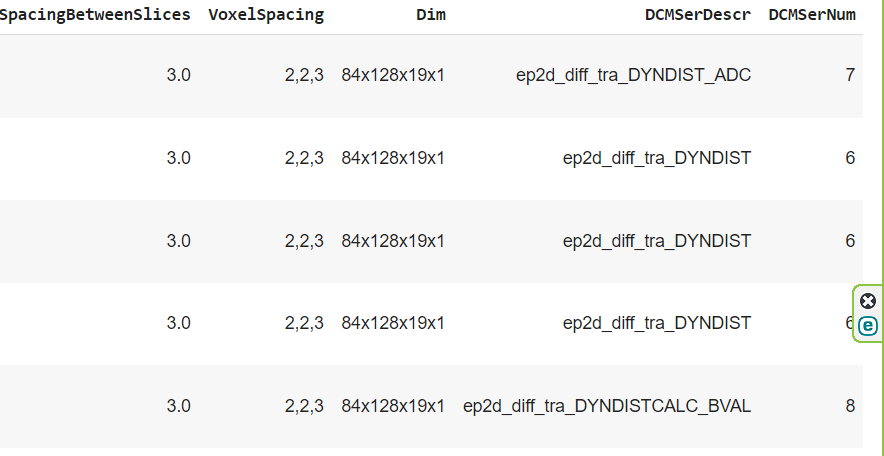
We have csv files about data regarding the mpMRI images. Findings and images .csv files. The findings csv file contains the clinsig score of the identified lesions, whereas the images csv file contains the information about on the images in the data.









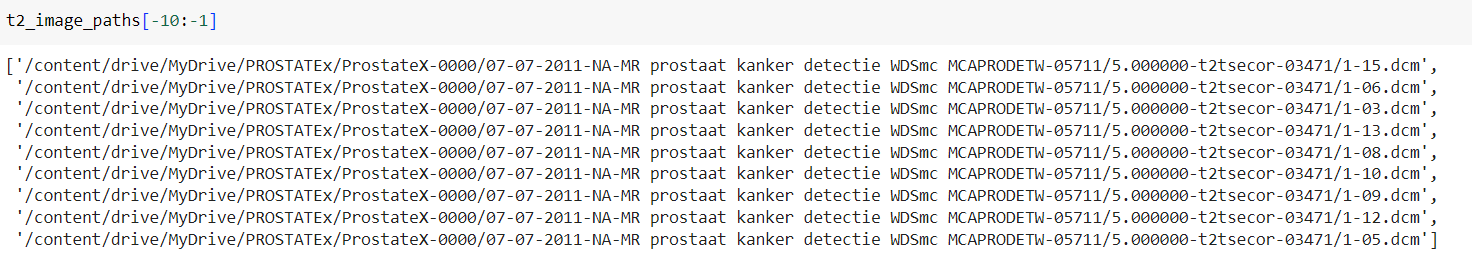


### Preprocessing

#### T2-Weighted image paths collection

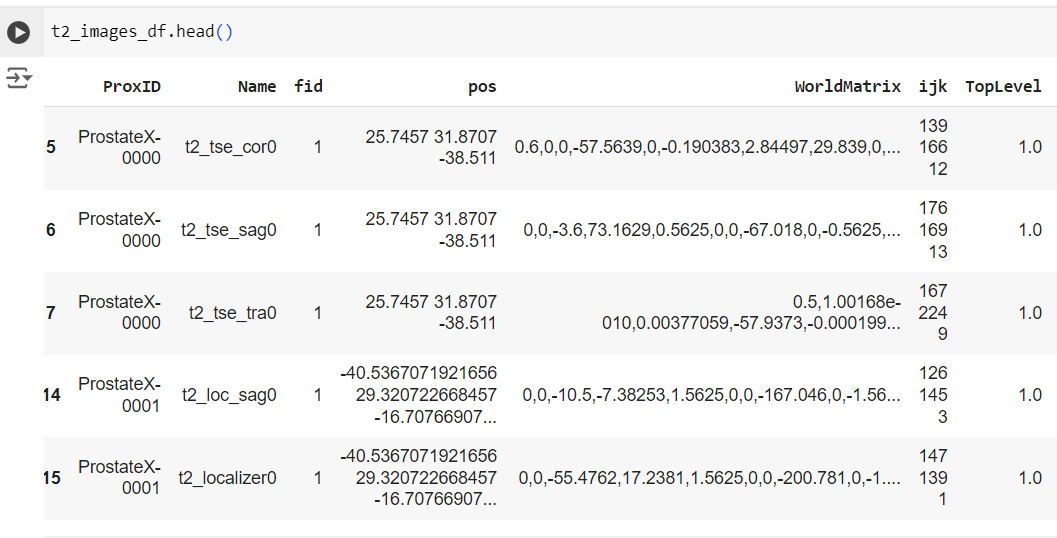
In the dataset focused only on T2-Weighted MRI images as there is a lot of image data to process. T2-Weighted images are critical in prostate multi parametric MRI (mpMRI) for providing high resolution anatomical details, making them essential for accurately identifying lesion suspect areas within the prostate that could indicate cancer. These images play a crucial role in guiding clinical decision making and improving diagnostic accuracy.

