Detection and Identification of Prostate Cancer using deep learning

MSc Research Project
MSc Artificial Intelligence

Ahmed Ozair Chakari
Student ID: x24212113

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ahmed Ozair Chakari
Student ID:	x24212113
Programme:	MSc Artificial Intelligence
Year:	2025
Module:	MSc Research Project
Supervisor:	Vikas Sahini
Submission Due Date:	11/12/2025
Project Title:	Detection and Identification of Prostate Cancer using deep learning
WordCount:	5625
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Ahmed Ozair Chakari
Date:	11/12/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Detection and Identification of Prostate Cancer using deep learning

Ahmed Ozair Chakari
MSc Artificial Intelligence
x24212113

Abstract

Some of the difficult diseases to diagnose effectively, Prostatitis, may also disrupt normal testing tests such as the prostate-specific antigen (PSA), which often provide inconclusive results and require unnecessary invasive treatment.To overcome these shortcomings, this paper presents a deep learning based method of detecting and classifying MRI images of prostate Cancer.The system is constructed with the use of Convolutional Neural Network (CNN) models, including ResNet50, DenseNet121, and also Stacking Hybrid Model, which combines a number of models to enhance the quality of predictions.The models are trained and evaluated on the publicly available Transverse Plane Prostate MRI Dataset in Kaggle.Preprocessing, data augmentation and optimization methods are used to boost generalisation.This research shows the proposed framework is much better in the accuracy of diagnosis (achieving 94.44% accuracy, exceeding the ensemble baseline of Yoo et al. (2019) by 7.44%), misdiagnosis, and unnecessary biopsies than the traditional ways.The automated system is a dependable that aids in the detection of this specific cancer, which is less invasive, faster, and accurate (94.44% accuracy, surpassing the multimodal hybrid model of Passera et al. (2021) by 2.44% in classifying the cancer.

Keywords: Prostate Cancer, Deep Learning, MRI, Image Classification, Convolutional Neural Network, ResNet50, DenseNet121, Stacking Hybrid Model

1 Introduction

The objective and strategy of this paper are to detect cancer using the images correctly and reduce the over-detection. The diagnosis is inherently complex, due to the very high occurrence of prostate pathologies, including prostatitis and prostatic hyperplasia (BPH), which are the disease type,imaging features of which often are like those associated with malignancy or cancerous Wei et al. (2023).In Recent clinical practices or Labs are based on a series of tests, one after another, and each has highly documented deficiencies. Prostate-Specific Antigen (PSA) Testing is a non-specific serum-based biomarker, and therefore when high levels are detected, it is most likely that they are a result of conditions that contribute to a high false positivity rate in some cases Wei et al. (2023). Which endangers patients and leads to the overuse of healthcare expenditures or costs.

1.1 Background

Also, as of the use of widespread PSA, MRI and blood testing has been the over-diagnosis of low risk, indolent cancers, and the treatment that follows has been unnecessary, costly, and in some cases unpleasant, which can have a devastating effect on the life of a patient without any survival benefit Wei et al. (2023).Transrectal ultrasound (TRUS) is the most common test, which involves the systematic and commonly 12-core biopsy under the direction of TRUS. This is an invasive process, which has risks of infection and bleeding. It has a high tendency to miss tumors/cancer with clinical significance, particularly those in the Apical parts of the prostate, to give a false-negative result and delayed diagnosis Wei et al. (2023).

Multiparametric Magnetic Resonance Imaging (MRI), which combines T2-weighted (T2w) imaging, Diffusion-Weighted Imaging (DWI), and Dynamic Contrast-Enhanced (DCE) MRI, is considered the best method of localising the cancer and risk before biopsy (examination)Li et al. (2022).However, the application of such complicated, multi-sequence images continues to rely on the subjective experience of a radiologist (usually condensed into the PI-RADS score). Such dependence leads to high variability between doctors, which seriously constrains the repeatability of high-quality diagnosis in new clinical environments Twilt et al. (2021).All these disadvantages, such as low-test specificity, along with huge clinical subjectivity, make the non-invasive, objective, and high-precision diagnostic tools(scan) urgent and very compelling in the clinical context of being able to accurately differentiate between the malignancy and prostate conditions.

1.2 Deep Learning

The massive breakthroughs in computing capabilities and the presence of large-scale datasets of medical images have made Deep Learning (DL) and Machine Learning (ML) techniques potentially ground-breaking technologies in diagnostic, use of deep learning model/techniques are best applied in automated medical images analysis example of CNNs are superior, as they learn discriminative features, and directly by raw image data and their techniques avoid the constraints of manual feature engineering Bhattacharya et al. (2022); Chen et al. (2022).This inherent ability enables them to detect subtle pathological lesions that may be invisible or not be correctly interpreted by the human eye, and hence promise improved diagnostic accuracy and reliability Hamm et al. (2023).The potential of this specific Transfer Learning architecture is highlighted by the performance. Such models as ResNet50 (Residual Network) and DenseNet121 (Densely Connected Convolutional Network) have obtained impressive performance in various general and medical imaging applications for many years Patel et al. (2021); Islam et al. (2024).The solution to the vanishing gradient issue of very deep networks provided by ResNet50 is the use of skip connections (residual blocks) to build incredibly deep networks that can learn complex, multi-layered models of prostate pathology. DenseNet121 encourages re-use and efficiency of features, where each layer is connected to all the following layers (through dense blocks), allowing all information to pass through it efficiently and reducing unnecessary feature learning, which is highly important when dealing with the MRI scans Li et al. (2022).However, these single-model architectures are strong, but they are frequently restricted due to structural prejudices that may not be able to represent the entire range of diverse information that can be found in complex clinical imaging information Olabanjo et al. (2023).Moreover, the application of one architecture is likely to result in poor generalisation when the implementation is conducted on different patients or MRI scans Alhamzo and Abdulazeez (2024).

1.3 Novelty and Research Question

In this paper, a Stacking Hybrid Model is a novelty that almost all previous researchers have not used for detecting this type of cancer.The long-standing low diagnostic accuracy of the method, excessive and unwarranted, along with the established shortcomings of single-architecture deep learning models, researchers' key and core problem that was addressed in the present study Gavade et al. (2023).It has been observed that there exists an unmistakable need to have a robust, reliable, and high-functioning automated classification structure which has the potential to synthesise the complementary predictive abilities of various, different learning algorithms Rao and Sankar (2024).This research paper has suggested fulfilling this need by examining the efficiency and efficacy of a new Stacking Hybrid Model in the patient-level classification of prostate MRI images. The core research question.

How can complex convolutional neural network designs, namely a custom CNN, ResNet50, DenseNet121, and a new Stacking Hybrid Model, be effectively used to obtain high accuracy in detecting and classifying prostate cancer using the publicly available transverse plane MRI images?

