Enhancing Prostate Cancer Diagnosis Accuracy with Artificial Intelligence: A Multimodal Approach Using Ensemble Techniques

Mihir Sontake (ID: 23241713)

Supervisor: Dr. Meghana Kshirsagar

Department of Computer Science and Information Systems

University of Limerick

Submitted to the University of Limerick for the partial fulfilment of degree

Master of Science In Software Engineering With Data Analytics (MSc) 2023-24

August 15, 2024

Abstract

This is an era of rapid advancements and increased adoption of Artificial Intelligence across industries. Artificial Intelligence is bringing transformations by enhancing efficiency and decision-making capabilities. Artificial Intelligence is being used in healthcare, customer service, finance, manufacturing, transportation, agriculture, retail, education, energy, human resources, environment, law services, and space exploration industries.

Artificial Intelligence in the healthcare industry is useful for detecting cancer. Particularly, Prostate cancer is being diagnosed with the help of Artificial Intelligence algorithms. Prostate Cancer diagnosis is an active area of Artificial Intelligence research, where advanced algorithms such as deep neural networks networks and convolutional neural networks are being utilized.

In this dissertation, we analyse whether using an Artificial Intelligence model trained using a multimodal dataset leads to better predictive outcomes for Prostate cancer detection. We aim to enhance the accuracy of the Artificial Intelligence model using fine-tuning and transfer learning.

The multimodal dataset we use contains both tabular and image data. We use logistic regression on the tabular data to predict a binary outcome of whether a patient has clinically significant Prostate cancer. We use a support vector machine algorithm on the tabular data to predict the grade of Prostate cancer. We use convolutional neural network algorithms to predict a binary outcome of whether a patient has clinically significant Prostate cancer.

We utilize pre-trained models of Inception v3 and ResNet 50 for transfer learning.

Finally, we merge the accuracy of both modalities using different weight ratios. We choose the weight ratio that gives the highest accuracy of the ensemble model.

Declaration of Authenticity

Title: Enhancing Prostate Cancer Diagnosis Accuracy with Artificial Intelligence: A Multimodal Approach Using Ensemble Techniques

Author: Mihir Sontake

Award: Master of Science

Supervisor: Dr. Meghana Kshirsagar

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other Irish examination board.

Mihir Sontake

Consent for Publication

I hereby consent that this dissertation document can be made publicly available in electronic format for research purposes.

Mihir Sontake

15th August, 2024

Use of AI/GenAI

I declare that I used Generative AI tools - ChatGPT and Grammarly to generate ideas for the dissertation, generate code for the notebooks and rephrase the dissertation content.

Acknowledgements

I want to express my sincere gratitude to my supervisor, Dr. Meghana Kshirsagar. She consistently guided and supported me from the beginning of my dissertation journey. Her efforts to help me improve my understanding of Artificial Intelligence and its role in healthcare were invaluable in enhancing my knowledge. Her valuable constructive feedback has played a crucial role in shaping and enriching my dissertation.

Further, I would like to thank my professor Dr. Nikola Nikolov who inspired me to delve deeper into the field of data mining. His lectures on data mining were beneficial in making basic ideas of data manipulations, feature engineering, model training, and model evaluation clear and provided a base for further research.

Working on a subject I am passionate about has been extremely rewarding and has provided me with many learnings and insights into its application in the healthcare industry. I am thankful to the University of Limerick for this opportunity.

I thank my parents for their unwavering support and motivation, which helped me complete this dissertation.

Mihir Sontake

Contents

List of Tables

LIST OF TABLES

List of Figures

markbothGLOSSARYGLOSSARY

LIST OF FIGURES

Glossary

AI Artificial Intelligence

ANN Artificial Neural Network

CNN Convolutional Neural Network

DL Deep Learning

LR Logistic Regression

ML Machine Learning

SVM Support Vector Machine

PCa Prostate Cancer

csPCa Clinically Significant Prostate Cancer

MRI Magnetic Resonance Imaging

bpMRI Biparametric MRI

mpMRI Multiparametric MRI

ISUP International Society of Urological Pathology

PI-RADS Prostate Imaging-Reporting and Data System

1

Introduction

1.1 Overview

This dissertation is focused on building an Artificial Intelligence model to diagnose prostate cancer or PCa. PCa is a serious condition that is a leading cause of cancer-related deaths in men. With an expected 1.5 million new cases and 397,000 deaths globally in 2022, prostate cancer (PCa) is the second most common cancer in the world and the fifth largest cause of cancer-related death in males Bray et al. (2024). Prostate cancer that is clinically significant, or csPCa, is frequently aggressive and needs to be detected early Pecoraro et al. (2021).

To diagnose PCa magnetic resonance imaging or MRI scanned images NIBIB (n.d.) are utilised. MRI scanned images create detailed images of soft tissues in the body using radio waves and strong magnets. MRI is a non-invasive tool that is used to detect, localize, and stage PCa, allowing prostate biopsy planning NIBIB (n.d.) Pecoraro et al. (2021). MRI scanned images give a clear picture of the prostate and nearby areas and they also can be used to identify the seriousness, spread, and location of PCa.

There is complexity involved in reading MRI scanned images which causes doctors to face difficulty in giving an accurate diagnosis He et al. (2023). Interpreting MRIs takes a lot of time, requires knowledge, and is typically accompanied by a significant inter-observer variability Pellicer-Valero et al. (2022). An Artificial Intelligence model developed in this dissertation will give an accurate and quick diagnosis, which can help doctors and clinicians verify their readings.

1. INTRODUCTION

Artificial Intelligence techniques such as Machine Learning or ML, and Deep Learning or DL, are being employed to accurately detect PCa, while also reducing the cost of detecting PCa. These techniques are used to analyze MRI scanned images and also patient data such as age, race, family history, and lifestyle factors according to Talaat et al. (2024).

The Artificial Intelligence model described in this dissertation utilised a multimodal dataset containing tabular and image data. It predicts whether a patient has csPCa and the PCa grade (0 to 5) Srigley et al. (2016). To achieve this the algorithm is trained on data from a publicly available dataset. This is the PI-CAI (Prostate Imaging: Cancer AI) grand challenge PI-CAI (n.d.) [1] The dataset utilised in this dissertation was developed by Saha et al. (2024). It contains data and labels from 1500 Patient exams.

The tabular data is in the form of a .csv file that contains features of patient ID, study ID, patient age, prostate volume, prostate-specific antigen, prostate-specific antigen density, histopath type, lesion Gleason score, lesion ISUP score, patient ISUP score, and patient csPca label.

The image data is from MRI scanned images in the form of .mha files. The dataset consists of 5 MRI scanned image types of .mha images for each patient which are described in this table 3.1. The .mha images are 3-dimensional, hence we utilise a Python library known as SimpleITK, to extract 2-dimensional .jpg images or slices from each .mha image.

The Artificial Intelligence algorithms that are utilised in this dissertation are logistic regression and support vector machine which take as input the .csv file, and convolutional neural networks which take as input the .jpg images. There are 1500 records in the .csv file, which is equivalent to 1 record for each patient. There are an overall 74,050 images for the model to train on in the ratio of: (28.69%:Yes and 71.31%: No).

The results from the tabular data and image data models are combined to give a prediction. The logistic regression algorithm and support vector machine algorithm are trained on the tabular data and the convolutional neural network

[1] Grand Challenges are difficult but important problems set by various institutions or professions to encourage solutions or advocate for the application of government or philanthropic funds especially in the most highly developed economies Gould (2010).

algorithm is trained on the MRI scanned 2-dimensional images extracted from the 3-dimensional MRI scanned images. The logistic regression algorithm is utilised here because it predicts only binary classes, hence it is useful for predicting whether csPCa is detected or not. The Support vector machine classifier predicts among 6 classes the PCa grade from ISUP 0 to ISUP 5. The convolutional neural network algorithm is selected as it is a widely used algorithm for computer vision applications. It autonomously learns features from images. The convolutional neural network will perform classification to determine whether the image contains csPCa or not. Our aim here is to improve the diagnostic accuracies of models from tabular and image datasets and combine them to enhance the detection accuracy.

The models are assessed on the test accuracy and loss. The accuracy scores of models of both modalities are combined by applying three different weight ratios. The weight ratio that gives the best prediction accuracy will be chosen for the Artificial Intelligence model.

1.2 Problem Statement

Developing an Artificial Intelligence model that can enhance diagnostic accuracy of clinically significant prostate cancer using an ensemble model trained on a multimodal dataset.

1.3 Research Objective

The primary objective of this research is to develop an Artificial Intelligence model utilising techniques such as convolutional neural networks, logistic regression, and support vector machines. By leveraging a multimodal dataset, this research aims to enhance the diagnostic accuracy of clinically significant prostate cancer.

1.4 Scope

The scope of the problem can be defined below:

1. An Excel file containing various patient medical information such as age, PSA, PSAD, Prostate volume, gleason score, ISUP grade, and a label of 'yes' and 'no'

2. A LR model trained using Excel file data to classify a record as having csPCa or not

3. A SVM model trained using the Excel file data to classify the ISUP grade of the patient

4. Leveraging SMOTE technique to oversample the minority classes

5. Evaluating model results using metrics: accuracy, specificity, sensitivity, ROC, AUC, F1-score

6. An image dataset containing 5 MRI sequences per patient, namely, axial T2W, sagittal T2W, coronal T2W, DWI, and ADC

7. A CNN model built using pre-trained CNNs of Inception v3 and Resnet 50 to classify a patient as having csPCa or not

8. Adjusting hyperparameters such as batch size, epochs, learning rate, and adding layers to observe performance changes

9. Evaluating model results using metrics: accuracy and loss

10. Evaluating three different weights assigned to both models that give the highest prediction accuracy

1.5 Research Question

- Can we enhance the accuracy of Artificial Intelligence models for Prostate Cancer Diagnosis using a multimodal dataset, fine-tuning, and ensemble techniques?

1.6 Research Contribution

The research contributions of this study are the following:

- Training a logistic regression algorithm on tabular data from 1500 patient MRI exams.
- Training a support vector machines classifier algorithm on tabular data from 1500 patient MRI exams.
- Training convolutional neural networks on MRI images from 1500 patient MRI exams using transfer learning from Inception v3 and Resnet 50 architectures.
- Combining the predictions of logistic regression, support vector machines and convolutional neural network.
- Assessing the performance of the combined model in different ways by assigning 50:50, 40:60, and 60:40 weight ratios.

1.7 Methodology

- Identify the research area for the study.
- Identify a large image dataset with labelled data.
- Approximate a systematic literature review.
  - Importance of prostate cancer and early detection.
  - Overview of logistic regression, support vector machine and convolutional neural network algorithms.
  - Overview of large-scale convolutional neural networks.
  - Current Gaps in research
- Define the research question.
- Design the study.

- Select the machine learning and deep learning algorithms for the study.
- Coding the algorithms selected in Python written on Jupyter Notebooks.
- Prepare an ensemble model combining the models by assigning weights.
- Evaluate the performance of the ensemble model using 3 different weight ratios - 50:50, 40:60, and 60:40.

1.8 Overview of Dissertation

The rest of the dissertation is organized as follows. Chapter 2 delves into a comprehensive literature review, exploring in-depth Machine Learning and Deep Learning Techniques and why they were chosen. Chapter 3 details the techniques of data loading, preprocessing, model building, model training, and model evaluation. This chapter also discusses the optimisations done on the model using fine-tuning on the hyperparameters, Chapter 4 discusses the results of the dissertation. Chapter 5 concludes the dissertation and outlines potential directions for future research.

1.9 Motivation

In my career in software comprising 4.8 years, I have focused on developing web applications. While I have no previous experience in building Artificial Intelligence models within a corporate setting, I would like to gain practical experience in this area. I understand the role of Artificial Intelligence technologies such as Machine Learning and Deep Learning algorithms in making society better. I am motivated to delve into learning these skills more deeply. Consequently, I have decided to concentrate in this study on understanding how Artificial Intelligence models can be effectively utilized in real-world healthcare scenarios.

During the preliminary research for my dissertation, I came across the concepts of Logistic Regression, Support Vector Machine, and Convolutional Neural Network. I was intrigued by the capabilities of Convolutional Neural Networks to learn features on their own from images and classify them. Hence, my goal

is to develop an Artificial Intelligence model that can be applied in real-world healthcare situations using these concepts.