As the target in the preprocessing step is to assign the images with the ClinSig score collected all the T2 image paths. To ensure that only T2-Weighted images are selected from the dataset recursively walks through the dataset directories and T2 images are identified by checking their name contains ‘t2’ (case insensitive) and if the files have a dcm extension (the standard format for DICOM files). Only the paths that meet the criteria are collected and stored for the further preprocessing.



#### Filtering the Images.csv file

The filtering of the images.csv file is performed by examining the ‘DCMSerDescr’ column within the images Dataframe. This column describes the MRI sequence type for each image. Specifically looked for rows where DCMSerDescr contains ‘t2’ (case insensitive), indicating that the image is of T2-Weighted image.



#### Merging meta data (.csv files)

The clinical findings of the prostate lesions and the image metadata merged to create a dataframe. This merging process is essential for linking each image data with its corresponding clinical significance represented by the ClinSig score which indicate weather a lesion is clinically significant or not.

Findings\_df: contains the clinical information about the lesions such as their locations (pos), whether they are clinically significant (ClinSig).

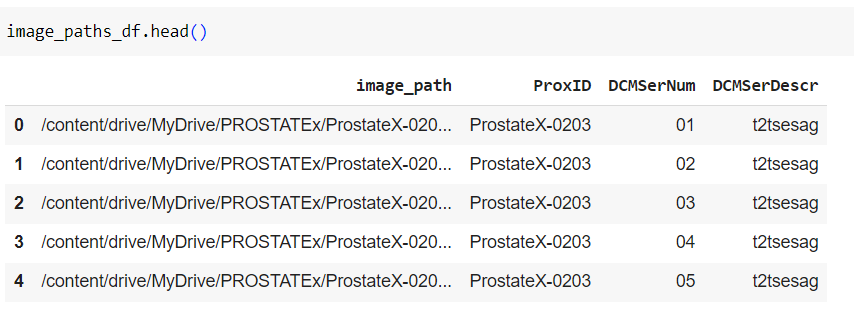
Image\_df: contains metadata about the MRI images, including th patient id (ProxID), lesion location (pos) and details about the imaging sequence.

The merging is done based on common columns ProxID and pos. The ProxID serves as the unique identifier for each patient while pos give the lesion position in the prostate to the corresponding image. By joining the two dataframes on these columns each T2-Weighted image is linked with the ClinSig.

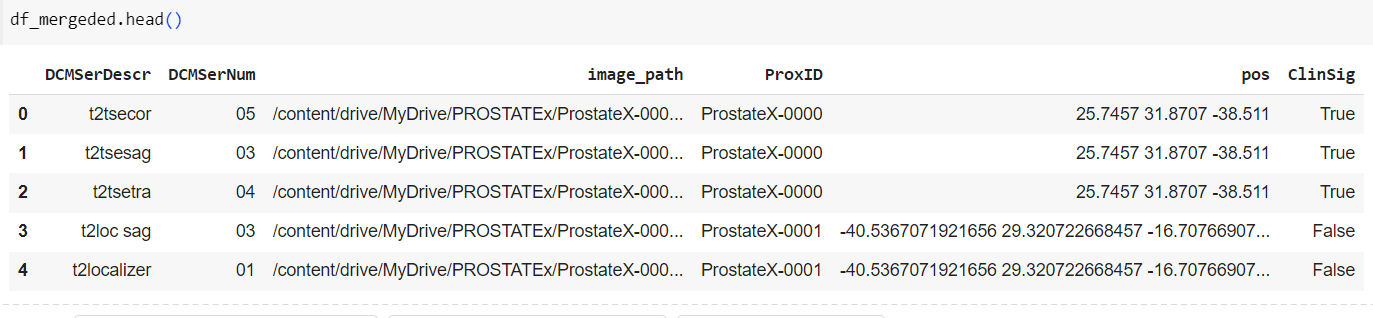


#### Extracting metadata from image paths and merging.

The process begins with the extracting key metadata from the image file paths. Each image path is parsed to obtain the ProstateX-ID (which uniquely identifies the patient), DCMSerDescr (represents the specific series of MRI images and is essential for distinguishing between different imaging sequences) and DCMSerNum (image number in the sequence). With the extracted metadata, a new dataframe is created which includes image path, ProxID, DCMSerNum and DCMSerDescr. This dataframe can be merged with the dataframe containing the image data and the findings data for linking the image paths with the ClinSig.