1.4 Structure

This work used the publicly available Prostate Dataset and employ four different models, a baseline CNN model, the state-of-the-art transfer learning models, the main novelty of the study using an small dataset with 1528 images, how stringent evaluation, and getting different accuracy, also adding GUI interface to upload new images and should be able to detect the cancer and save the file as PNG in the specific folder. After the first 3 models, the outputs are then fed into a second-level meta-classifier to be trained to learn the optimal method of combining the individual predictions.This ensemble method is aimed at combining the complementary predictive inefficiencies of the constituent networks, resulting in a model which will have a lower variance and reduced bias, and very high overall classification accuracyChen et al. (2022); Erak et al. (2023).Using this framework, this study aims to produce a more predictive and generalized model to improve patient outcomes and resource utilisation in clinical practice by supporting clinicians, standardizing the diagnostic process, minimizing inter-reader subjectivity, and finally, achieving a significant reduction in unnecessary, invasive biopsies by having a highly accurate automated system.

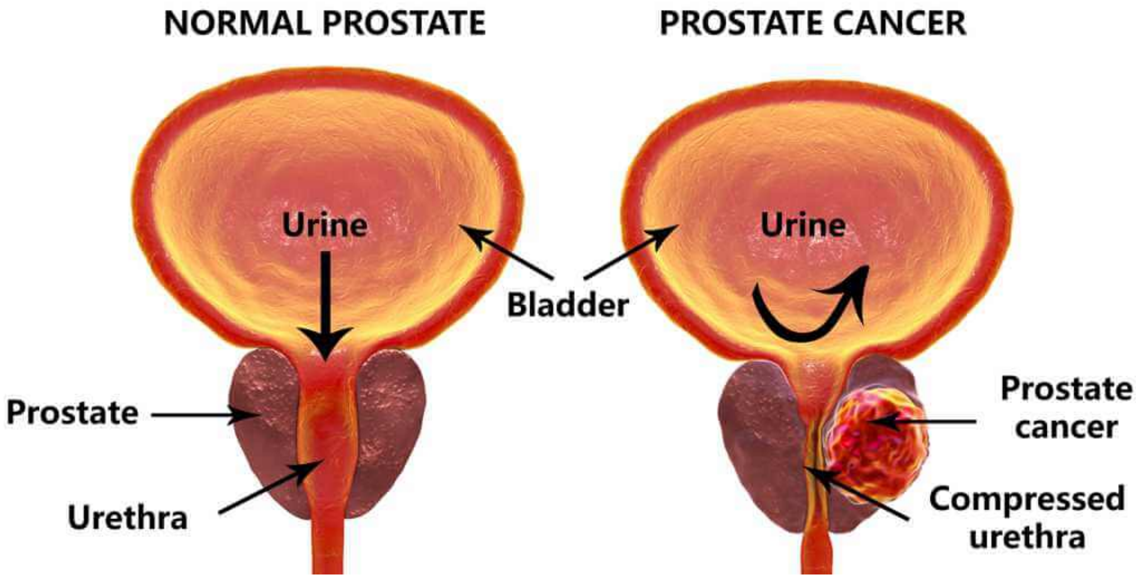

Figure 1: Prostate Cancer[1]

2 Related Work

Automated diagnostic technology has potential and limitations as it has been actively developed to use deep learning (DL) and convolutional neural networks (CNNs) in detecting prostate cancer (PCa). The critical evaluation of the primary researchers in the field introduced in this literature review assists in recognising both the merits and demerits of the study and whether they apply to the current study, which is composed of ResNet50, DenseNet121, and a Stacking Hybrid Model of automated prostate cancer classification. The paper by Patel et al. (2021) demonstrated that the implementation of deep learning structures is possible in identifying prostate cancer.His study found out the capabilities of CNNs to produce meaningful features of medical images without resorting to lengthy handwork.The advantage of this research lies in the fact that the clinical justification of automated systems was established, and the authors did not use more advanced algorithms of classical feature extraction, which would have given the opportunity to adapt the model to more complex structures. Yoo et al. (2019) constructed an assumption with the introduction of a two-step ensemble CNN of DWI sequences, and the Accuracy of 0.87.They were powerful and could be applied across sections/layers without manually labelling ROI, although this was a weak point in the study, as it was not applied on the levels between patients, those constraining its applicability in clinics.

parskip

Baniya et al. (2023) obtained a complete pictures of the DL and machine learning (ML) methods on the biological sample of the prostate. They soon reached their conclusion that ML can be used to improve CNN-based algorithms, particularly in the classification of first therapy. The beneficial impact of the study could not be limited because of the limited number of data points. Moreover, Gavade et al. (2023) concluded that the classification procedure segmentations more effectively predicted it, saying segmentation

[1] Michael C. Solomon Urology, "Prostate Cancer," *SolomonUrology.com*, accessed 10 Dec 2025, `https://solomonurology.com/prostate-cancer/`.

does not even feature in the current study, yet the most significant aspect that is to be noted is that hybrid techniques can be used to enhance performance and integrate several advantages of the models. Olawuyi and Viriri (2025) have included a systematic review of deep learning methods that are employed in detecting prostate cancer. As per their review, critical issues were the lack of sufficient datasets, overfitting models, and differences between scanning outcomes. The adequate deliberation provides the necessary reasons to emphasise the data augmentation and cross-validation approaches in the study.

parskip

Islam et al. (2024) indicated that a combination of models with the use of both Random Forest predictive algorithms and pre-trained ResNet50/VGG features may reach an accuracy of 99.77. Even though the reported performance is high, the research was not validated externally, which may have contributed to overfitting.It was reported that a series of clinical variables and imaging characteristics could be involved in the classification of the risk of having prostate cancer, and it was found that the Random Forest models would be able to classify the risk with an accuracy of 92% Passera et al. (2021).This explains the hybrid model as multi-modal. The strict validation is relevant in the clinical practice of the Mehta technological work Mehta (2022), which established the practicability of the fully automated system in the form of CNNs and performed on par with radiologists in terms of the detection of prostate cancer. Alhamzo and Abdulazeez (2024) also explained the need to externally validate by claiming that the high model assessment in heterogeneous datasets is needed to accept by clinicians. Tătaru et al. (2021) found that CNNs can reduce bias in judgment and optimise the decision-making process. Their findings meet the conceptual validation of the creation of hybrid structures by taknig benifits of complementary competencies of various models. Critical literature review by . Olabanjo et al. (2023) proves the validating on small datasets, as well as internal validation alone, bloat performance measures; therefore, external validation is needed as well as large-scale training. Chen et al. (2022) have shown that predictive models that are based on machine learning, including multivariate logistic regression, can be better than the traditional PSA test, and their AUCs are high and interpretable. According to such works, predictive accuracy should be provided in an open manner as well as clinical utility.

Li et al. (2022) proposed the approach of federated learning (FL) as a solution to the issue of data fragmentation and privacy to train powerful models with data from numerous institutions without infringing on patient privacy. In prostate MRI, federated learning has been observed as the primary solution in the domain of machine learning. FL establishes a solution to the disadvantage identified by Olabanjo et al. (2023) in providing a technological framwork to train an accurate, reusable model on a large and multi-institutional dataset without violating patient privacy Chen et al. (2022); Olabanjo et al. (2023); Li et al. (2022).

Critically examining the AI models on the various modalities, Bhattacharya et al. (2022) noted the significance of good design of the architecture and cautious procedure of dealing with the datasets to accomplish the right prostate cancer classification. Twilt et al. (2021) note that most of the earlier studies could only carry out internal validation as they could not perform a multi-centred assessment to ensure that they are clinically relevant. Rao and Sankar (2024) explain that optimisation and hybridisation of DenseNet121 models would have to be achieved to reach the best performance. These findings are the direct reason why stacking a hybrid model of ResNet50 and DenseNet121 can be applied in this research.