2

Literature Review

This chapter presents a comprehensive literature review of prostate cancer and the importance of early detection. It discusses the current state-of-the-art Artificial Intelligence algorithms such as logistic regression, support vector machine, and deep learning. It also discusses large-scale convolutional neural networks that will be used in this dissertation for prostate cancer detection. Further, it discusses the current gaps in research.

2.1 Introduction

2.1.1 Prostate Cancer, Early Detection, and Diagnostic Techniques

2.1.1.1 Early Detection

The male reproductive and urinary systems include the prostate as a gland. Just below the bladder and ahead of the rectum is where you'll find it. It is comparable in size to a chestnut. Base, apex, anterior, posterior, and two lateral surfaces make up its composition MayoClinic (n.d.).

One in nine men is thought to develop prostate cancer or PCa, at some point in their lives. Men should have regular tests for PCa to check for any signs or symptoms because the disease can often be successfully treated if discovered early Talaat et al. (2024).

2.1.1.2 Diagnosis

There are many tests that a patient undergoes when detecting clinically significant Prostate Cancer or csPCa. Initially, a doctor decides on a test based on the patient's preferences and local expertise CancerResearchUK (n.d.). Prostate-specific antigen levels are measured by a blood test called a PSA (prostate-specific antigen). One kind of protein that the tissues in the prostate produce is called PSA. Elevated levels of PSA are an indication, however, not conclusive evidence of csPCa AmericanCancerSociety (n.d.). A digital rectal examination is performed to examine the prostate size and volume. Then depending on the test results further biopsy is recommended. A biopsy of the tissue is used to confirm csPCa AmericanCancerSociety (n.d.).

MRI exams can be used as a diagnostic tool to confirm csPCa by determining the PI-RADS score. PI-RADS (Prostate Imaging–Reporting and Data System) v2.1 Turkbey et al. (2019) is a structured reporting scheme for MRI to evaluate suspected csPCa. It provides a comprehensive set of standards for scanning, interpreting, and reporting multiparametric MRI (mpMRI). PI-RADS gives a score from 1 to 5 that determines the probability of csPCa.

PI-RADS scheme is defined as follows:

1. **PI-RADS 1:** very low probability (clinically significant cancer is doubtful to be present)

2. **PI-RADS 2:** low probability (clinically significant cancer is unlikely to be present)

3. **PI-RADS 3:** intermediate probability (the presence of clinically significant cancer is equivocal)

4. **PI-RADS 4:** high probability (clinically significant cancer is likely to be present)

5. **PI-RADS 5:** very high probability (clinically significant cancer is highly likely to be present)

Despite substantial improvements in MRI techniques, there are still limitations to prostate MRI interpretation in clinical practice. Inter-observer variability has been observed in MRI readings, even in the most recent PI-RADS version 2.1 standard Brembilla et al. (2020). Disparities in scanning parameters and image quality among scanners also pose a challenge in accurately comparing MRI studies Leake et al. (2014).

Biopsy should generally be considered for PI-RADS 4 or 5 lesions or PI-RADS 1 or 2 lesions to confirm csPCa Turkbey et al. (2019).

2.1.1.3 MRI techniques

The current recommendation for MRI imaging for prostate cancer detection is Biparametric MRI (bpMRI) Brembilla et al. (2020). bpMRI is a type of MRI that has the following MRI image sequences as given below:

- T2-weighted imaging (T2W)
- Diffusion-weighted imaging (DWI)
- Apparent diffusion coefficient maps (ADC)

These MRI sequences are described in the table 2.1.

Our dataset from the PI-CAI grand challenge PI-CAI (n.d.) Saha et al. (2024) utilised bpMRI imaging.

2.2 Review of Literature

2.2.1 Theoretical Foundations

2.2.1.1 Machine Learning

Machine learning or ML is a branch of computer science and a subset of Artificial Intelligence that deals with algorithms that extract information from data and make predictions. Machine learning includes several algorithms that perform Classification, Regression, and Clustering tasks. The term Machine Learning means that the machine is learning from patterns. Machine Learning algorithms

ID	Name	Description
1	T2-Weighted imaging	T2W is used to demonstrate the zonal anatomy of the prostate. It is ideal for the transitional zone of the prostate gland. Tumors can be well delineated and their location can be easily identified RadiologyAssistant (n.d.) Radiopaedia (n.d.*c*).
2	Diffusion-Weighted Imaging	DWI can be a measure of cellular density. Malignant tumours have a high cellular density hence, the ability of water to freely move is decreased, hence, the diffusion is impaired. Suspicious areas can be identified as those having high b-values RadiologyAssistant (n.d.) Radiopaedia (n.d.*b*).
3	Average Diffusion Coefficient maps	ADC maps give a quantitative measure of cellular density and can be considered a molecular imaging tool for tumour aggressiveness RadiologyAssistant (n.d.) Radiopaedia (n.d.*a*).

Table 2.1: Imaging Techniques and Their Descriptions RadiologyAssistant (n.d.)

are fed data from which they learn and optimise their operations to improve performance developing intelligence over time Datacamp (n.d.) IBM (n.d.*d*).

Machine Learning can be supervised or unsupervised. Supervised ML consists of Classification and Regression tasks. Supervised ML is a method that uses an algorithm to understand the relationship between independent and dependent variables. Supervised ML includes algorithms such as Random Forest, Linear Regression, K-Means Classifier, Logistic Regression and Support Vector Machine. Unsupervised ML consists of Clustering tasks. Unsupervised ML is a method that doesn't require any labelled data for training. It includes K-means clustering Datacamp (n.d.) IBM (n.d.*d*).

Machine Learning is gaining rapid adoption in various use cases across industries such as predicting credit card fraud, anomaly detection, and predicting the

spread of diseases IBM (n.d.*d*).

2.2.1.2 Logistic Regression

One type of supervised machine learning algorithm is logistic regression. It forecasts the likelihood of a certain result, occurrence, or observation between two classes or two potential By examining the correlation between one or more independent variables, logistic regression divides data into several groupings IBM (n.d.*c*). It is widely employed in predictive modeling, in which the model calculates the mathematical likelihood of an event falling into a particular category or not. It determines the weights across the dataset's features using a straightforward mathematical calculation IBM (n.d.*c*).

Here are a few instances of these groupings and situations where the binary response is either inferred or expected:

1. **Determine the probability of heart attacks:** Medical professionals can use a logistic model to ascertain how an individual's weight, exercise level, and other factors relate to one another and use that information to estimate the likelihood that the patient will experience a heart attack spiceworks (n.d.).

2. **Possibility of enrolling into a university:** Application aggregators can ascertain a student's likelihood of being accepted into a specific university course by examining the correlation between factors like GRE, GMAT, or TOEFL scores and acceptance spiceworks (n.d.).

A logistic or sigmoid function is used in logistic regression to map the probabilities of the predictions. An S-shaped curve known as the sigmoid function transforms any real number into a range between 0 and 1 IBM (n.d.*c*).

Moreover, the model predicts that the instance belongs to that class if the estimated probability produced by the sigmoid function exceeds a predetermined threshold on the graph. The model anticipates that the instance does not belong in the class if the calculated probability is less than the predetermined threshold IBM (n.d.*c*).

For instance, the output of the sigmoid function is regarded as 1 if it is more than 0.5. Conversely, the output is categorized as 0 if it is less than 0.5. Furthermore, if the graph continues to the negative end, y is expected to be equal to 0 and vice versa. Put otherwise, a sigmoid function output of 0.76 indicates that there is a 76% chance that the event will occur IBM (n.d.*c*).

Known as an activation function for logistic regression, the sigmoid function 2.1 is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{2.1}$$

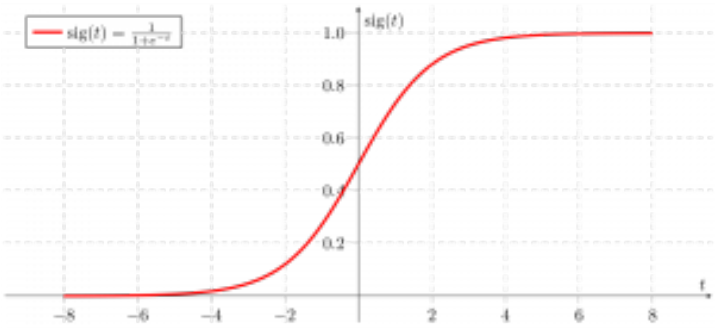


Figure 2.1: Sigmoid Function

The logistic regression equation is typically expressed as:

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \tag{2.2}$$

where, x = input vector, β = model coefficients, y = output variable (0 or 1)

The output value modeled here is binary (0 or 1) as opposed to a numeric value, in contrast to linear regression.

2.2.1.3 Support Vector Machine

For classification problems, a Supervised Machine Learning technique called Support Vector Machine (SVM) is employed. In an N-dimensional space, it finds the best hyperplane or line to maximize the distance between each class IBM (n.d.*a*).

The decision boundaries in a feature space that are used to divide the data points of various classes are called hyperplanes. It will be a linear equation in the case of linear classifications IBM (n.d.*a*).

If the hyperplane is a line in a two-dimensional space or a plane in an n-dimensional space, it is determined by the number of features in the input data IBM (n.d.*a*). The algorithm finds the optimal decision boundary between classes by maximizing the margin between points since numerous hyperplanes can be identified to differentiate classes IBM (n.d.*a*). It may therefore effectively generalize to new data and produce precise classification predictions as a result. Since they pass through the data points that establish the greatest margin, the lines that are next to the ideal hyperplane are referred to as support vectors IBM (n.d.*a*).

The 2.2 algorithm of SVM is characterized by its ability to disregard outliers and choose the optimal hyperplane that optimizes margin. Outliers don't affect SVM IBM (n.d.*a*).

Because it can handle both linear and nonlinear classification tasks, the SVM algorithm is frequently employed in machine learning IBM (n.d.*a*). Nevertheless, in cases where the data cannot be separated linearly, IBM (n.d.*a*), kernel functions are employed to convert the data into a higher-dimensional space, therefore permitting linear separation. The particular use case and the properties of the data will determine which kind of kernel function is used, such as linear, polynomial, radial basis function (RBF), or sigmoid kernels IBM (n.d.*a*).

2.2.1.4 Deep Learning

Deep learning is a subfield of machine learning that employs neural networks. The neural networks function similarly to the human brain IBM (n.d.*b*). Neural networks are also referred to as Artificial Neural Networks (ANN) due to their artificial creation 2.3. The majority of AI applications nowadays rely on deep learning IBM (n.d.*b*).

The topology of the underlying neural network architecture is the primary distinction between machine learning and deep learning IBM (n.d.*b*). Deep learning models use three or more layers, up to hundreds or thousands of layers, to train

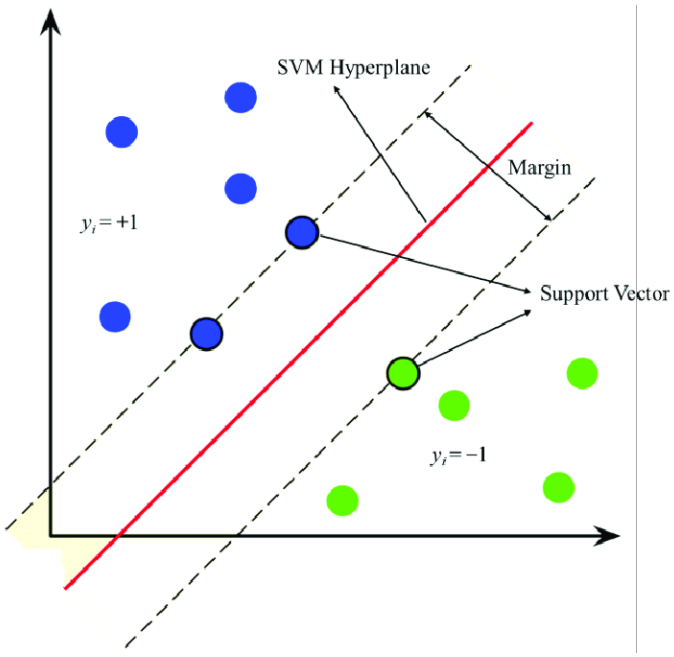


Figure 2.2: Support Vector Machine

the models, whereas typical machine learning models use simple neural networks with one or two computational layers IBM (n.d.*b*).