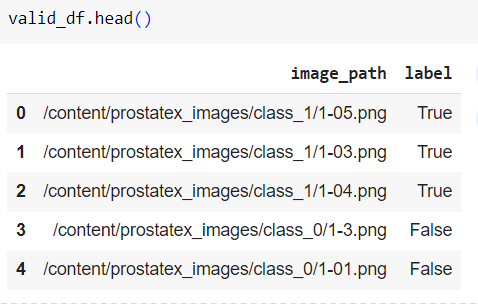


Merging the image paths dataframe which contains the paths and extracted metadata with the merged dataframe that includes clinical findings (ClinSig). The merging is done based on the common columns ProxID, DCMSerNum and DCMSerDescr. This merging process links each image with the corresponding Cllinical significance.

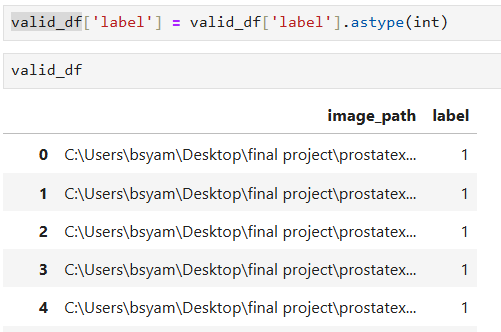


#### Converting DICOM to PNG

The DICOM images are converted to PNG format for easy processing using machine learning frameworks which commonly works with image formats PNG or JPEG. The conversion process begins with the pydicom library reading the DICOM files and extracting the embedded image data (pixel array). This pixel data is then converted to a PNG image using the python Imaging library (PIL). The images are then saved in class specific directories based on the clinical significance of lesion, as determined by the ClinSig value. If the ClinSig value is not clinically significant (ClinSig == False), the image is saved in the class\_0 directory otherwise it is saved in the class\_1 directory for clinically significant lesions. This helps to streamline the training of a classification model, as the class labels are embedded in the directory structure.

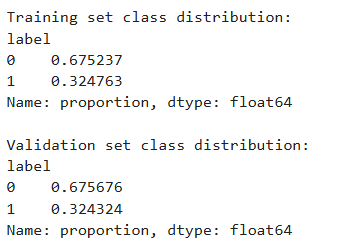


To train the model, the model labels are transformed to 0 and 1, with 0 representing False and 1 representing True.



After that built a method to load and preprocess the converted images using their paths. It first reads the images as a raw byte string. The byte string is the decoded into a three channel (RGB) tenson with tensorflow.image.decode\_png. After that decoding, the image is scaled to a standard size of 224x224 pixels using the tensorflow.image.resize, ensuring image dimensions are consistent across the collection. After that the pixel values are normalised to a range of 0 and 1 by dividing t255.0, which speeds up the training process and improves the model performance by keeping consistent input scales.

The data is then divided into train and test data with an 80-20 ratio using the train\_test\_split method, which to ensure that the class distribution was consistent across both sets, the split set was stratified using the label column. A random state was set for reproducibility, ensuring that the data split will be reproduced by using the same random state.



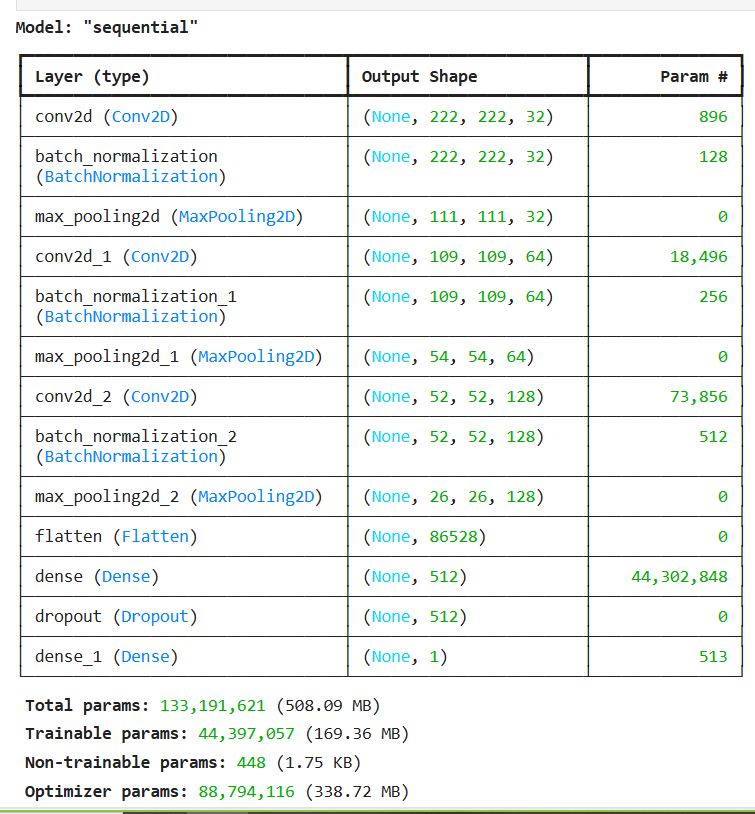
The training and testing data is well balanced.