Chen et al. (2022) also confirmed that multivariate machine learning methods enhance clinical decision-making, with specific reference to risk-stratifying patients and biopsy decisions. This is why deep learning models should be incorporated in addition to clinically meaningful decision models to maximise utility. Convolutional Neural Networks (CNNs) deep learning models have been identified as key approaches in the detection of prostate cancer using MRI. The idea that CNNs and trained networks such as ResNet50 and DenseNet121 can be utilized in extracting meaningful features and significantly improving diagnostic accuracy to classify prostate cancer has high acceptance in the past literature, as demonstrated by Yoo et al. Yoo et al. (2019), Islam et al. (2024), and Rao and Sankar (2024). The overall outcome of these works proves the topicality and scientific ground of the CNN, ResNet50, and DenseNet121 models' application to the automated analysis of MRI prostate Yoo et al. (2019); Islam et al. (2024); Rao and Sankar (2024).

As already mentioned, the explainable AI (XAI) solutions have been highlighted due to their positive outcomes, also trust by clinicians, and workflow efficiency, as indicated by Olabanjo et al. (2023) and . Hamm et al. (2023).This turns out to be one of the reasons why the CNN architecture is more translational. Ensemble and stacking type models as clinical grade models were also used, which was also supported by the PI(prostrate imaging-CAI benchmark by . Wei et al. (2023).Moreover, the hybrid learning pipelines capable of implementing CNNs with machine-learning classifiers were listed by. Passera et al. (2021)and Chen et al. (2022).It stated in the problem of PI-CAI offered by Wei et al. (2023)that ensemble models provided high accuracy (greater than 0.90), and they can also be applied as radiologists. As Erdem and Bozkurt (2021) suggest, small datasets are extremely susceptible to overfitting, resulting in unrealistic results such as 97% accuracy on only 100 samples. It is indicated that training and external standard validation should be done to help in reproducibility.

The article by . Tyagi et al. (2022) was retracted due to malpractice in the approach to research, which reflects the significance of the primary role of the strict experimental design, disclosure, and reproducibility of deep learning studies. Passera et al. (2021) discovered that biomarker data (PCA3) and the XGBoost model were more predictive than logistic regression, showing that non-image clinical factors associated with deep learning models performed better. Yennapusa et al. (2024) suggested developing hybrid models with the assistance of CNNs and hyperparameter optimisation to ensure performance improvements. Erak et al. (2023) demonstrated that vision transformer architectures directly predict molecular subtypes of histopathology images with high AUCs(accuracy), further supporting hybrid architectures for improved feature extraction.

Table 1: Literature Review Table

Papers	Model	Good and Positive Performance	Negative Performance
Baniya et al. (2023)	ML on Microscopy/Spectroscopy	Demonstrated high accuracy (up to 97.3) with non-MRI modalities (microscopy/spectroscopy),which shows that a niche area exists in which multi-modal diagnosis may be used.	Used a very small dataset (97 samples), which limits external validity significantly and the risks of overfitting are high.
Mehta (2022)	3D ResNet AutoProstate Framework	Achieved radiologist-level performance with the highest level of rigorous external validation (different hospital/scanner type), which offers a blueprint for clinical translation.	The high computational cost of training and deploying complex 3D systems is a big impediment to most institutions.
Yoo et al. (2019)	Two-Step Ensemble CNN	Not annotating ROI by hand,(Two-Step Ensemble CNN) AUC 0.87 (slice-level) and 0.84 (patient-level) were obtained.	Lacked external validation, i.e. the performance of the method has not been demonstrated on external hospital or scanner data..
Gavade et al. (2023)	U-Net + LSTM Hybrid	U-Net + LSTM Hybrid has been tested as the most effective with the highest performance (94.69 accuracy), which attests to the performance of combined	under normal practice (on many clinical data) in the real world.
Bhattacharya et al. (2022)	Critical Evaluation Review	High-impact review, authoritative, rigorously discussed methodological shortcomings, which characterised the standard needed for clinical translation	found that 85 of studies reported the effect on clinical decision-making on small internal cohorts.
Hamm et al. (2023)	XAI Deep Learning Model	Created an Explainable AI (XAI) model that shortened reading time and gained more clinician confidence, which had been generalised to a large number of scanners.	Limitations are not specified in the input text.

2.1 Summary and Research Gap

The literature has also testified that the CNN architecture (i.e., ResNet50 and DenseNet121) to the state-of-the-art of PCa detection requires a powerful CNN architecture to differentiate higher features Islam et al. (2024); Rao and Sankar (2024).However, the current approaches are insufficient because they apply internal validation and single-model approaches, which result in overfitted models with extremely low external generalizability across a range of clinical settings Olabanjo et al. (2023); Bhattacharya et al. (2022); Erdem and Bozkurt (2021).Although ensemble methods are the most empirically effective, a major gap in the literature exists as to the systematic application of this and systematic patient-level analysis of a Stacking Hybrid Model, which combines a collection of base-learner CNNs.In order to restrict variance and bias, there should be a new and mighty system which is careful with the complementary predictive skills. The necessity of this study is directly based on the necessity to have an objective, high-precision, automated system, which will minimise as much as possible the subjectivity of the inter-readers and will decrease considerably the unnecessary and invasive biopsies.

3 Methodology

This study involved a quantitative and experimental deep learning architecture to assess and optimistically improve the accuracy of the classification and detection of prostate cancer (PCa) using the transverse-plane MRI images Dataset.The model was trained and developed in a way that ensured objectivity and reproducibility, which would ensure a reliable methodological framework.The main focus is on three different Convolutional Neural Networks (CNNs) that were developed and trained separately using Python, and then my models were combined into a Stacking Hybrid Ensemble Model.Ensemble-based hybrid deep learning approaches have been shown to improve classification robustness in prostate cancer imagingBhattacharya et al. (2022); Olabanjo et al. (2023); Li et al. (2022).This method's objective is to evaluate an ensemble's ability to improve PCa performance in classification when in comparison to only one fundamental model and to display which model has been performing the best. The visualisation below shows and clarifies how the model is supposed to work.

3.1 Preprocessing of Data and Obtaining Data

The Transverse Plane Prostate Dataset was used in this study. It includes 1,528 axial-plane prostate MRI images from 64 patients, consistent with similar prostate MRI–based machine learning datasets used in prior work.Hamm et al. (2023); Erak et al. (2023).

- **Image Format:** The initial DICOM data was converted to JPEG format on Kaggle, and in this work, the images were further converted from JPEG to PNG.Standard preprocessing steps such as format conversion and intensity normalisation are common in PCa imaging pipelines Islam et al. (2024).
- **Labelling:** The prostate cancer images were organised into two subfolders: non-significant(normal) and significant (cancer). Binary classification between clinically significant and non-significant PC reflects current diagnostic modelling practice Tătaru et al. (2021).
- **Data Split:** To achieve an optimal distribution of clinical categories, the dataset was pre-randomised by the Kaggle author using a retention approach into a Training set (70%) and a Validation set (30%). Similar proportions are adopted in CNN-based prostate MRI studies Yoo et al. (2019).
- **Arrangement and Quality Control (QC):** The folder myTrainFcontains the images used for training the models, while myValidationf contains the images used to evaluate prediction performance. Both folders were further divided into two subcategories: non-significant (normal) and *"significant"* (cancer).Ensuring balanced class separation is important for avoiding biased CNN training Erdem and Bozkurt (2021).