The fundamental concept of machine learning and deep learning, shared by both, is that the algorithms and models utilize input data and correct labels to learn patterns over time. Through multiple iterations, the model improves and learns the underlying patterns. This learning process is achieved by implementing a cost function and an optimizer. The optimizer's role is to minimize the loss or cost function using a set learning rate across numerous iterations IBM (n.d.*b*).

Deep neural networks have layers of connected nodes that work together to make predictions or categorize data. The process of passing data through the network is called forward propagation. The input and output layers are known as visible layers. The layers in the middle are known as hidden layers. The data is input by the input layer, and the final prediction or classification is generated by the output layer IBM (n.d.*b*).

Backpropagation modifies the model's weights and biases to train it after calculating prediction errors using techniques like gradient descent. This, along with forward propagation, helps the neural network make predictions and fix

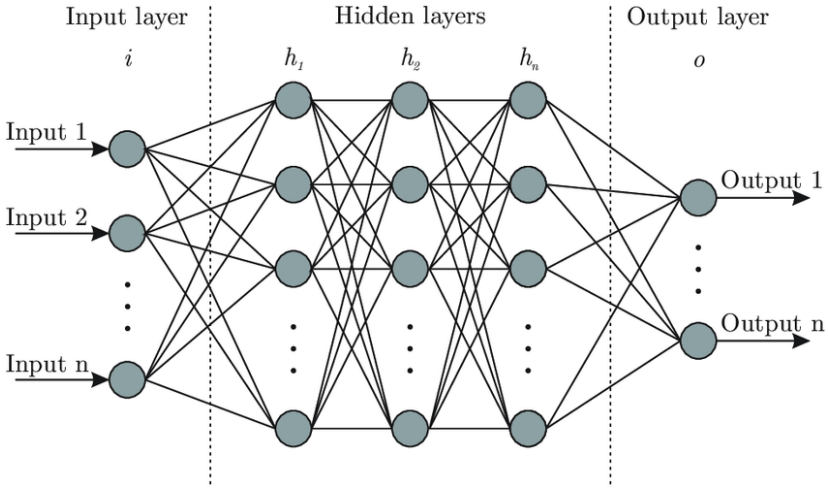


Figure 2.3: Artificial Neural Network

mistakes. After many rounds of adjustments, called epochs, the model becomes more accurate IBM (n.d.*b*).

Deep learning algorithms are mainly of the following types:

1. **Convolutional neural networks:** The main use of these is in applications related to computer vision and image categorization IBM (n.d.*b*).

2. **Recurrent neural networks:** Since they depend on sequential or time-based data, these are mainly used in applications for speech and natural language recognition IBM (n.d.*b*).

3. **Long Short-Term Memory Networks:** They are a particular kind of recurrent neural network that is particularly well-suited for tasks like speech recognition and time series data prediction because they can learn long-term dependencies and are expressly made to handle the problem of long-term dependence Simplilearn (n.d.*b*) IBM (n.d.*b*).

4. **Generative adversarial networks:** are neural networks that are used to generate new data that resembles the original training data, both inside and outside of artificial intelligence (AI). Some examples of these are pictures that seem to be human faces, but they are artificial and weren't taken of actual individuals. The back-and-forth between the generator and

discriminator, the two parts of the GAN, is where the word "adversarial" originates Simplilearn (n.d.*b*) IBM (n.d.*b*).

5. **Transformer Networks:** Transformers are a fundamental component of many contemporary models of natural language processing. By processing input data via self-attention methods, these models allow for parallelization and improved long-range dependency management IBM (n.d.*b*).

2.2.1.5 Convolutional Neural Network

Convolutional neural networks are feed-forward neural networks. Convolutional neural networks (ConvNets) have several hidden layers that extract information from images Simplilearn (n.d.*a*). It is used in computer vision tasks to recognize objects in images Simplilearn (n.d.*a*).

The construction of a ConvNet consists of the following layers explained in this diagram 2.4:

1. **Input Layer:** This is the input layer where ConvNet receives raw picture data in the form of pixel arrays Simplilearn (n.d.*a*).

2. **Convolutional Layer:** An image's features are initially extracted using a convolution layer. The convolution operation is carried out by a number of its filters Simplilearn (n.d.*a*).

   To create a feature map, each filter that is applied to the input image by this layer slides or convolves over the image. This aids in the detection of a variety of features, including patterns, textures, and edges Simplilearn (n.d.*a*).

3. **Activation layer or ReLU:** The rectified linear unit, or ReLU for short, is a kind of activation layer. ReLU sets all of the negative pixels to 0 by doing an element-wise operation. The next step is to shift the extracted feature maps to an activation layer Simplilearn (n.d.*a*).

   A corrected feature map is the result of adding non-linearity to the network through an activation layer. The ReLU activation function diagram can be seen at 2.5.

4. **Pooling layer:** This layer preserves the important information while decreasing the dimensionality of the feature maps Simplilearn (n.d.*a*). Another name for this process is down-sampling. To create a pooled feature map, the rectified feature map is passed through a pooling layer. Average and maximum pooling are two popular forms of pooling. This is done to lower memory usage and speed up computations.

5. **Fully Connected Layer:** The output is flattened and fed into one or more fully connected or dense layers following many convolutional and pooling layers. This results in the output layer that determines the final classification or prediction Simplilearn (n.d.*a*).

6. **Output Layer:** The ultimate classification result is shown in this layer.

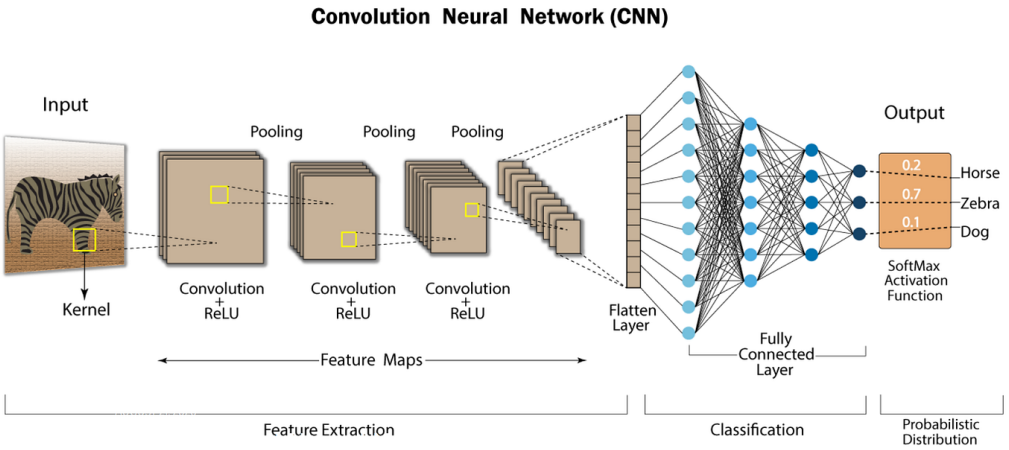


Figure 2.4: ConvNet Architecture

2.2.2 Current AI Technologies used for Prostate Cancer Detection

In machine learning (ML) technology, deep learning has become the most used AI technique He et al. (2023). Convolutional neural network models, one type of deep learning approach, have the potential to help physicians by increasing detection speed and accuracy.

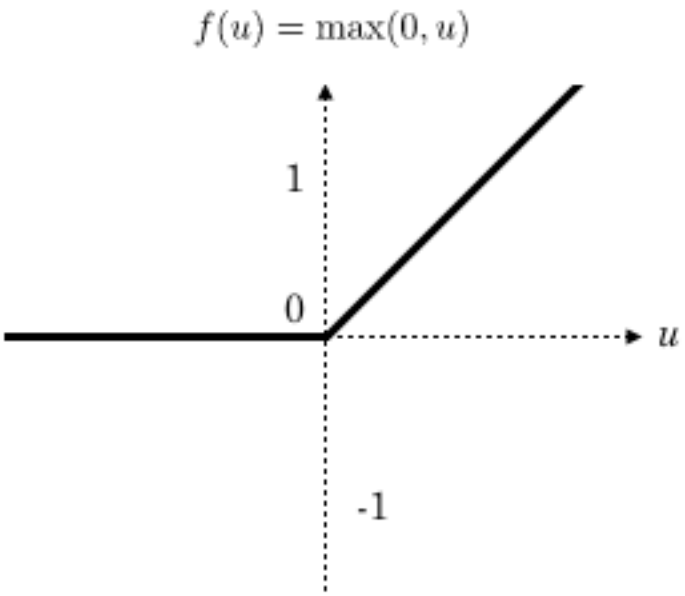


Figure 2.5: ReLU Activation Function

Deep learning can significantly improve the detection rates of PCa. It can lead to more accurate and faster diagnosis preventing over-diagnosis and under-diagnosis. By eliminating the need for costly tests and intrusive biopsies, deep learning can help save expenses.

MRI images present a complexity that poses challenges for accurate human interpretation He et al. (2023). This observation is supported by studies which have documented substantial intra-observer and inter-observer variability in readings Pellicer-Valero et al. (2022) Brembilla et al. (2020). Deep learning offers a promising solution in this context, as it serves to expedite interpretation processes, improve the quality of image interpretation, and mitigate the potential for overtreatment.

Deep Convolutional neural network (CNN) models are one of the many Deep learning models that have attracted a lot of interest because of their promising results in medical imaging He et al. (2023). CNN's multilayered neural network architecture, which draws inspiration from the human visual system, can process convolutional operations. Convolutional, max-pooling, and fully connected layers make up the three main layers of CNN structures. These tiers deal with certain

functions or computation techniques that take in, process, and provide pertinent data He et al. (2023). The core of CNNs is the convolution layer, which extracts picture characteristics using multiple convolution kernels. The max-pooling layer, also known as the downsampling layer, lowers computational work by combining data from a specific range. The fully connected layer, employed as a classifier, incorporates all local information collected from the previous max-pooling or convolutional layer that is class-distinctive, finally giving the appropriate class predictions He et al. (2023).

CNN-based networks have been built for specialized computer vision tasks, such as Inception, AlexNet, and ResNet for image classification, YOLO and Faster R-CNN for object identification, and U-Net and Mask R-CNN for semantic segmentation He et al. (2023).

2.2.2.1 A Review of Deep Learning Models in Prostate Cancer Detection

Wang et al. (2018) created an end-to-end CNN with two sub-networks: one for aligning apparent DWI and T2W, and the other as a convolutional neural classification network. The end-to-end CNN model was trained and tested on 360 patients using a fivefold cross-validation procedure, and it achieved a sensitivity of 0.89 for detecting high-risk PCa cases.

Ishioka et al. (2018) used a CNN architecture combining U-net with ResNet50 He et al. (2016) for PCa detection. De Vente et al. (2020) used network architectures that were adaptations of U-Net. Pellicer-Valero et al. (2022) used a Retina U-Net architecture that is intended for use with medical pictures, combining the Retina Net detector with the U-Net segmentation CNN. Talaat et al. (2024) suggested the Prostate Cancer Detection Model (PCDM) model for prostate cancer automatic diagnosis. To enhance the detection process' performance, PCDM used a modified ResNet50-based architecture that incorporates dual optimizers and a faster R-CNN. The suggested model performs better than both the ResNet50 and VGG19 architectures, according to the experimental results. The model was trained on a sizable dataset of annotated medical images. High rates of sensitivity, specificity, precision, and accuracy were attained by the PCDM model, which were 97.40%, 97.09%, 97.56%, and 95.24%, respectively.

2.2.3 Large Scale ConvNets

Using millions of photos and thousands of item categories, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) Russakovsky et al. (2015) serves as a benchmark for object category detection and classification. From 2010 to 2017, the competition was held yearly and attracted participation from over fifty institutions.

Since then, very deep convolutional networks have gained popularity and produced notable improvements across a range of benchmarks.

2.2.3.1 Inception v3

A CNN architecture called Inception v3 Szegedy et al. (2016) was unveiled in 2016 to analyze and classify images. The goal of Inception v3's architecture was to support deeper networks without going overboard with the number of parameters. Despite having 25 million parameters, its performance is comparable to algorithms with more parameters.

On the ILSVRC 2012 Deng et al. (2009) dataset, Inception V3 outperforms earlier Inception models and is on par with other modern architectures, achieving state-of-the-art performance. It has enhanced generalization, computing efficiency, and accuracy.