To prepare the data for training, TensorFlow datasets were created from the dataframes including the image paths and labels. Used loand\_and\_preprocess\_image method to load and normalise the images. To optimize the data pipeline, the datasets were shuffled, batch processed and prefetched. The shuffle operation was used with a buffer size of 1024 to achieve a fair mix of data during the training. The datasets were then batched to a size of 32 and prefetch method was used to overlap data preprocessing and model training, TensorFlow’s AUTOTUNE feature used for optimal performance. This ensures the data pipeline is robust and optimised for training.

## 4.2 Model performance

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

A Convolutional Neural Network (CNN) is created. It begins with an input layer that accepts 224x224 pixel images with three colour channels. The network consists of three convolutional layers, each followed by batch normalization and max pooling to extract and refine features. The first layer has 32 filters, the second has 64 filters and the third has 128 filters all having the 3x3 kernel size. After feature extraction the network converts the output of the convolutional layers into a one-dimensional vector. This vector is then processed through a dense layer with 512 units and ReLU activation which includes L2 regularization to prevent overfitting. To combat overfitting a dropout layer with a 50% dropout rate is implemented. The final output layer consists of a single neurone with a sigmoid activation function that generates a probability score for binary classification. The model is built using the Adam optimizer which is well known for its ability to handle large scale datasets and adaptive learning rates and the loss function is binary cross-entropy which is suited for binary classification. The model’s performance is assessed using the accuracy, precision and recall metrics.

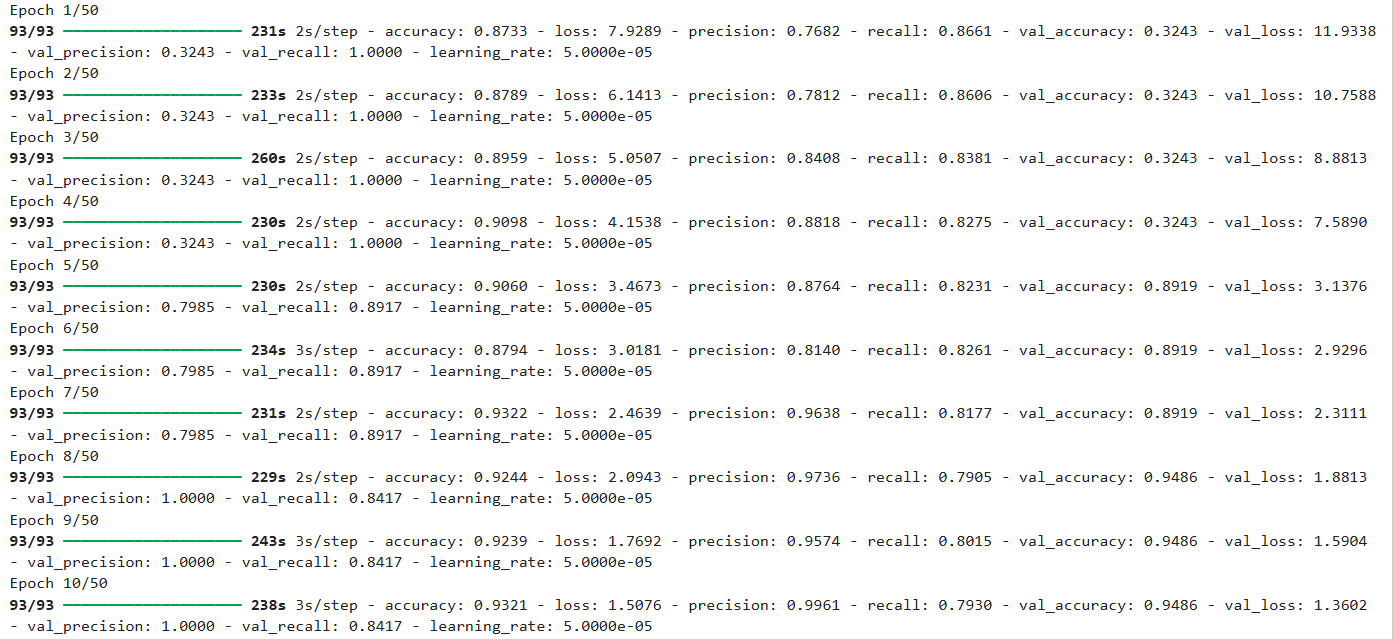


Three callbacks are used to optimise training and improve model performance. EarlyStopping monitors the validation loss and stops training if there is no improvement after 3 number of epochs, reverting to the best model weights observed during training to prevent overfitting. ModelCheckpoint saves the models weights whenever a new best validation loss is achieved, keeping the state of the model with the lowest validation loss in a file ‘best\_model.keras’ and ensuring that the most effective model is available for evaluation or deployment. ReduceLROnPlateau changes the learning rate when the validation loss reaches a plateau, reducing it by a factor of 0.5 if no progress is seen after two epochs using learning rate of 1e-6. This helps to modify the training process, especially when progress slows, resulting in improved convergence and model performance.