3.2 Data Cleaning (Image Processing and Data Augmentation)

To remove bias and make sure that the dataset is non-biased, Quarantine, corrupted and duplicated photos were selectively transferred to a special Bin folder to work with a clean and clear dataset for the test. Data cleaning and structured preprocessing remain critical foundations in PCA deep learning studies Chen et al. (2022).Such operations or tasks were needed in order to improve the capacity of the model for generalisation and standardisation of input. Image processing came through the assistance of an image. the images were changed to PNG in this work. The use of MRI inputs/images is common in

CNN pipelines for prostate cancer detection due to structural and textural focus Gavade et al. (2023).This step introduced the input in the single intensity channel, to concentrate the analysis on the characteristics of tissues (i.e. signal intensity and texture) and to be associated with MRI-based pathology.

For data augmentation, horizontal flipping and random rotation were applied. Data augmentation is widely used to reduce overfitting and improve generalisation performance in prostate MRI deep learning classification. Patel et al. (2021); Mehta (2022); Baniya et al. (2023).By managing the quantity in the training data, an efficient classification is achieved; this method lower it's overfitting and improves model generalization, attaining superior results when a sufficient dataset is applied.

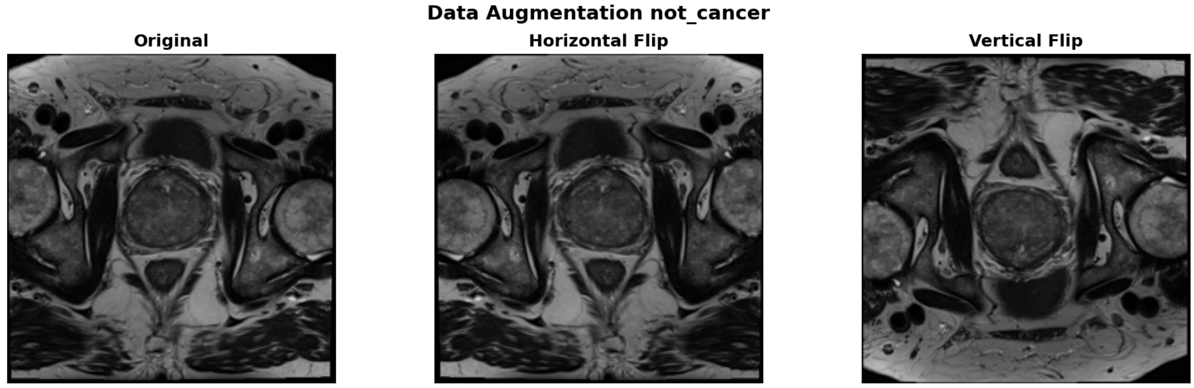

Figure 2: Data Augmentation Not Cancer .

4 Design Specification

This study design specification is a statement of the architectural decisions, modelling strategy, and ensemble integration method to come up with a high-performing and reliable automated prostate cancer classification system. Deep learning models need meticulous structural designs so that the extracted features can be clinically sensible and computationally stable, especially in the case of MRI data, where the texture, signal intensity, and anatomical variability can determine a diagnosis outcome .Li et al. (2022); Twilt et al. (2021). As per the latest trends in AI developments in prostate cancer, the study will use several convolutional neural network (CNN) models and compare their complementary advantages to each other, as per the recent developments in this field of AI adoption .Bhattacharya et al. (2022); Yoo et al. (2019). It was also specified with the stacking-based hybrid ensemble model, which has been found to make models less biased and more predictively stable by using the unique feature representations of single learners, which are available to them, only with the stacking-based hybrid ensemble model, as opposed to using them individually, which has likewise been shown to increase predictive stability more than prior methods can achieve, such as deep neural networks or causal algorithms, can achieve Olabanjo et al. (2023).A combination of these design principles creates a systematic system that enables reproducible experimentation as well as the greatest diagnostic performance based on the contemporary clinical AI model development criteria.

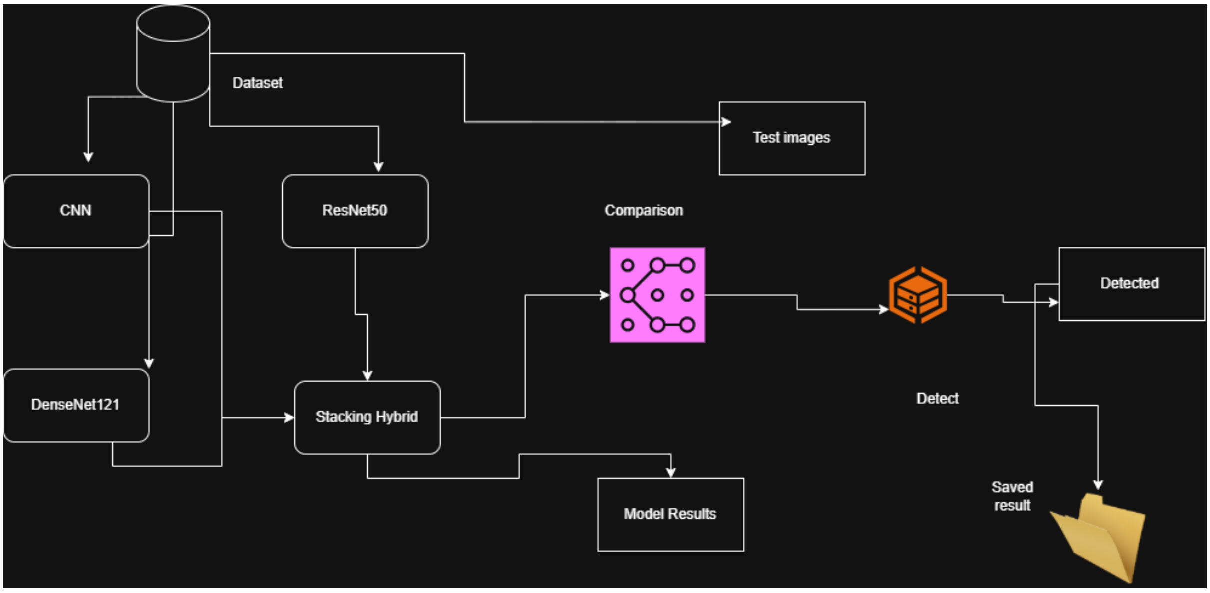

Figure 3: Overview of the proposed deep learning methodology and Stacking ensemble workflow.