The Inception v3 architecture is frequently utilized as pre-trained architecture from ImageNet in libraries like Tensorflow, Keras, and PyTorch. It has been reused in numerous applications.

The architecture of Inception v3 represented in this image 2.6 can be defined as follows:

1. Consists of 42 layers deep network Opengenus (n.d.).

2. To effectively capture information at various scales, the Inception module combines a variety of parallel convolutional layers with various kernel sizes (1x1, 3x3, and 5x5) and pooling operations Opengenus (n.d.).

3. It helps in learning rich representations while controlling computational costs Opengenus (n.d.).

4. Using Label Smoothing, Factorized 7 x 7 convolutions, and an auxiliary classifer to propagate label information lower down the network (along with batch normalization for layers in the sidehead), Inception-v3 is an improved convolutional neural network architecture from the Inception family Opengenus (n.d.) Sciencedirect (n.d.).

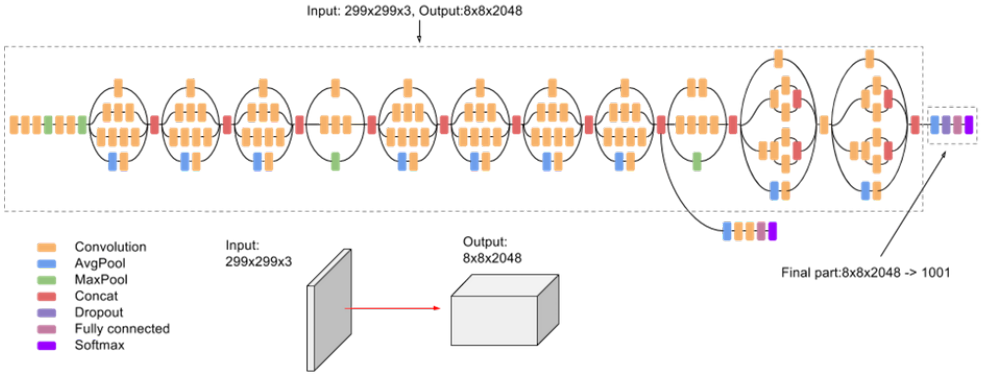


Figure 2.6: Inception v3 Architecture

2.2.3.2 ResNet 50

The CNN architecture known as ResNet 50 He et al. (2016) is a member of the ResNet (Residual Networks) family, a group of models created to tackle the difficulties involved in training deep neural networks Roboflow (n.d.). ResNet 50, created by researchers at Microsoft Research Asia, is well-known for its complexity and effectiveness in image classification applications. ResNet designs are available at several depths: ResNet 18, ResNet 32, and so on. ResNet 50 is a mid-sized version Roboflow (n.d.).

Deep neural network deterioration was the main issue that ResNet resolved. Their accuracy saturates and then rapidly declines as networks get deeper. It's not overfitting that's causing this decline, but rather the challenge of training process optimization Roboflow (n.d.).

Residual Blocks, which enable information to flow directly over skip links and mitigate the vanishing gradient problem, are how ResNet handled this issue Roboflow (n.d.).

Using residual networks, characterized by the usage of skip connections and residual blocks, has not only altered the way we train these networks but has also accelerated the development of more sophisticated and efficient models Roboflow (n.d.).

Deeper neural network training has been made possible by ResNet-50's remarkable ability to overcome vanishing gradient issues with its 50 bottleneck residual blocks Roboflow (n.d.).

The architecture of ResNet 50 represented in this image 2.7 can be defined as follows:

1. **ReLU Activation Layer:** Following the batch normalization layers and each convolutional layer, the ReLU (Rectified Linear Unit) activation function is applied. ReLU introduces non-linearity into the network by allowing only positive values to flow through. This is necessary for the network to learn intricate patterns in the data Roboflow (n.d.).

2. **Bottleneck Convolution Layers:** Three convolutional layers make up the block, each followed by batch normalization and ReLU activation Roboflow (n.d.).

   (a) By using a 1x1 filter size, the first convolutional layer lowers the number of channels in the input data. Without losing too much information, this reduction in dimensionality aids in data compression and increases computational efficiency.

   (b) Utilizing a 3x3 filter size, the second convolutional layer extracts spatial characteristics from the input.

   (c) Before the output is added to the shortcut connection, the third convolutional layer once more applies a filter size of 1x1 to get the number of channels back to what they were Roboflow (n.d.).

3. **Skip Connection:** Similar to a typical residual block, the skip connection is the most important component. It makes it possible to add the original input straight to the convolutional layers' output. This bypass connection makes sure that even in cases where the convolutional layers find it difficult

to pick up new features in that particular block, crucial information from previous layers is retained and distributed across the network Roboflow (n.d.).

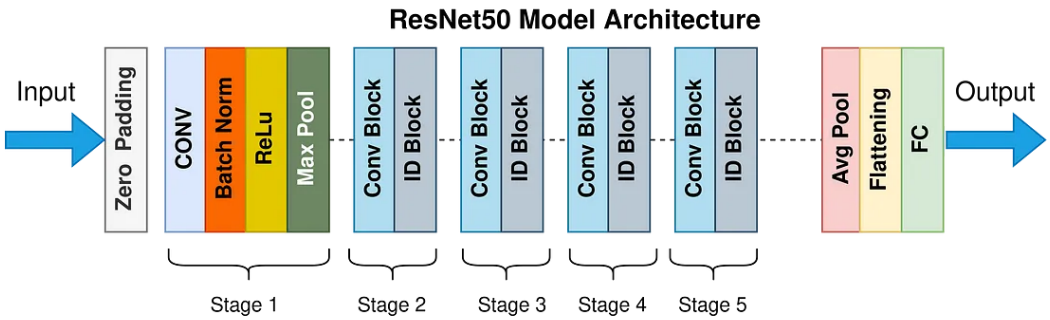


Figure 2.7: ResNet 50 Architecture

2.2.4 Current Gaps in Research

Because prostate tissue is complex and imaging data might vary, it can be difficult for AI algorithms to discern between benign and malignant tumors Alshuhri et al. (2023).

Presently, most available DL Agorithms rely on 2D images for feature extraction and analysis. In lesion identification tasks, two-dimensional (2D) slice-wise CNNs are known to typically perform worse than real 3D CNNs Pellicer-Valero et al. (2022).

Enhancing CNN architectures may make DL Algorithms more computationally capable He et al. (2023) Michaely et al. (2022). We are fine-tuning our DL models to achieve better performance and fill this research gap.

DL CNNs are known for being a 'black box' in regards to how decisions are made. CNNs may not be as transparent as ML-based techniques, leading some clinicians to avoid utilizing them on real patients outside of research Michaely et al. (2022).

The quantity of annotated sample data used is relatively small Alshuhri et al. (2023) He et al. (2023). Because of the use of tiny datasets and the lack of consistency and repeatability in the obtained imaging data, AI techniques exhibit

difficulty in overfitting Alshuhri et al. (2023) Corradini et al. (2021). We are using a large dataset with annotated data to fill this research gap.

The utility of DL for PCa diagnosis and therapy will be improved by model and algorithm optimization, the addition of multi-omics data, and the growth of medical database resources He et al. (2023). We are utilizing a large multimodal dataset and ensemble techniques to enhance the AI model's diagnostic accuracy and fill this research gap.

3

Methodology

3.1 Introduction

We want to diagnose prostate cancer from a multimodal dataset. Since prostate cancer diagnosis is a tough problem to solve even for experienced doctors, our Artificial Intelligence model aims to diagnose prostate cancer with high accuracy.

3.2 Research Design

Per the research objective, we will build models to diagnose cancer using multiple modalities and combine their accuracy scores with a weight ratio.

The first mode is through tabular data that will use algorithms of logistic regression, a type of binary classifier, and support vector machines, a type of multi-class classifier. Logistic regression will classify the patient into either of two categories - having clinically significant prostate cancer or not having clinically significant prostate cancer. Support vector machines will predict the ISUP grade of the patient.

The second mode is through image data that will use the algorithm of a convolutional neural network. It will classify an image into 2 categories - having clinically significant prostate cancer or not having clinically significant prostate cancer.

Finally, we combine multiple mode model's accuracies. The study uses 3 weight ratios by assigning the weights to the logistic regression/SVM and CNN model accuracies. Then, the weight ratio with the highest accuracy is chosen as the best performing on the dataset. This image 3.1 describes this workflow.

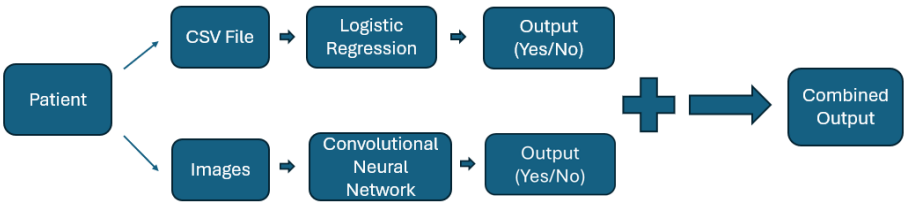


Figure 3.1: Classification of Clinically Significant Prostate Cancer (csPCa)

In logistic regression and SVM, we use exploratory data analysis, data preprocessing, model training, and evaluation metrics.

In the CNN model, we use exploratory data analysis, image preprocessing (resizing, augmentation), transfer learning, hyperparameter optimisation, model training, and evaluation metrics.

3.3 Data Collection

The dataset used for this thesis was taken from the PI-CAI (Prostate Imaging: Cancer AI) challenge PI-CAI (n.d.). The dataset contains a Public Training and Development Dataset (1500 cases).

It is available for all participants and researchers, to train and develop AI models Saha et al. (2024).

All data is fully anonymized and made available under a non-commercial CC BY-NC 4.0 license.

Imaging data has been released via: zenodo.org/record/6624726 (DOI: 10.5281/zenodo.6624726)

Annotations have been released and are maintained via: github.com/DIAGNijmegen/picai_labels

Each patient case includes three standard imaging sequences: axial T2W, axial DWI, and axial ADC scans (files ending in _t2w.mha, _hbv.mha, _adc.mha, respectively). In addition to these, patients may also have sagittal and coronal

T2W scans (_sag.mha, _cor.mha files), but they are not mandatory. Notably, no patient case will contain dynamic contrast-enhanced (DCE) sequences.

Out of the 1500 cases in the Public Training and Development Dataset, 1075 cases have benign tissue or indolent PCa, while 425 cases have csPCa.

3.4 Implementation

3.4.1 Logistic Regression Model

The first step is importing the required libraries in Python. The next step is to load the dataset.

3.4.1.1 Exploratory Data Analysis

Then, we perform Exploratory Data Analysis to understand the data and plot various graphs. We use exploratory data analysis to get the summary of the dataset, find the number of missing values per column, find out the numerical and categorical features, measure the class imbalance, and draw various plots for numerical features - histograms, boxplots 3.6, and correlation heatmap 3.4. The categorical features are visualised with countplots 3.5 to understand the data through visualisations.

We plot the class imbalance of 'case_csPca' values in this diagram 3.2.

We plot the class imbalance of 'case_ISUP' values in the diagram 3.3.

3.4.1.2 Data Preprocessing

The next step is to prepare the data for the model. We encode the 'csPCa' labels of 'YES' and 'NO' as 1 and 0, respectively. Then we count the number of missing values per column. We drop the columns of 'study_id', 'patient_id', 'mri_date', 'histopath_type', 'lesion_GS', 'lesion_ISUP', and 'case_ISUP'. Next, we impute the missing values in column 'psad'. We divide the column 'psa' by 'prostate_volume' to impute missing values in column 'psad'. Next, we drop all rows with missing values. Now we apply MinMaxScaler to normalise all features except the 'case_csPCa' as that is the predicted value.