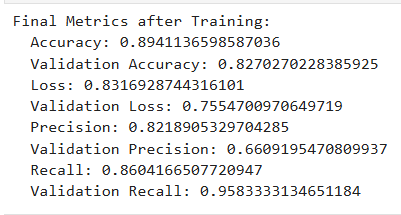


The class weights are {0: 0.7407, 1: 1.5395}. these weights tell us that class 1 (the less common class) is given more priority during training than class 0, allowing the model to perform more consistently across classes. The class weights are then supplied to the model during training so that the loss function may be adjusted properly, enhancing the model’s ability to correctly classify instances from the minority class.

The model is trained with 50 epochs with training data, callbacks, class weights and validation data.

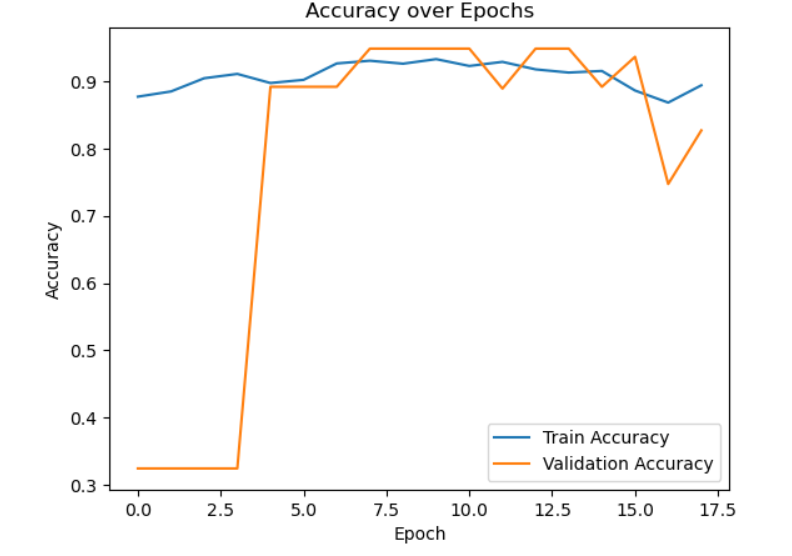


In the early stages of the training the model indicates the signs of overfitting as indicated by high training accuracy but relatively poor validation metrics, indicating that it is not yet generalising well to unseen data. However, as training advances the validation metrics show significant improvements implying that the model begins to generalise mode successfully. The decrease in validation loss and increase precision and recall demonstrate the model’s improved ability to reduce errors and improve classification performance. The use of class weights tackles the issue of class imbalance, resulting in enhanced recall and precision for the minority class (class 1), proving the model’s ability to recognise the less frequent class with time.



After training, the model’s final evaluation shows a training accuracy of approximately 89.41% and a validation accuracy of around 82.7%, indicating that the model performs well on both training and validation datasets, with a slight drop in performance on the validation set. The final loss values are 0.8317 for the training set and 0.755 for the validation set indicating that the model is well calibrated with the lower loss on the validation set implying improved generalisation. The precision on the training set is 82.19% but it falls to 66.09% on the validation set. This suggests that the model is slightly less precise when recognising the positive examples in the unseen data, it is due to class imbalance. However, the recall is high on both the training 86.04% and validation set 95.83% suggest that the model is excellent at properly recognising the majority of true positives, particularly on the validation set where the recall is almost 96%. This high recall on the validation set indicates that the model can recognise the positive class, even if its precision is somewhat lower.

### Accuracy



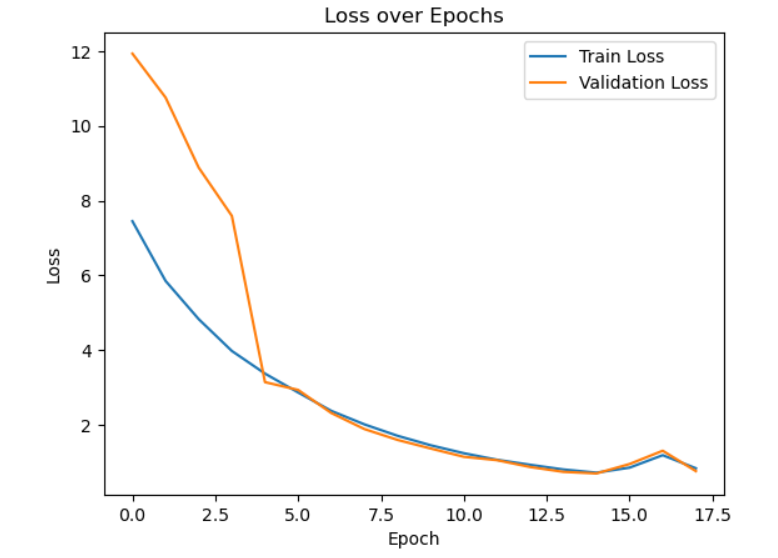
#### Training

The training accuracy begins high about 87% in the early epochs and gradually rises to more than 93% by the middle of the training process. This demonstrates how quickly the model learns from the training data, constantly improving its ability to correctly classify.