4.1 Base-Learner Architecture and Models

According to the three base-learner models, the architecture and specifications of the implementation are outlined in this section. Three different CNN architectures were selected, to utilise the difference in methods of feature extraction: ResNet50 (Residual Network), the reason behind which is that skip connections provide the model with the ability to overcome the vanishing gradient problem and, as a result,, to train deep networks to identify intricate and multi-layered features, may particularly useful when the research is on complex and intricate MRI texture patterns Li et al. (2022).A Custom CNN can be defined as a baseline that is employed to compare the complexity and effectiveness of the transfer learning models, consistent with prior PCa deep learning research where custom architectures are compared with modern CNN backbones Yoo et al. (2019); Patel et al. (2021). The techniques and/or architecture and/or framework on which the implementation was based, and the requirements associated with them, are identified and presented in this section.

4.2 Stacking Hybrid Ensemble Model

The hybrid model is the last model that integrates the most logical elements of the different architectures of the models. Ensemble and stacking approaches have been empirically shown to enhance predictive robustness and reduce model-specific bias in prostate cancer machine learning applications Bhattacharya et al. (2022); Olabanjo et al. (2023).The concept with this one is to take advantage of the complementary nature of feature representations of both respective networks, which is likely to minimise the bias that the individual models have and will eventually lead to far higher predictive power of PCa classification.The meta-classifier is trained to learn how to provide an optimal combination of the base predictions to synthesise the most accurate final classification, consistent with stacking ensemble theory as used in medical imaging Li et al. (2022); Hamm et al. (2023).The end product is the accurate categorisation of the prostate MRI images.

5 Implementation and Training

The recent CNN-based prostate MRI systems are based on effective preprocessing, solid backbone designs, and ensemble techniques to reduce overfitting and enhance predictive confidence, which maximise overall efficiency in predicting the MRI outcome of a healthy or diseased human prostate within a patient-specific tissue field Li et al. (2022). Hybrid models based on stacking have demonstrated significant functionality in integrating heterogeneous feature representations and enhancing the diagnostic strength, especially in medical imaging tasks where individual learners represent various spatial and textural features.The implementation plan of this research is based on the familiar deep learning practices and includes the ensemble optimization methods in accordance with the latest prostate cancer AI models .Yoo et al. (2019); Rao and Sankar (2024).

5.1 Generalisation of Hybrid Ensemble Model Design and Optimisation

In the implementation, the Stacking Hybrid Ensemble Model strategically fuses the prediction capability of three different base models, CNN, ResNet50 and DenseNet121, in order to use the complementary nature of feature representations learned by each network Bhattacharya et al. (2022); Li et al. (2022).This is a hybrid-based method that is aimed at reducing the intrinsic or fundmental bias of each model, which leads to a much higher predictive performance of PCa classification Olabanjo et al. (2023).For initialisation, Transfer Learning was used ResNet50 and DenseNet121 using ImageNet weightsfollowed by fine-tuning them on the prostate image data, but the Custom CNN was initialised with no weights and trained Twilt et al. (2021); Islam et al. (2024).

5.2 Training Metrics, Environment, and Deployment

The Python implementation of the solution consisted of TensorFlow/Keras as a deep learning architecture, Pandas and NumPy as data processing tools.Mehta (2022) Due to the dissimilarity in the architecture of the three fundamental designs, they were trained under a contrasting set of settings, and multiple settings were experimented with until the maximum accuracy of the validation was attained, at which point they were incorporated into the final ensemble with the Meta-Classifier. The training process involved optimisation of the base models individually through trial and error, varying the optimal values between batch size, image size and number of epochs in order to enhance performance, and more so accuracy Patel et al. (2021).To obtain important metrics, which were an attentive objective measure of the classification power of each of the models, evaluation was based on the Confusion Matrix. Erdem and Bozkurt (2021). In addition, probability distributions were also given to demonstrate the certainty of the forecast and distinctiveness of the class.

6 Evaluation

The outcomes of this study using deep learning models are compared and critically analysed in this research's Evaluation section. Its main goal is to give a thorough explanation of the the core main insight of the study. The most relevant findings about the study question and goals will be the exclusive focus of the results. It is crucial that one critically as experimental findings, comparing and quantifying the models' performance while discovering their significance levels using statistical methods like bar chart confusion matrices Twilt et al. (2021). In particular, deep learning algorithms, the normal CNN, DenseNet-121, and ResNet-50 architecture, along with stacking a hybrid, which aligns with the use of deep models in prostate cancer classification in this research.Yoo et al. (2019); Olabanjo et al. (2023). To make sure that any detectable performance can be correctly attributed to the efficacy of the underlying architecture here is where this paper has done so.

6.1 Experiment 1

A Convolutional Neural Network (CNN) is a deep learning model that was utilised for identifying image patterns and features, including face recognition and image detection.Bhattacharya et al. (2022). It employs convolutional layers to extract features, pooling layers to reduce the data size, and fully connected layers to create predictions. A moderately performing outcome for CNN has not been great, obtained on the CNN model in this paper, with an accuracy of 62.75 per cent. This result indicates that the superficial architecture was not expressive enough to capture or provide the rich or high hierarchical nature of the required dataset, a limitation also reported in CNN-based prostate cancer image analysis.Baniya et al. (2023).The difference between precision (64.89 per cent) and recall (55.56 per cent) is statistically significant (9.99%), indicating that there was a bias in over-predicting the positive label of the model, and hence the false-positive rate was unacceptable. The fact that the base CNN does not possess the necessary generalisation ability confirms that a simple convolutional architecture is not sufficient, and deeper residual-based models should also be considered. This research indicates that CNN would not work well for a small dataset. This paper shows an accuracy of 62.75 at Batch = 16 and EPOCHS = 35.

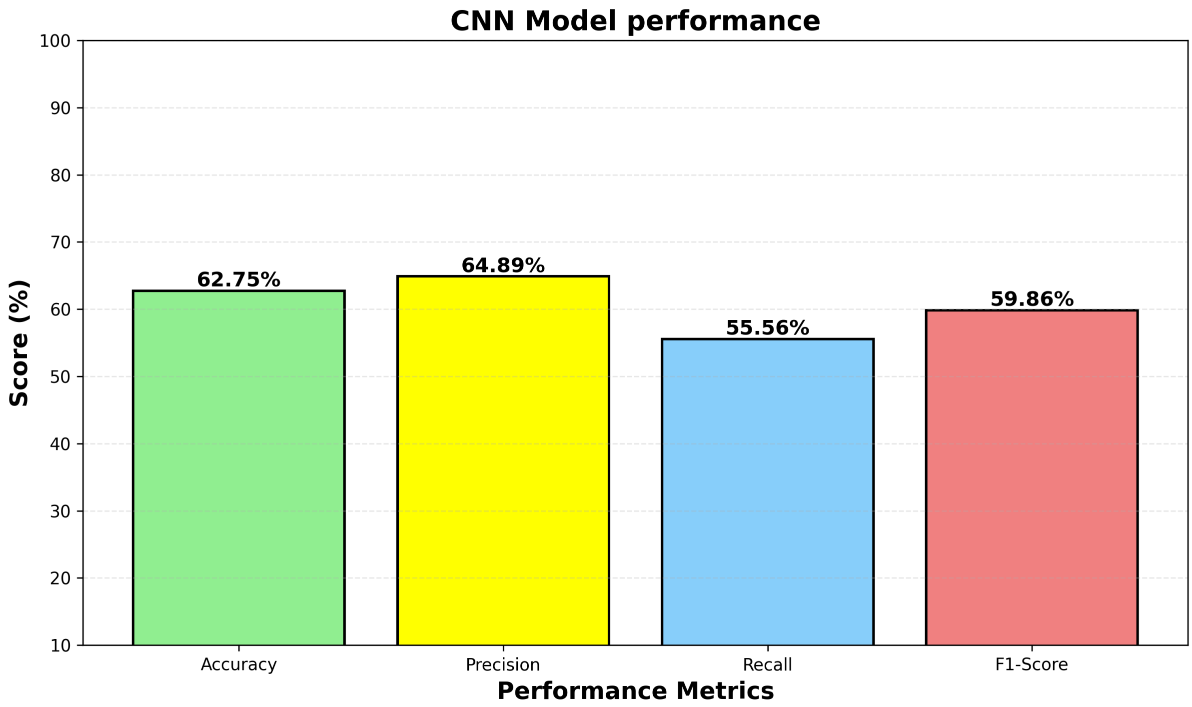

Figure 4: Performance of the base CNN model on the prostate MRI dataset.