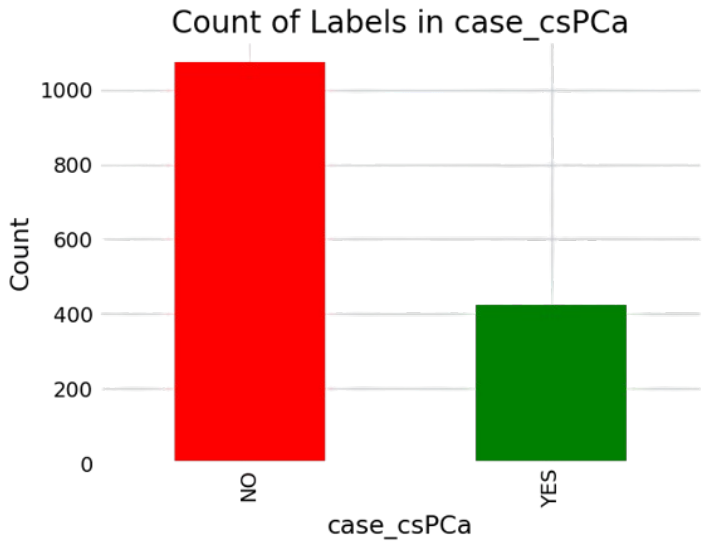


Figure 3.2: Class Imbalance

Now we need to convert all categorical columns to numerical columns. We use one-hot encoding on the 'histopath_type' column to make it numerical. Now we convert the 'lesion_GS' column to a new column 'max_lesion_GS' to a number by summing the numbers for each lesion, and taking the maximum sum of all lesion sums to get a single number. Now we convert the 'lesion_ISUP' to a new column 'max_lesion_ISUP' to find the maximum value of a lesion ISUP. We drop columns - 'lesion_ISUP' replacing it with 'max_lesion_ISUP', and 'lesion_GS' replacing it with 'max_lesion_GS'. Finally, we drop the column 'case_ISUP' as it is the same as 'max_lesion_GS'.

3.4.1.3 Model Training

Next, we split the columns into X = independent variable and Y = dependent variable. We split the dataset into 4 variables - X_train, X_test, Y_train, and Y_test with an 80:20 train:test ratio.

Next, we balance the 2 classes to have an equal number of labels in the training data for both classes. For this SMOTE or synthetic minority oversampling technique generates synthetic data for the minority class - YES or 1 class here.

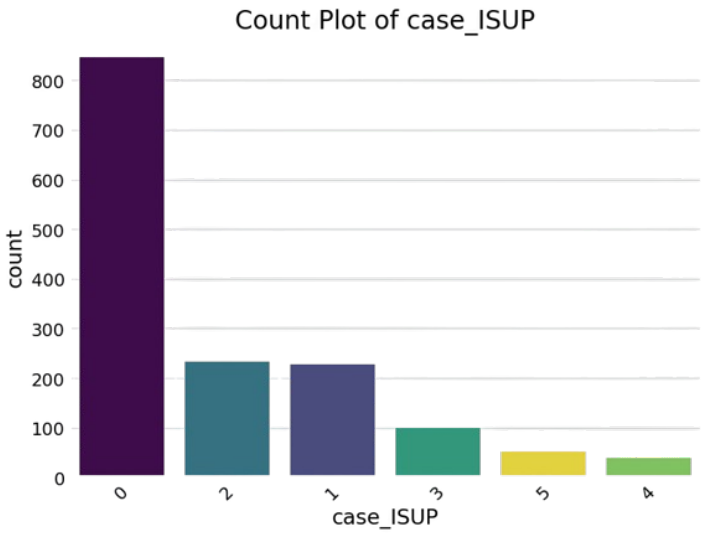


Figure 3.3: SVM Class Imbalance

Next, we the logistic regression model, the code snippet is given in this diagram 3.8.

3.4.1.4 Evaluation Metrics

Next, we evaluate metrics such as accuracy, precision, recall, and f1-score. For the logistic regression model metrics are as given in this image 3.9.

Finally, we plot the ROC curve and calculate the value of AUC for logistic regression 3.10.

3.4.2 Support Vector Machine Model

3.4.2.1 Data Preprocessing

We drop the columns 'study_id', 'patient_id', 'mri_date', and 'case_csPCa'. We drop any rows with missing values in columns 'histopath_type', 'lesion_GS', and 'lesion_ISUP'. We impute the missing values of column 'psad' similar to the logistic regression notebook.

Now our task is to convert the categorical features to numerical features. We one hot encode the 'histopath_type' column. We convert the 'lesion_GS' column to a new column called 'max_lesion_GS' to a number by summing the numbers and

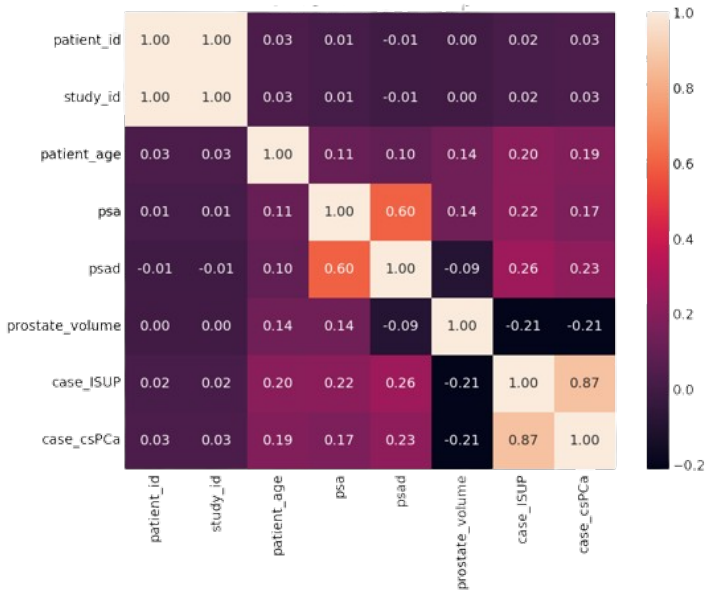


Figure 3.4: Heatmap

taking the maximum sum of all lesion sums. Next, we convert the 'lesion_ISUP' to a new column 'max_lesion_ISUP' to find the maximum value of a lesion ISUP. Finally, we drop columns - 'lesion_ISUP' replaced by 'max_lesion_ISUP', and 'lesion_GS' replaced by 'max_lesion_GS'. Next, we drop 'case_ISUP' as it is the same as 'max_lesion_ISUP'.

As all features are numeric now we can apply normalisation on all except the 'case_ISUP' label feature using MinMaxScaler.

3.4.2.2 Model Training

Next, we split the columns into X = independent variable and Y = dependent variable. We split the dataset into 4 variables - X_train, X_test, Y_train, and Y_test with an 80:20 train:test ratio.

Next, we balance the 6 label classes to have an equal number of labels in the training data for all 6 classes. For this SMOTE or synthetic minority oversampling technique generates synthetic data for the minority classes.

Next, we use grid search cross-validation to select the best parameters for the SVM model. We use parameters to maximise 'accuracy'. It is given in this

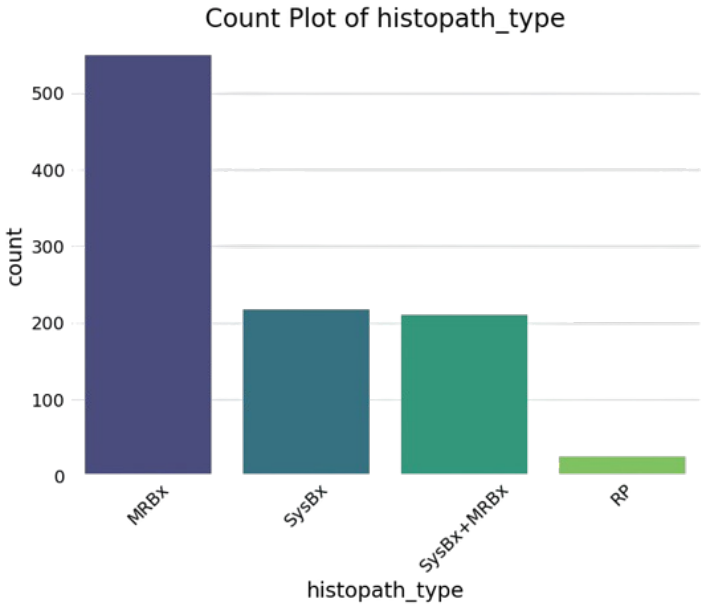


Figure 3.5: Count Plot of Histopath Type

code snippet 3.11. After we get the best parameters, we train the support vector machine model given in this code snippet 3.12.

3.4.2.3 Evaluation Metrics

Next, we evaluate metrics such as accuracy, precision, recall, and f1-score. For the SVM model metrics are as given in this image 3.13.

Finally, we plot the ROC curve and calculate the value of AUC for support vector machine models 3.14.

3.4.3 Image data Model

3.4.3.1 Data Loading

First, we import the libraries such as Tensorflow, Sklearn, and SimpleITK. Then we download the image dataset as ZIP files from the URLs. Then we unzip these files to get the images for each Patient. Next, we divide all the images according to the csPCa label for the patient - either 'YES' or 'NO', present in the CSV file.

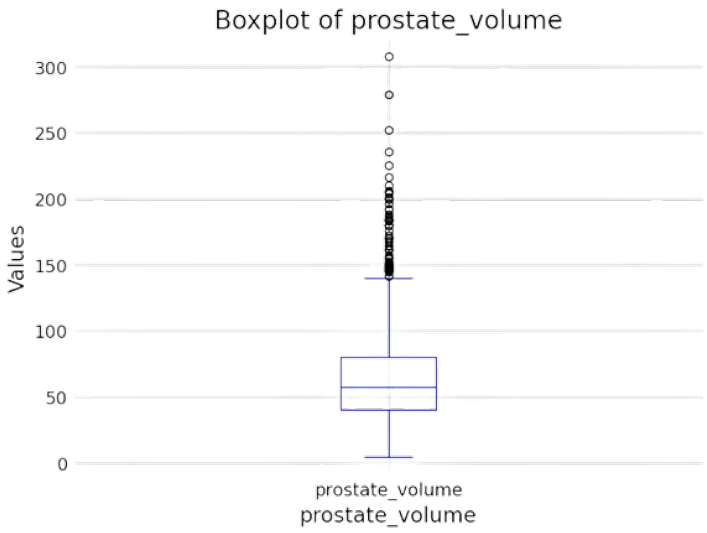


Figure 3.6: Boxplot of Prostate Volume

```
from imblearn.over_sampling import SMOTE
smote = SMOTE()
X_train_res, y_train_res = smote.fit_resample(X_train, y_train)
```

Figure 3.7: Synthetic Minority Oversampling Technique (SMOTE)

Next, we extract slices from the MHA images in the dataset. We create a directory called 'input_images'. Inside 'input_images' we create folders 'Yes' and 'No'. Within 'Yes' and 'No' we create more folders for the MRI sequences - 't2w', 'adc', 'cor', 'hbv', and 'sag'. Explained in the table 3.1 are their meanings.

We store the slices in their respective folders according to the label and sequence type. We extract only the top 10 slices from each MHA image and use it for model training. We also update the number of slices per image and patient in the DataFrame and store it as a CSV file. A figure of the CSV file is given 3.15 with information on the number of slices per patient and MRI sequence type.

3.4.3.2 Data Visualisation

The following figures show the prostate cancer positive 3.16 and negative 3.17 T2W images.

```
model = LogisticRegression()
model.fit(X_train_res, y_train_res)
```

Figure 3.8: Logistic Regression Model Training

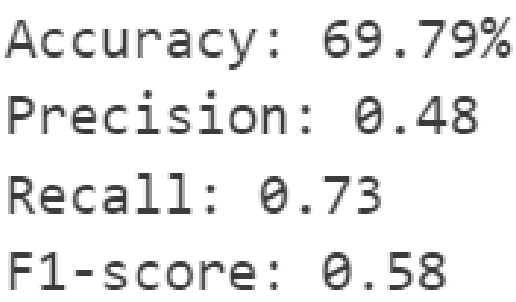


Figure 3.9: Logistic Regression Metrics

3.4.3.3 Exploratory Data Analysis

We use exploratory data analysis (EDA) to generate plots showing the class imbalance in the number of images per label for both t2w images and all images combined. The plot for the t2w class imbalance is 3.18. The plot for all images' class imbalance is 3.19.

3.4.3.4 Data Preprocessing

The data and label arrays are used for the model input. We use the label as 1 for Cancer Images and 0 for No Cancer Images. The data array appends all images as arrays with the numpy library after resizing them to (256, 256) dimensions. We use libraries matplotlib, PIL, numpy, and OpenCV (cv2) for this purpose.

For Experiments 1 and 2, we use only t2w images. For Experiment 3 we use the complete image dataset.

Then a splitting ratio of train and test sets given by the 0.85:0.15 ratio is applied to get variables - X_train, X_test, Y_train, and Y_test.