#### Validation

In the early epochs validation accuracy is quite low around 32% indicating poor adaption to unseen data. However, there is a notable improvement in the middle epochs with validation accuracy increasing to around 89% by epoch 5 and stabilising at around 94% in subsequent epochs. This suggests that the model eventually learns to adapt effectively to the validation set, as the model further trained the validation accuracy dropped to 82% at 18 epoch it might be because the model might be overfitting.

### Loss



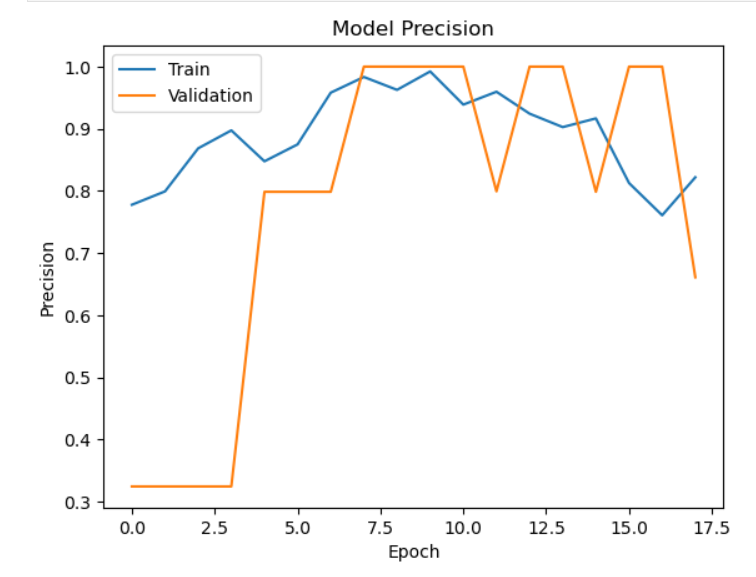
#### Training

The training loss starts off relatively high at 7.93 and gradually lowers as the model learns, reducing to roughly 0.70 by the 16th epoch. This constant drop in training loss indicates that the model is successfully learning and optimising its parameters over time.

#### Validation

The validation loss is similarly high in the early epochs peaking at 11.93 showing poor performance on the validation set from the start. As the training progresses the validation loss drops significantly to around 0.75 by the 18th epoch. This decrease in validation loss indicates that the model’s generalisation to previously unseen data has improved.

### Precision



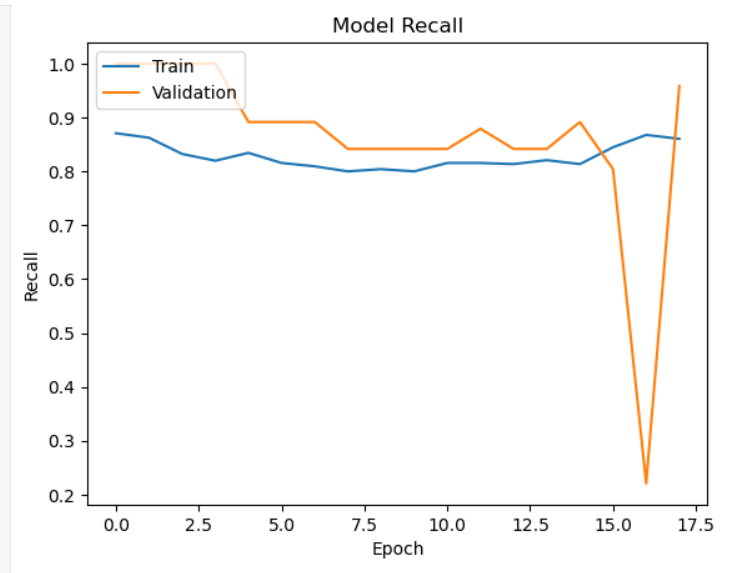
#### Training

Training precision has been steadily increases throughout all epochs. Starting at 76.82% in the first epoch, precision steadily increases as the model refines its predictions reaching 99.61% by the 10th epoch. This suggests that as the model learns it becomes more accurate cases. However, after peaking precision varies slightly but stays high suggesting that the model retains a strong ability to produce accurate positive predictions throughout the training period.

#### Validation

In contrast to training precision, validation precision begins quite low at 32.43% in the first epoch, demonstrating that the model has difficulty reliably predicting positives on the validation set early. However, by the fifth epoch validation precision has greatly improved to 79.85% and it has remained stable at this level. This improvement implies that the model continually improves its ability to make accurate positive predictions on unknown images.