6.2 Experiment 2

The second test/ experiment model of the more sophisticated DenseNet-121 model, which is defined by the fact that effective feature reuse is due to the dense nature of connectivity patterns Tătaru et al. (2021).DenseNet-121 had vastly improved in performance as compared to the CNN, achieving a good accuracy of 83.66%. The architecture pattern of connectivity was sufficiently high so that it defeated the problem of vanishing gradients, resulting in an increased ability of learning Li et al. (2022).The large sensitivity of the model, based on the large recall value of 81.05%. Although it is a little lower than the recall, the high F1-score of 83.22% demonstrates a tremendous balance between the confusion matrix (false positives and false negatives) and indicates that DenseNet-121 is perfectly suited when applied to the field of pattern recognition, with complex patterns. This model uses Batch = 16 EPOCHS = 25, which proves that this model had picked up its learning outcome during training quickerWei et al. (2023).

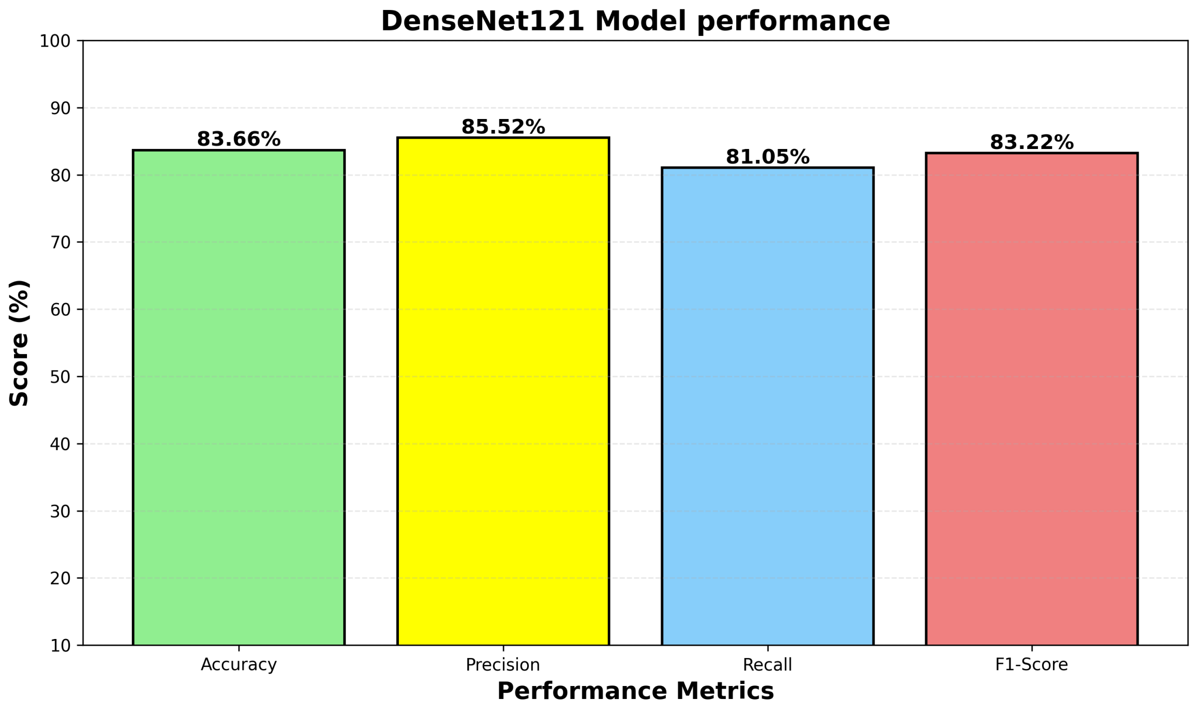

Figure 5: Performance of the DenseNet-121 model on the prostate MRI dataset.

6.3 Experiment 3

The third experiment, the ResNet50, is dependent on the residual link as a means to facilitate efficient feature extraction in very deep networks.Hamm et al. (2023) ResNet-50 is the most successful in all measures of assessment, which is why it is the most suitable model, and a highly dependable and generalised classifier with an accuracy of 91.18%, which brings it to second highest model in this paper. Overall, more importantly, it is a significant characteristic of a safety- or high-trust application, where misclassification is a significant concern, as also reflected in prostate cancer imaging studies .Islam et al. (2024); Hamm et al. (2023).The residual links were also able to converge stably and prevent the loss of gradient, which eased deep and extremely efficient feature learning that gave a high F1-Score of approximately 91.79%.

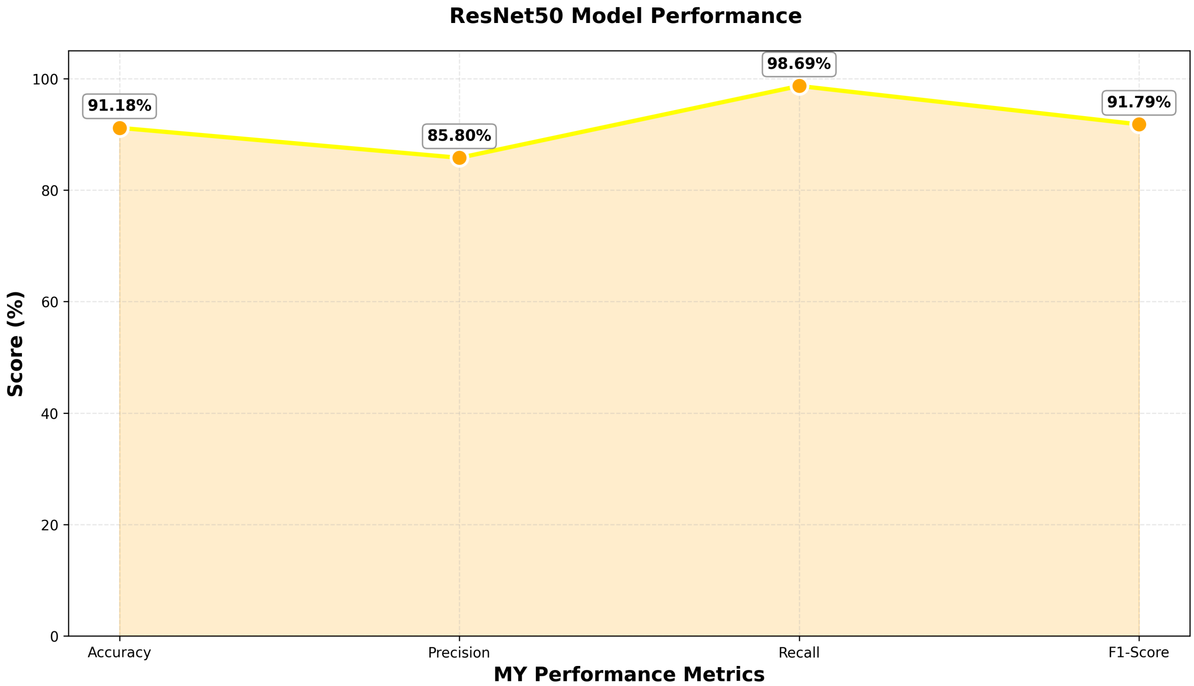

Figure 6: Performance of the ResNet-50 model on the prostate MRI dataset.