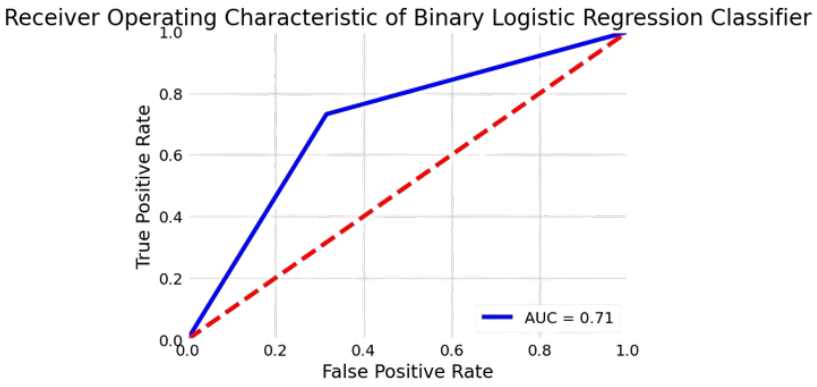


Figure 3.10: Logistic Regression ROC Curve

ID	Name	Description
1	t2w	Axial T2-weighted Imaging
2	adc	Apparent Diffusion Coefficient map
3	hbv	Diffusion-Weighted Imaging or DWI
4	cor	Coronal T2-weighted Imaging
5	sag	Sagittal T2-weighted Imaging

Table 3.1: MRI Sequences

3.4.3.5 Model Building

Then, we build the models using Transfer Learning using Inception v3 Szegedy et al. (2016) and ResNet 50 He et al. (2016).

The 4 models we use are described in this table 3.2.

3.4.3.6 Data Pipeline

The Image Data Generator is used to augment only the training data. Re-scaling is done for both train and test data to normalise the pixel values by dividing by 255. We take care not to horizontally flip the data while augmenting the train data as medical features may be destroyed by flipping horizontally.

```
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
from sklearn.preprocessing import StandardScaler


# Define the parameter grid
param_grid = [
    {'C': [0.1, 1, 10, 100, 1000],
    'gamma'  [1, 0.1, 0.01, 0.001, 0.0001],
    'kernel': ['linear', 'rbf', 'polynomial', 'sigmoid']}
]


# Create a GridSearchCV object
grid_search = GridSearchCV(SVC(), param_grid=param_grid, cv=5, scoring='accuracy')

# Fit the model on the training data
grid_search.fit(X_train_res, y_train_res)

# Print the best parameters and the best score
print("Best parameters found: ", grid_search.best_params_)
print("Best cross-validation accuracy: {:.2f}".format(grid_search.best_score_))
```

Figure 3.11: Support Vector Machine Grid Search Cross Validation

```
from sklearn.svm import SVC
svm_model   SVC(kernel = 'rbf', C = 1000, gamma=1).fit(X_train_res, y_train_res)
svm_predictions : svm_model.predict(X_test)
```

Figure 3.12: Support Vector Machine Model Training

3.4.3.7 Model Training

All 4 models are trained with the image data with 10 - 20 epochs and a batch size of 32.

The first 2 experiments are trained on only t2w images. The third experiment is trained on the complete dataset of images.

Experiment 1 We use an Adam optimizer with a learning rate of 0.001. We use a pre-trained Inception v3 model from the Tensorflow library with 'imagenet' weights. We don't use the first and last layers of Inception v3 as we don't want to modify them. Then, we Freeze the layers of Inception v3 by not training the

```
Accuracy: 0.845360824742268
Precision: 0.8849261632766787
Recall: 0.845360824742268
F1-Score: 0.8508836524300443
```

Figure 3.13: Support Vector Machine Metrics

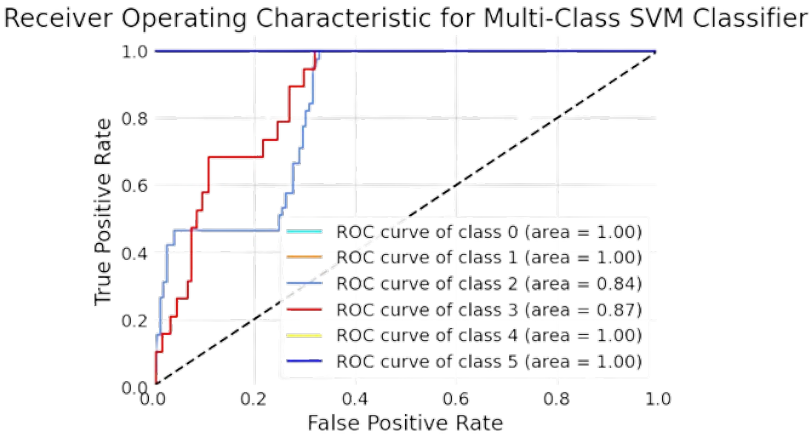


Figure 3.14: Support Vector Machine ROC Curve

ID	Model Name	Description
1	Inception 1	Inception v3 model with a fully connected output layer having 1 neuron.
2	Inception 2	Modified Inception 1 model's architecture with 6 more layers, dropout and regularization added. 3.20.
3	ResNet 1	ResNet 50 model with a fully connected output layer having 1 neuron.
4	ResNet 2	Modified ResNet 1 model's architecture with 6 more layers, dropout, and regularization added. 3.21.

Table 3.2: Summarizing all models

existing weights. Then we flatten the output of Inception v3 and apply a dense layer with 1 neuron and a sigmoid function for binary classification.

patient_id	study_id	mri_date	patient_ag	psa	psad	prostate_v	histopath_	lesion_GS	lesion_ISU	case_ISUP	case_csPC	t2w	adc	cor	hbv	sag
10000	1000000	02-07-2019	73	7.7		55	MRBx	0+0	0	0	NO	31	31	23	31	29
10001	1000001	27-05-2016	64	8.7	0.09	102				0	NO	21	21	19	21	19
10002	1000002	18-04-2021	58	4.2	0.06	74				0	NO	22	22	22	22	19
10003	1000003	05-04-2019	72	13		71.5	SysBx	0+0	0	0	NO	23	23	23	23	23
10004	1000004	21-10-2020	67	8	0.1	78	SysBx+MRI	0+0,0+0	0,0	0	NO	21	21	15	21	19
10005	1000005	18-07-2012	64	12.1	0.24	51	MRBx	4+3,0+0	3,0	3	YES	19	19	15	19	19
10006	1000006	23-10-2020	73	6.2	0.23	27	SysBx+MRI	0+0,3+3	0,1	1	NO	21	21	21	21	19

Figure 3.15: DataFrame With Slices Information

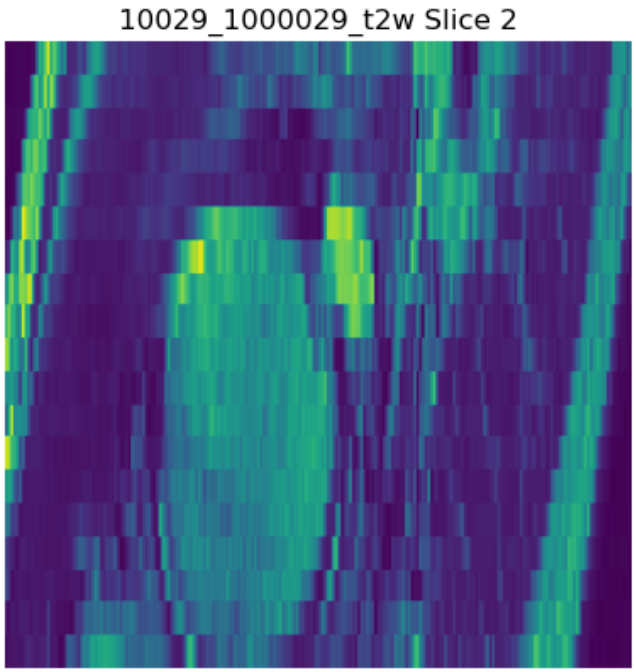


Figure 3.16: Prostate Cancer Positive

Then, we use a flattened layer on the Inception v3 output. Then, the output of the flattened layer is inputted to a Dense layer which has 1 neuron and a sigmoid activation function for predicting the 2 classes.

The second model adds layers after the inception's output. It is given in the image 3.20.

The third model uses ResNet 50 pre-trained model and similar to the first model of Inception v3 uses freezing the layers and Adam optimiser with a learning rate of 0.001.

The fourth model adds similar modifications to the ResNet 5 model that the second model added in the first model with the same Adam optimizer and learning

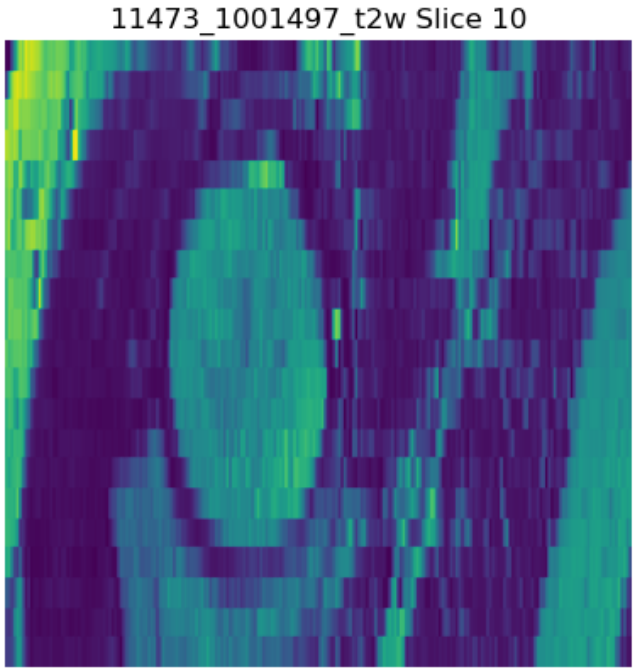


Figure 3.17: Prostate Cancer Negative

rate. It is given in this image 3.21.

Experiment 2 Based on the results from Experiment 1 we perform optimisations on the models to correct the overfitting problem.

1. Increasing Epochs to 20 for Inception 2 and ResNet 2 due to the larger number of units and more layers.

2. Reducing the Adam optimiser learning rate by a factor of 10 to learn more features and improve accuracy.

3. Using a callback function for Early Stopping the training process.

Experiment 3 Now we train the models on the complete image dataset comprising all MRI sequences as per Experiment 2.

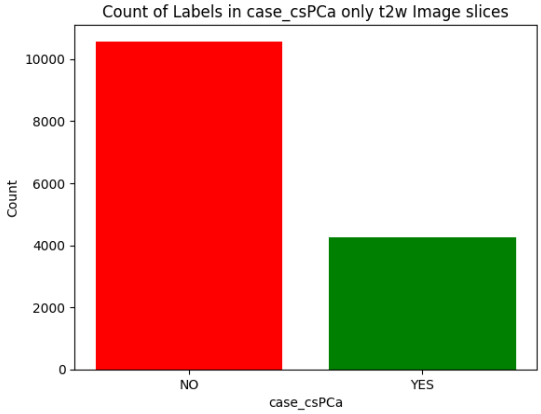


Figure 3.18: T2W Class Imbalance for CNN model

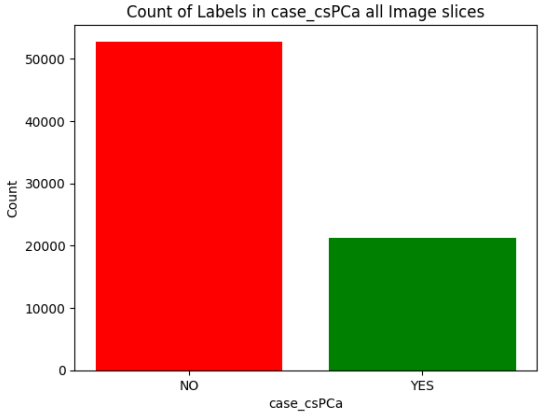


Figure 3.19: Class Imbalance for CNN model for all types of images

3.4.3.8 Results

The model accuracies and losses are evaluated on the test and train data. The train and test accuracies and losses every epoch are plotted. For experiment 1 they are given here 3.22. For experiment 2 they are given here 3.23. For experiment 3 they are given here 3.24.

The ResNet 2 model performs better in all three experiments. The loss function goes down with the number of epochs and the test accuracy value increases with the number of epochs. Hence, we choose the ResNet 2 model as our final CNN model.