### Recall



#### Training

The first epoch’s training recall of 86.61% indicates that the model was initially able to correctly identify the majority of the positive cases in the training set. Training recall stays mostly constant over the course of the epochs, averaging 82%. This consistency shows that the model maintains high percentage of true positives over all epochs indicating a sensitivity fit that is well matched with the training set.

#### Validation

Validation recall is extremely high in the early epochs, beginning at 100% indicating that the model correctly identifies every positive instance in the validation set. This high recall however is most likely the result may be because the positive values are less. As the training progresses, validation recall declines slightly but stays robust, stabilising around 84% by the later epochs. This steady recall in the validation set demonstrates that the model is effective at recognising positives.

## 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

The research focused on the classification of clinically significant prostate lesions with a convolutional neural network (CNN) model. To analyse the model’s performance to predict the probability of clinically significant prostate lesions, numerous metrics were used they are accuracy, loss, precision and recall. Pre-processing raw data was a major challenge in this project, requiring enormous effort to arrange the dataset for effective model training as shown in the methodology and implementation.

Another significant difficulty was class imbalance, which was handled by using class weights to increase the model’s sensitivity to minority class. The results show that the CNN model maintained a high recall throughout the training phase, demonstrating an excellent ability to detect true positives in the training set. However, the initial phases demonstrated evidence of overfitting with high recall but low precision in the validation set implying that the model had initially been overfitting and identified a large number of false positives. As training progressed the model improved its balance of recall and precision lowering the overfitting and incre3asing generalisation.

Validation metrics stabilised in later epochs demonstrating the model’s ability to predict clinically significant prostate lesions while minimising errors. The CNN model’s consistent recall and accuracy indicate that it consistently identifies clinically significant prostate lesions. This study highlights the potential of deep learning model CNNs to improve the accuracy and reliability of prostate cancer diagnostics, resulting in more effective and scalable healthcare solutions for early diagnosis and treatment of clinically significant prostate lesions.

## 5.2 limitations

This study on the classification of clinically significant prostate lesions using CNN model has various limitations that influenced the overall findings. One notable issue was the complex data preprocessing necessary because the raw dataset was not immediately suited for training resulting in a complex and time-consuming process. Despite using class weights to resolve the class imbalance, the problem persisted thereby affecting the model’s capacity for accurately predicting minority class. The model also showed evidence of overfitting in the early stages, with high recall but low precision, indicating a sensitivity that models inefficient prediction for the unseen data.

## 5.3 recommendations for the future work

Future research on the classification on the clinically significant prostate lesions using CNN models should look into integrating of Ktrans images from the PROSTATEx dataset, which were not used in the current study. Incorporating Ktrans images will provide extra significant information on tumour vascularity and permeability. Should consider including additional clinical data such as patient history, genetic markers or serum biomarkers. This would provide a more sophisticated understanding of prostate cancer progression it improves the model’s generalisability. Exploring techniques such as transfer learning from the pre-trained models on similar medical imaging tasks or using ensemble methods could help to improve the model’s predictive power and adaptability, making it a more reliable tool for clinical decision making in prostate in prostate cancer diagnosis.