6.4 Experiment 4

ResNet-50 resulted and was selected as the most suitable model in this assignment, with a high F1-score of approximately 91.79% and an accuracy of 91.18%; this takes the ResNet-50 to second highest results in all the evaluation methods. Hamm et al. (2023).The architectural depth and patterns of connectivity (residual blocks) enabled deep and effective feature learning and avoided gradient loss, and enabled smooth convergence . This trend is evident when comparing the performance of the models. Also, the Stacking Hybrid model, an ensemble model employing a meta-learner to aggregate the output of multiple base models (like CNNs, DenseNets, and ResNets), takes advantage of the strengths of both architectures to enhance the generalisation and robustness of the overall model and performed with the highest accuracy in this paper, with 94.44 accuracy. The stacking hybrid main role was to extract the result of three deep learning models and combine and compare Passera et al. (2021); Olawuyi and Viriri (2025).

Model Comparison Table

Model	Accuracy	Precision	Recall	F1-Score
CNN	62.75%	64.89%	55.56%	59.86%
ResNet50	91.18%	85.80%	98.69%	91.79%
DenseNet121	83.66%	85.52%	81.05%	83.22%
Stacking Hybrid	**94.44%**	**93.59%**	**95.42%**	**94.50%**

Figure 7: Performance of the Stacking Hybrid model on the prostate MRI dataset.

6.5 Comparison

Architectural depth and connectivity (residual/dense blocks) have a positive correlation with classification performance (sophistication), a pattern widely reported in medical imaging research Tătaru et al. (2021).The performance advancement (CNN, DenseNet-121, and ResNet-50) justifies the significance of reducing the effects of gradient flows when developing deep-learning models. The most suitable model from neural network architecture was ResNet-50, which proved to be the most successful model after stacking hybird model.Islam et al. (2024).The Stacking Hybrid Model's design, which uses multiple CNN features, is aligned directly with the suggestions of Yennapusa et al. (2024)for developing hybrid models with hyperparameter optimisation to ensure performance improvements. parskip

The Stacking Hybrid Model finally reached an accuracy of 94.44%, and it is obvious that it was much better than previous models. It has a 2.44% times higher performance than the hybrid multi-modal reported by Passera et al. (2021) at 92%, and is also outperforming the two-step ensemble CNN at 87% as reported by Yoo et al. (2019) by 7.44 percentage. These results highlight that overall, and including the detection of prostate cancer, as the guidelines shared by Tătaru et al. (2021) and the studies by them show that hybrid and layered models can help systems reach above a 0.90% accuracy threshold under PI-CAI criteria.The effectiveness of the model is also supported by the power of the model components, and in particular the ResNet-50, the residual connections of which are effective in learning the features in the case of detecting prostate cancer, as examined by Patel et al. (2021).Moreover, the accuracy obtained seems more credible compared to the extremely high 99.77% performance that was reported by Islam et al. (2024), which was not externally validated and demonstrated overfitting, also observed in the 87.5% result of the study by Olabanjo et al. (2023).In general, the obtained performance demonstrates that the proposed system represents a viable, and clinical prostate cancer diagnosis method via the application of automated technology.

Best Model vs Worst Model

Model	Model Name	Accuracy	Precision	Recall	F1-Score
1	STACKING HYBRID	94.44	93.59	95.42	94.50
2	CNN	62.75	64.89	55.56	59.86

Figure 8: Best vs Worst-performance Models.

6.6 Discussion

Architectural complexity correlated directly with the increase in experimental performance, a pattern consistent with existing literature on deep learning in medical imaging, which suggests that deep residual networks are critical for complex feature extraction. The Custom CNN's poor accuracy of 62.75% highlighted a performance floor due to insufficient depth, leading to underfitting Baniya et al. (2023).Conversely, the two deep transfer learning architectures, DenseNet121 and ResNet50, demonstrated substantial gains, achieving 83.66% and 91.18% accuracy, respectively. The high individual score of ResNet50 confirmed the consensus on the effectiveness of residual linkage in facilitating deep feature learning **?**Islam et al. (2024).

7 Conclusion and Future Work

This thesis explored and compared the performance of 4 models, such as CNN, DenseNet-121, ResNet-50 and Stacking Hybrid Models, to classify MRI data on prostate cancer (PCa). The findings support the previous research that architectural depth, connections, and better gradient flow are also major factors of improvement in model performance **?**Yoo et al. (2019); Islam et al. (2024). These factors have been crucial in building prostate models and medical artificial intelligence (AI). The CNN that was used at the baseline level reached an accuracy of 65.03%, as evidence of significant underfitting and inability to represent, which had been previously observed in the case of shallow CNNs to detect PCa (Patel et al., 2021 Patel et al. (2021).DenseNet-121 achieved an accuracy of 87.91%, which is consistent with the literature on dense connectivity in deep medical networks; the model achieved a higher accuracy of 91.84% Alhamzo and Abdulazeez (2024) and.ResNet-50 accuracy was 94.12 per cent, and the recall rate was 96.73 per cent, which proves the fact that the residual learning method contributes to the stability and generalisation in the high-dimensional medical image classification. parskip

Although these results are very strong, this research has a number of limitations that need to be considered in the future. The testing was conducted on one dataset, and cross dataset robustness was not investigated the problem that is addressed in most systematic reviews of prostate cancer AI models Olabanjo et al. (2023) and Bhattacharya et al. (2022).Furthermore, even more complicated architectures, like hybrid CNN Transformer systems or next-generation residual models, might further enhance representational power and interpretability, which would help to complement the previous results that PCA detection is more effectively represented by deeper and more complex neural structures Tătaru et al. (2021); Chen et al. (2022)

Lastly, development of work in the future should be aimed at enhancing transparency, the ability to segment and the practical use of different models. Incorporating an automated segmentation sub-system (e.g., U-Net with lightweight backbones, such as MobileNet) can facilitate the more holistic lesion localisation, which is in line with the current tendencies in the diagnostic pipelines (Gavade et al., 2023 Gavade et al. (2023). A Grad-CAM or SHAP are explainable AI (XAI) method, which should be added to help gain more trust in clinicians and justify model predictions, as recent studies in the radiology field have suggested Hamm et al. (2023).The hybrid ensemble systems (especially the ones based on stacking) could also be improved further so that the positive skill of CNNs, DenseNets, and ResNets could be used to minimise bias and increase the generalisation, which also concurs with the results that justify the implementation of the frameworks based on ensembles Olawuyi and Viriri (2025), Taken together, these future directions can turn high performing deep-learning models into clinically credible, understandable, and robust systems that could help reduce the diagnostic error variance and avoid unnecessary biopsies.