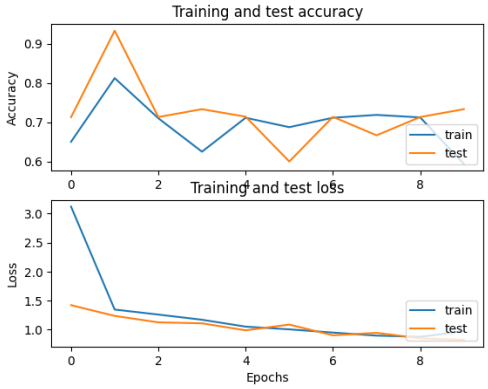
```
x = layers.Flatten()(inception.output)

x = layers.Dense(1024, activation='relu', kernel_regularizer=L2(0.001))(x)
x = layers.Dropout(0.2)(x)
x = layers.Dense(512, activation='relu', kernel_regularizer=L2(0.001))(x)
x = layers.Dropout(0.2)(x)
x = layers.Dense(512, activation='relu', kernel_regularizer=L2(0.001))(x)
x = layers.Dropout(0.2)(x)
prediction = layers.Dense(1, activation='sigmoid')(x)
```

Figure 3.20: Model: Inception 2

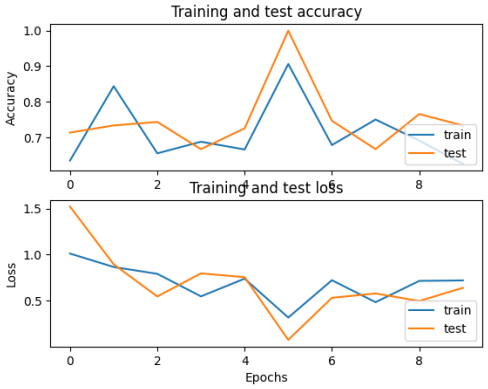
```
x = layers.Flatten()(resnet50_model.output)

x = layers.Dense(1024, activation='relu', kernel_regularizer=L2(0.001))(x)
x = layers.Dropout(0.2)(x)
x = layers.Dense(512, activation='relu', kernel_regularizer=L2(0.001))(x)
x = layers.Dropout(0.2)(x)
x = layers.Dense(512, activation='relu', kernel_regularizer=L2(0.001))(x)
x = layers.Dropout(0.2)(x)
prediction = layers.Dense(1, activation='sigmoid')(x)
```

Figure 3.21: Model: ResNet 2

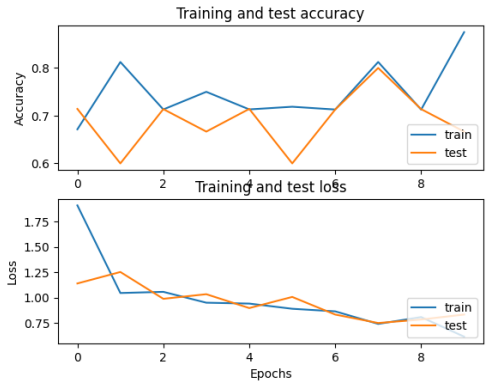
(a) Inception 1 Test and Train Accuracy and Loss



(b) Inception 2 Test and Train Accuracy and Loss



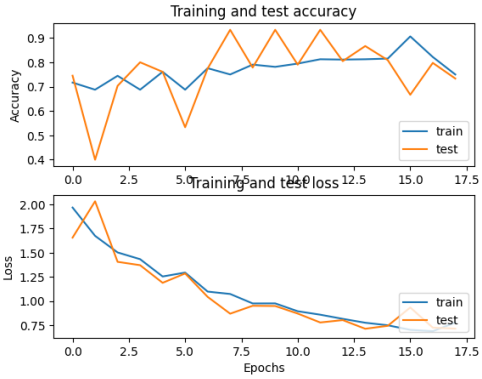
(c) ResNet 1 Test and Train Accuracy and Loss



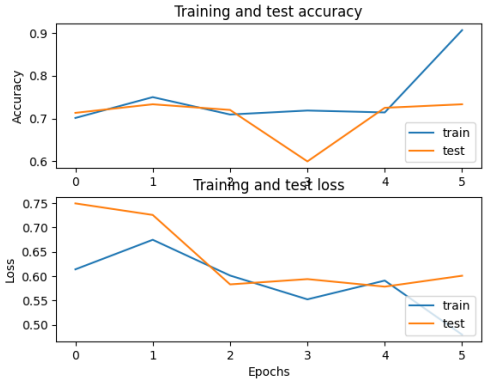
(d) ResNet 2 Test and Train Accuracy and Loss

Figure 3.22: Test and Train Accuracies and Losses each epoch for Experiment 1

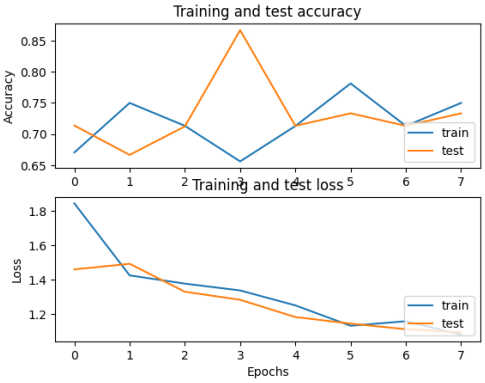
(a) Inception 1 Test and Train Accuracy and Loss on Experiment 2



(b) Inception 2 Test and Train Accuracy and Loss on Experiment 2

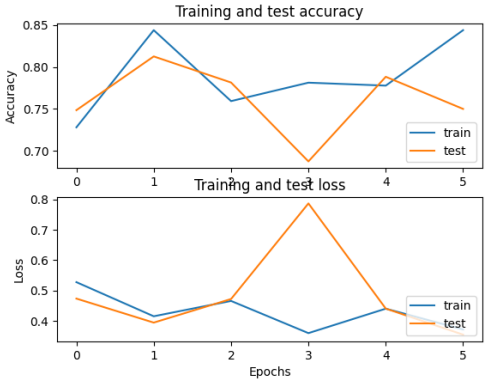


(c) ResNet 1 Test and Train Accuracy and Loss on Experiment 2

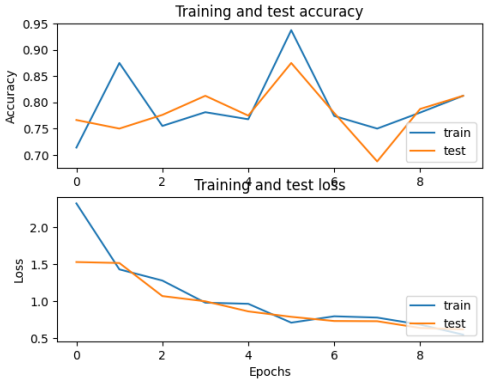


(d) ResNet 2 Test and Train Accuracy and Loss on Experiment 2

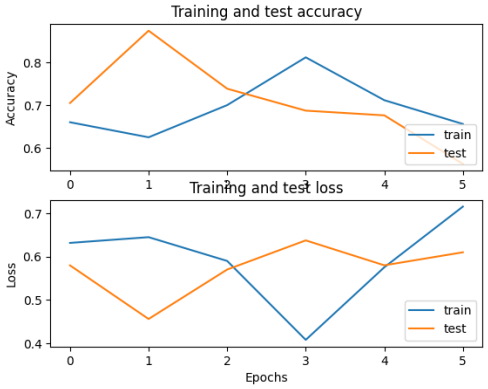
Figure 3.23: Test and Train Accuracies and Losses each epoch for Experiment 2



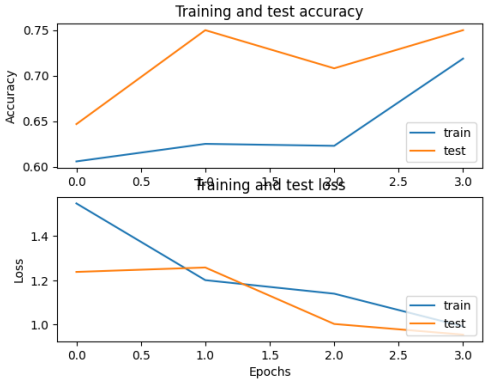
(a) Inception 1 Test and Train Accuracy and Loss on Experiment 3



(b) Inception 2 Test and Train Accuracy and Loss on Experiment 3



(c) ResNet 1 Test and Train Accuracy and Loss on Experiment 3



(d) ResNet 2 Test and Train Accuracy and Loss on Experiment 3

Figure 3.24: Test and Train Accuracies and Losses each epoch for Experiment 3

3.5 Weight Ratios Selection

We calculate a weighted average of the tabular and image model accuracies to get a final accuracy of the ensemble model. The final accuracy of the ensemble model is compared on three weights. These weights are given in 3.3.

We know the accuracy for the Logistic Regression model is 69.79% 3.9. The ResNet 2 CNN models' test accuracies and test losses are given in 3.4.

We choose the ResNet 2 model as our final CNN model due to its better performance. It showed a decline in both training and test loss with the number

ID	Tabular Model Weight (%)	Image Model Weight (%)
1	50	50
2	40	60
3	60	40

Table 3.3: Weights for Tabular and Image Models

of epochs, as well as an increase in both training and test accuracy with the number of epochs.

ID	Model	Test Accuracy	Test Loss
1	Inception 1	0.75	0.35
2	Inception 2	0.73	0.72
3	ResNet 1	0.56	0.61
4	ResNet 2	0.73	1.09

Table 3.4: Results of Experiment 3

From the table of accuracies, we observe the accuracy score of the ResNet 2 model = 0.73%. The final accuracy is given in this table 3.5.

ID	Logistic Regression Weight (%)	ResNet 2 Weight (%)	Final Accuracy (%)
1	50	50	71.5
2	40	60	71.8
3	60	40	71.2

Table 3.5: Final Accuracy Scores by Weights

As we can see from the table 3.5, the weights 40:60 give the highest accuracy score of 71.8%. This is because the weight assigned is higher for the ResNet 2 model which has higher accuracy than the Logistic Regression model.

3.6 Implementation Tools

I used Python in Jupyter Notebooks in Anaconda Environment and Google Colab Environment for the coding. Following are the libraries in Python which I used:

1. pandas
2. sklearn
3. seaborn
4. matplotlib
5. numpy
6. SimpleITK: Simple Insight Segmentation and Registration Toolkit is a powerful library for medical image analysis.
7. TensorFlow: Created by Google the library is used for training and inferring deep learning models. It allows for the easy creation of deep learning models.
8. PIL: Python Imaging Library
9. OpenCV: Open Source Computer Vision Library

4

Results and Discussion

4.1 Introduction

The research problem we are trying to solve is to enhance the accuracy of diagnosing Prostate Cancer using a multimodal dataset. We want to use a weight ratio to combine the accuracies of each mode to get a final accuracy of the ensemble model.

4.2 Presentation of Results

The machine learning model using tabular data and employing a logistic regression algorithm to predict whether the patient has clinically significant prostate cancer has an accuracy of 69.79% 3.9.

The machine learning model using tabular data and employing a support vector machine algorithm to predict the ISUP grade of Prostate cancer has an accuracy of 84.53% 3.13.

The deep learning model with the highest accuracy and best performance for predicting whether a patient has clinically significant prostate cancer is the ResNet 2 3.2 model which has an accuracy of 73% 3.4.

In the previous chapter's Weight Ratio Selection section, we selected the weight of 40:60 as the ideal weight for our ensemble AI model since it has the highest accuracy of 71.8% 3.5.

4.3 Analysis and Interpretation

We observe that our AI model's accuracy score is 71.8%, which is accurate for practical use in the real-world scenario. It would be beneficial for doctors and clinical staff to use this AI model to verify their diagnosis in very little time and with sufficient accuracy.

In the Literature Review Section, we saw the current gaps in research. They were concerned with the usage of smaller datasets with incomplete MRI sequences being used. To address this research gap we are using a large image dataset with complete MRI sequences. Another research gap was with the use of multimodal data which we are using in our model and addressing this gap. Another research gap was with regards to optimising models to have high accuracy. We are addressing this research gap as we are using grid search cross-validation for the SVM model and fine-tuning our deep-learning models to achieve better performance.

4.4 Limitations

The image dataset was very large (25GB) and extracting all images as .jpg from all .mha images and storing them was a challenge which resulted in a large amount of time spent in extracting the images.