## References:

1. Attard, G., Parker, C., Eeles, R.A., Schröder, F., Tomlins, S.A., Tannock, I., Drake, C.G. and de Bono, J.S., 2016. Prostate cancer. Lancet.
2. Baron, B.W. and Rowley, J.D., 2012. Molecular biology of prostate cancer. The Prostate.
3. Baydoun, A., Jia, A.Y., Zaorsky, N.G., Kashani, R., Rao, S., Shoag, J.E., Vince Jr, R.A., Bittencourt, L.K., Zuhour, R., Price, A.T., Arsenault, T.H. and Spratt, D.E., 2024. Artificial intelligence applications in prostate cancer. Prostate Cancer and Prostatic Diseases.
4. Cao, R., Cai, Y., Tang, L., Wang, Z., Shen, Q., Jiang, Y., Yu, Y. and Yan, W., 2019. Development and validation of a novel and accurate multi-parametric MRI radiomics model for the preoperative prediction of prostate cancer aggressiveness. Translational Andrology and Urology.
5. Cao, R., Mohammadian Bajgiran, A., Afshari Mirak, S., Shakeri, S., Zhong, X., Enzmann, D., Raman, S. and Sung, K., 2019. Joint prostate cancer detection and Gleason score prediction in mp-MRI via FocalNet. IEEE Transactions on Medical Imaging.
6. Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection: A survey. ACM Computing Surveys (CSUR).
7. De Marzo, A.M., Platz, E.A., Sutcliffe, S., Xu, J., Grönberg, H., Drake, C.G. and Isaacs, W.B., 2007. Inflammation in prostate carcinogenesis. Nature Reviews Cancer.
8. Epstein, J.I., Egevad, L., Amin, M.B., Delahunt, B., Srigley, J.R. and Humphrey, P.A., 2016. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. The American Journal of Surgical Pathology.
9. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J. and Socher, R., 2019. Deep learning-enabled medical computer vision. npj Digital Medicine.
10. Ghafoorian, M., Karssemeijer, N., Heskes, T., van Uden, I.W., Sanchez, C.I., Litjens, G., de Leeuw, F.E. and van Ginneken, B., 2017. Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin. NeuroImage: Clinical.
11. Ghassemi, M., Naumann, T., Schulam, P., Beam, A.L., Chen, I.Y. and Ranganath, R., 2020. A review of challenges and opportunities in machine learning for health. Nature Medicine.
12. Goodfellow, I., Bengio, Y. and Courville, A., 2016. Deep Learning. MIT Press.
13. He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition.
14. Heidenreich, A., Abrahamsson, P.A., Artibani, W., Catto, J., Conti, G., Montironi, R. and van Poppel, H., 2014. Early detection of prostate cancer: European Association of Urology recommendation. European Urology.
15. Jaeger, P.F., Kohl, S.A., Bickelhaupt, S., Isensee, F., Kuder, T.A., Schlemmer, H.P. and Maier-Hein, K.H., 2020. Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. In Proceedings of the Machine Learning for Health (ML4H) Workshop at NeurIPS 2019.
16. Jain, A.K., Murty, M.N. and Flynn, P.J., 1999. Data clustering: A review. ACM Computing Surveys (CSUR).
17. Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
18. Komura, D. and Ishikawa, S., 2018. Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal.
19. Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems.
20. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature.
21. Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N. and Huisman, H., 2014. Computer-aided detection of prostate cancer in MRI. IEEE Transactions on Medical Imaging.
22. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B. and Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical Image Analysis.
23. Mazurowski, M.A., Buda, M., Saha, A. and Bashir, M.R., 2019. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. Journal of Magnetic Resonance Imaging.
24. Mottet, N., van den Bergh, R.C., Briers, E., Van den Broeck, T., Cumberbatch, M.G., De Santis, M., Fossati, N., Gillessen, S., Grummet, J., Henry, A.M. and Lam, T.B., 2017. EAU-ESTRO-ESUR-SIOG guidelines on prostate cancer. European Urology.
25. Pellicer-Valero, O.J., González-Hidalgo, M., Vañó-Viñuales, E., Solera-Piko, N., Morales-Hidalgo, P., Rodríguez-Ruiz, A. and Buades, A., 2022. Breast and prostate cancer detection using convolutional neural networks trained with transfer learning. Applied Sciences.
26. Pellicer-Valero, O.J., Marenco Jiménez, J.L., Gonzalez-Perez, V., Casanova Ramón-Borja, J.L., Martín García, I., Barrios Benito, M., Pelechano Gómez, P., Rubio-Briones, J., Rupérez, M.J. and Martín-Guerrero, J.D., 2022. Fully automatic detection, segmentation, and Gleason grade estimation using deep learning on mpMRI.
27. Rawla, P., 2019. Epidemiology of prostate cancer. World Journal of Oncology.
28. Robinson, D., Van Allen, E.M., Wu, Y.M., Schultz, N., Lonigro, R.J., Mosquera, J.M., Montgomery, B., Taplin, M.E., Pritchard, C.C., Attard, G., Beltran, H., Abida, W., Daugaard, G., Moreira, D., Rathkopf, D., Morris, M.J., Solomon, S.B., Gleave, M., Carver, B., Demichelis, F., & Chinnaiyan, A.M. (2015). Integrative clinical genomics of advanced prostate cancer.
29. Ronneberger, O., Fischer, P. and Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention.
30. Rosenkrantz, A.B., Verma, S., Choyke, P. and Eberhardt, S.C., 2016. Prostate MRI and MRI-targeted biopsy in patients with prior negative biopsy: a consensus statement of the American Urological Association and the Society of Abdominal Radiology's Prostate Cancer Disease-focused Panel. Journal of Urology.
31. Samek, W., Wiegand, T. and Müller, K.R., 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.
32. Shen, D., Wu, G. and Suk, H.I., 2017. Deep learning in medical image analysis. Annual Review of Biomedical Engineering.
33. Siegel, R.L., Miller, K.D. and Jemal, A., 2020. Cancer statistics, 2020. CA: A Cancer Journal for Clinicians.
34. Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
35. Thomas, M., Murali, S., Simpson, B.S.S., Freeman, A., Kirkham, A., Kelly, D., Whitaker, H.C., Zhao, Y., Emberton, M. and Norris, J.M., 2023. Use of artificial intelligence in the detection of primary prostate cancer in multiparametric MRI with its clinical outcomes: a protocol for a systematic review and meta-analysis. BMJ Open.
36. Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine.
37. Wang, G., Ye, J.C., Mueller, K. and Fessler, J.A., 2018. Image reconstruction is a new frontier of machine learning. IEEE Transactions on Medical Imaging.
38. Wang, S., Summers, R.M. and Yao, J., 2016. Research challenges and opportunities in computational cancer imaging. IEEE Transactions on Biomedical Engineering.