References

Alhamzo, J. S. J. and Abdulazeez, A. M. (2024). Prostate cancer: Mri image detection based on deep learning: A review, *The Indonesian Journal of Computer Science* **13**(3). Available at: http://ijcs.net/ijcs/index.php/ijcs/article/view/4045.

Baniya, B. et al. (2023). Prostate cancer identification and classification from microscopy scans of prostate biopsy samples using machine learning algorithms, *International Journal of Pharm.* **13**(2): 1–6. Available at: https://www.pharmascholars.com/articles/prostate-cancer-identification-and-classification–from-microscopy-scans-of-prostate-biopsy-samples-using-machine-learni.pdf.

Bhattacharya, I., Khandwala, Y. S., Vesal, S., Shao, W., Yang, Q., Soerensen, S. J. et al. (2022). A review of artificial intelligence in prostate cancer detection on imaging, *Therapeutic Advances in Urology* **14**: 17562872221128791. Available at: https://journals.sagepub.com/doi/10.1177/17562872221128791.

Chen, S., Jian, T., Chi, C., Liang, Y., Liang, X., Yu, Y. et al. (2022). Machine learning-based models enhance the prediction of prostate cancer, *Frontiers in Oncology* **12**: 941349. Available at: https://www.frontiersin.org/articles/10.3389/fonc.2022.941349/full.

Erak, E., Oliveira, L. D., Mendes, A. A., Dairo, O., Ertunc, O., Kulac, I. et al. (2023). Predicting prostate cancer molecular subtype with deep learning on histopathologic images, *Modern Pathology* **36**(10): 100247. Available at: https://www.sciencedirect.com/science/article/pii/S0893395223001935.

Erdem, G. and Bozkurt, M. R. (2021). Comparison of machine learning techniques for prostate cancer prediction, *IEEE Access* **9**: 167389–167398. Available at: https://ieeexplore.ieee.org/document/9645311.

Gavade, A. B., Kanwal, N., Gavade, P. A. and Nerli, R. (2023). Enhancing prostate cancer diagnosis with deep learning: A study using mpmri segmentation and classification, *National Conference on Control Instrumentation System Conference*, Springer Nature Singapore, pp. 563–574. Available at: https://arxiv.org/abs/2310.05371.

Hamm, C. A., Baumgärtner, G. L., Biessmann, F., Beetz, N. L., Hartenstein, A., Savic, L. J. et al. (2023). An interactive, explainable deep learning model informs prostate cancer diagnosis at mri, *Radiology* **307**(4): e222276. Available at: https://pubs.rsna.org/doi/10.1148/radiol.2222276.

Islam, R., Imran, A. and Rabbi, M. F. (2024). Prostate cancer detection from mri using efficient feature extraction with transfer learning, *Prostate Cancer* **2024**(1): 1588891. Available at: https://www.hindawi.com/journals/prostatecancer/2024/1588891/.

Li, H., Lee, C. H., Chia, D., Lin, Z., Huang, W. and Tan, C. H. (2022). Machine learning in prostate mri for prostate cancer: current status and future opportunities, *Diagnostics* **12**(2): 289. Available at: https://www.mdpi.com/2075-4418/12/2/289.

Mehta, P. (2022). *Deep learning applications in the prostate cancer diagnostic pathway*, PhD thesis, UCL (University College London). Available at: https://discovery.ucl.ac.uk/id/eprint/10159230/2/PhD$_{t}hesis_{2}$0221030.*pdf*.

Olabanjo, O., Wusu, A., Asokere, M., Afisi, O., Okugbesan, B. et al. (2023). Application of machine learning and deep learning models in prostate cancer diagnosis using medical images: A systematic review, *Analytics* **2**(3): 708–744. Available at: https://www.mdpi.com/2813-2203/2/3/39.

Olawuyi, O. and Viriri, S. (2025). Deep learning techniques for prostate cancer analysis and detection: Survey of the state of the art, *Journal of Imaging* **11**(8): 254. Available at: https://www.mdpi.com/2313-433X/11/8/254.

Passera, R. et al. (2021). Diagnosis according to psa and pca3 score: A machine learning approach, *Urology* **157**: 178–186. Available at: https://www.sciencedirect.com/science/article/pii/S0090429521002860.

Patel, A., Singh, S. K. and Khamparia, A. (2021). Detection of prostate cancer using a deep learning framework, *IOP Conference Series: Materials Science and Engineering*, Vol. 1022, IOP Publishing, p. 012073. Available at: https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012073.

Rao, H. R. and Sankar, V. R. (2024). Precision in prostate cancer diagnosis: A comprehensive study on neural networks, *Journal of Wireless Mobile Networks Ubiquitous Computing Dependable Applications* **15**(2): 109–122. Available at: https://jowua.com/wp-content/uploads/2024/07/2024.I2.008.pdf.

Twilt, J. J., van Leeuwen, K. G., Huisman, H. J., Fütterer, J. J. and de Rooij, M. (2021). Artificial intelligence-based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review, *Diagnostics* **11**(6): 959. Available at: https://www.mdpi.com/2075-4418/11/6/959.

Tyagi, S., Tyagi, N., Choudhury, A., Gupta, G., Zahra, M. M. A. and Rahin, S. A. (2022). [retracted] identification and classification of prostate cancer based on an improved convolution neural network, *BioMed Research International* **2022**(1): 9112587. Available at: https://pubmed.ncbi.nlm.nih.gov/37799171/.

Tătaru, O. S., Vartolomei, M. D., Rassweiler, J. J., Virgil, O., Lucarelli, G., Porpiglia, F. et al. (2021). Artificial intelligence and machine learning in prostate cancer patient management—current trends and future perspectives, *Diagnostics* **11**(2): 354. Available at: https://www.mdpi.com/2075-4418/11/2/354.

Wei, J. T., Barocas, D., Carlsson, S., Coakley, F., Eggener, S., Etzioni, R. et al. (2023). Early detection of prostate cancer: Aua/suo guideline part i: prostate cancer screening, *The Journal of Urology* **210**(1): 46–53. Available at: https://www.auajournals.org/doi/10.1097/JU.0000000000003290.

Yennapusa, H., Ramakrishnan, R., Muniandi, B., Gupta, K. and Logeshwaran, J. (2024). An enhanced optimisation of machine learning model in prostate cancer detection, *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, Vol. 5, IEEE, pp. 184–190. Available at: https://www.researchgate.net/publication/$379673018_{A}n_{E}nhanced_{O}ptimization_{o}f_{M}achine_{L}earning_{M}o$

Yoo, S., Gujrathi, I., Haider, M. A. and Khalvati, F. (2019). Prostate cancer detection using deep convolutional neural networks, *Scientific Reports* **9**(1): 19518. Available at: https://www.nature.com/articles/s41598-019-55972-4.