The large number of images (75,000) required the usage of high-end GPUs such as A100 Nvidia GPU and T4 GPU which were purchased from Google Colab.

Due to the environment constraints of Google Colab, the runtime disconnecting leads to the loss of all variables and the need to repeat the entire experiment leading to time lost in repeating experiments.

Future research could include other pre-trained models such as VGG16, MobileNet V2, etc to improve diagnostic accuracy.

Due to the time constraint in completing the dissertation, further experimentation with multi-class classifier deep learning models to predict the Prostate cancer grade could not be completed.

4.5 Implications

The AI model we developed in this dissertation has the potential to be used in diagnosing clinically significant prostate cancer from the clinical information, and the MRI scans of the patient. It is a reliable and fast method to prevent misdiagnosis.

Deep learning models like the one we developed have a lot of potential for applications in real-world healthcare scenarios. There can be further improvements on metrics such as accuracy and recall by fine-tuning and optimising hyperparameters. Further, other transfer learning models can be used that can perform better for this use case.

5

Conclusion and Recommendations

5.1 Summary of Findings

Our AI model was able to diagnose a patient having clinically significant prostate cancer with an accuracy of 71.8%. It used a large multimodal dataset which leads to more reliable model performance, better test accuracy and lower test loss.

We successfully demonstrated several advanced Artificial Intelligence techniques such as Machine Learning and Deep Convolutional Networks to predict whether a patient has clinically significant Prostate cancer. We also predicted the grade of Prostate cancer using Machine Learning.

5.2 Conclusion

The AI model can help doctors and clinicians diagnose prostate cancer reliably and quickly. It can act as a second opinion for experienced doctors which can confirm their readings. There can be further improvements in hyperparameters, better models, and other transfer learning models.

5.3 Contributions

Our model has an accuracy score of 71.8%. The state-of-the-art model accuracy is 95.24% from the PCDM model Talaat et al. (2024).

We used a completely new prostate cancer dataset (PI-CAI grand challenge) PI-CAI (n.d.) with more patient data than previous research works. We used 1500 patients' data while previous research was limited to around 300 patients like the Prostate X challenge ProstateX (n.d.).

We also addressed several research gaps such as using complete MRI sequences for model training, using a multimodal dataset for model training, and fine-tuning models.

5.4 Future Work

Further research needs to be done to improve the model's accuracy on the image dataset. Other transfer learning models or newer models need to be experimented with.

In future, new models for predicting the ISUP grade of Prostate Cancer can be developed and experimented with to enhance the accuracy.

We also can explore using different MRI sequences in combination to achieve better accuracy as required by the project.

5.5 Final Remarks

I was satisfied with my learnings about logistic regression, support vector machine, convolutional neural networks, fine-tuning, and transfer learning from this research. I learned new Python libraries and the platform of Google Colab that facilitated this research.

I hope this research project helps to improve the accuracy of diagnosing prostate cancer models in future work.

References

Alshuhri, M., Al-Musawi, S. G., Al-Alwany, A. A., Uinarni, H., Rasulova, I., Rodrigues, P., Alkhafaji, A. T., Alshanberi, A. M., Alawadi, A. H. and Abbas, A. H. (2023), 'Artificial intelligence in cancer diagnosis: Opportunities and challenges', *Pathology-Research and Practice* p. 154996. 25, 26

AmericanCancerSociety (n.d.), 'Tests to diagnose and stage prostate cancer', `https://www.cancer.org/cancer/types/prostate-cancer/detection-diagnosis-staging/how-diagnosed.html`. Accessed: 02.07.2024. 10

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I. and Jemal, A. (2024), 'Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA: a cancer journal for clinicians* **74**(3), 229–263. 1

Brembilla, G., Dell'Oglio, P., Stabile, A., Damascelli, A., Brunetti, L., Ravelli, S., Cristel, G., Schiani, E., Venturini, E., Grippaldi, D. et al. (2020), 'Interreader variability in prostate mri reporting using prostate imaging reporting and data system version 2.1', *European radiology* **30**, 3383–3392. 11, 20

CancerResearchUK (n.d.), 'Tests for prostate cancer', `https://www.cancerresearchuk.org/about-cancer/prostate-cancer/getting-diagnosed/tests-for-prostate-cancer`. Accessed: 02.07.2024. 10

Corradini, D., Brizi, L., Gaudiano, C., Bianchi, L., Marcelli, E., Golfieri, R., Schiavina, R., Testa, C. and Remondini, D. (2021), 'Challenges in the use of artificial intelligence for prostate cancer diagnosis from multiparametric imaging data', *Cancers* **13**(16), 3944. 26

REFERENCES

Datacamp (n.d.), 'What is machine learning? definition, types, tools more', `https://www.datacamp.com/blog/what-is-machine-learning`. Accessed: 10.07.2024. 12

De Vente, C., Vos, P., Hosseinzadeh, M., Pluim, J. and Veta, M. (2020), 'Deep learning regression for prostate cancer detection and grading in bi-parametric mri', *IEEE Transactions on Biomedical Engineering* **68**(2), 374–383. 21

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009), ImageNet: A Large-Scale Hierarchical Image Database, *in* 'CVPR09'. 22

Gould, M. (2010), 'Giscience grand challenges: How can research and technology in this field address big-picture problems? arcuser, 13 (4), 64–65'. 2

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', pp. 770–778. 21, 23, 36

He, M., Cao, Y., Chi, C., Yang, X., Ramin, R., Wang, S., Yang, G., Mukhtorov, O., Zhang, L., Kazantsev, A. et al. (2023), 'Research progress on deep learning in magnetic resonance imaging–based diagnosis and treatment of prostate cancer: a review on the current status and perspectives', *Frontiers in Oncology* **13**, 1189370. 1, 19, 20, 21, 25, 26

IBM (n.d.*a*), 'What are support vector machines (svms)?', `https://www.ibm.com/topics/support-vector-machine`. Accessed: 15.07.2024. 14, 15

IBM (n.d.*b*), 'What is deep learning?', `https://www.ibm.com/topics/deep-learning`. Accessed: 16.07.2024. 15, 16, 17, 18

IBM (n.d.*c*), 'What is logistic regression?', `https://www.ibm.com/topics/logistic-regression`. Accessed: 13.07.2024. 13, 14

IBM (n.d.*d*), 'What is machine learning (ml)?', `https://www.ibm.com/topics/machine-learning`. Accessed: 10.07.2024. 12, 13

Ishioka, J., Matsuoka, Y., Uehara, S., Yasuda, Y., Kijima, T., Yoshida, S., Yokoyama, M., Saito, K., Kihara, K., Numao, N. et al. (2018), 'Computer-aided diagnosis of prostate cancer on magnetic resonance imaging using a convolutional neural network algorithm', *BJU international* **122**(3), 411–417. 21

Leake, J. L., Hardman, R., Ojili, V., Thompson, I., Shanbhogue, A., Hernandez, J. and Barentsz, J. (2014), 'Prostate mri: access to and current practice of prostate mri in the united states', *Journal of the American College of Radiology* **11**(2), 156–160. 11

MayoClinic (n.d.), 'Prostate gland', `https://www.mayoclinic.org/diseases-conditions/prostate-cancer/multimedia/prostate-gland/img-20006060`. Accessed: 01.07.2024. 9

Michaely, H. J., Aringhieri, G., Cioni, D. and Neri, E. (2022), 'Current value of biparametric prostate mri with machine-learning or deep-learning in the detection, grading, and characterization of prostate cancer: a systematic review', *Diagnostics* **12**(4), 799. 25

NIBIB (n.d.), 'Magnetic resonance imaging (mri)', `https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri`. Accessed: 15.06.2024. 1

Opengenus (n.d.), 'Inception v3 model architecture', `https://iq.opengenus.org/inception-v3-model-architecture/`. Accessed: 20.07.2024. 22, 23

Pecoraro, M., Messina, E., Bicchetti, M., Carnicelli, G., Del Monte, M., Iorio, B., La Torre, G., Catalano, C. and Panebianco, V. (2021), 'The future direction of imaging in prostate cancer: Mri with or without contrast injection', *Andrology* **9**(5), 1429–1443. 1

Pellicer-Valero, O. J., Marenco Jimenez, J. L., Gonzalez-Perez, V., Casanova Ramon-Borja, J. L., Martín García, I., Barrios Benito, M., Pelechano Gomez, P., Rubio-Briones, J., Rupérez, M. J. and Martín-Guerrero, J. D. (2022), 'Deep learning for fully automatic detection, segmentation, and gleason grade estimation of prostate cancer in multiparametric magnetic resonance images', *Scientific reports* **12**(1), 2975. 1, 20, 21, 25

PI-CAI (n.d.), 'The pi-cai challenge', `https://pi-cai.grand-challenge.org/`. Accessed: 15.06.2024. 2, 11, 28, 52

ProstateX (n.d.), 'Prostatex', `https://prostatex.grand-challenge.org/`. Accessed: 15.06.2024. 52

RadiologyAssistant (n.d.), 'Prostate cancer - pi-rads v2.1', `https://radiologyassistant.nl/abdomen/prostate/prostate-cancer-pi-rads-v2-1`. Accessed: 03.07.2024. xiii, 12

REFERENCES

Radiopaedia (n.d.*a*), 'Apparent diffusion coefficient', https://radiopaedia.org/articles/apparent-diffusion-coefficient-1. Accessed: 04.07.2024. 12

Radiopaedia (n.d.*b*), 'Diffusion-weighted imaging', https://radiopaedia.org/articles/diffusion-weighted-imaging-2. Accessed: 04.07.2024. 12

Radiopaedia (n.d.*c*), 'T2 weighted image', https://radiopaedia.org/articles/t2-weighted-image. Accessed: 04.07.2024. 12

Roboflow (n.d.), 'What is resnet-50?', https://blog.roboflow.com/what-is-resnet-50/. Accessed: 21.07.2024. 23, 24, 25

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L. (2015), 'ImageNet Large Scale Visual Recognition Challenge', *International Journal of Computer Vision (IJCV)* **115**(3), 211–252. 22

Saha, A., Bosma, J. S., Twilt, J. J., van Ginneken, B., Bjartell, A., Padhani, A. R., Bonekamp, D., Villeirs, G., Salomon, G., Giannarini, G. et al. (2024), 'Artificial intelligence and radiologists in prostate cancer detection on mri (pi-cai): an international, paired, non-inferiority, confirmatory study', *The Lancet Oncology* . 2, 11, 28

Sciencedirect (n.d.), 'Inception v3', https://www.sciencedirect.com/topics/computer-science/inception-v3. Accessed: 20.07.2024. 23

Simplilearn (n.d.*a*), 'Convolutional neural network tutorial', https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network. Accessed: 17.07.2024. 18, 19

Simplilearn (n.d.*b*), 'Top 10 deep learning algorithms you should know in 2024', https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm#how_deep_learning_algorithms_work. Accessed: 16.07.2024. 17, 18

spiceworks (n.d.), 'What is logistic regression? equation, assumptions, types, and best practices', https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/. Accessed: 13.07.2024. 13

Srigley, J. R., Delahunt, B., Egevad, L., Samaratunga, H., Yaxley, J. and Evans, A. J. (2016), 'One is the new six: The international society of urological pathology (isup) patient-focused approach to gleason grading', *Canadian Urological Association Journal* **10**(9-10), 339. 2

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016), Rethinking the inception architecture for computer vision, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 2818–2826. 22, 36

Talaat, F. M., El-Sappagh, S., Alnowaiser, K. and Hassan, E. (2024), 'Improved prostate cancer diagnosis using a modified resnet50-based deep learning architecture', *BMC Medical Informatics and Decision Making* **24**(1), 23. 2, 9, 21, 52

Turkbey, B., Rosenkrantz, A. B., Haider, M. A., Padhani, A. R., Villeirs, G., Macura, K. J., Tempany, C. M., Choyke, P. L., Cornud, F., Margolis, D. J. et al. (2019), 'Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2', *European urology* **76**(3), 340–351. 10, 11

Wang, Z., Liu, C., Cheng, D., Wang, L., Yang, X. and Cheng, K.-T. (2018), 'Automated detection of clinically significant prostate cancer in mp-mri images based on an end-to-end deep neural network', *IEEE transactions on medical imaging* **37**(5), 1127–1139. 21