The Applications of Deep Learning in Medical Imaging Diagnosis

by

Suryadipto Sarkar

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2021 by the
Graduate Supervisory Committee:

Teresa Wu, Chair
Antonia Papandreou-Suppappola
Alvin Silva

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

In this thesis, the applications of deep learning in the analysis, detection and classification of medical imaging datasets were studied, with a focus on datasets having a limited sample size. A combined machine learning-deep learning model was designed to classify one small dataset, prostate cancer provided by Mayo Clinic, Arizona. Deep learning model was implemented to extract imaging features followed by machine learning classifier for prostate cancer diagnosis. The results were compared against models trained on texture-based features, namely gray level co-occurrence matrix (GLCM) and Gabor. Some of the challenges of performing diagnosis on medical imaging datasets with limited sample sizes, have been identified. Lastly, a set of future works have been proposed.

Keywords: Deep learning, radiology, transfer learning, convolutional neural network.

DEDICATION

I dedicate this thesis to my family, to whom I shall be forever grateful.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor, Professor Teresa Wu, for guiding me with the utmost care and understanding. Next, I would like to thank my mentors—Professor Antonia Papandreou-Suppappola, Professor Jennie Si, Professor Baoxin Li and Dr. Alvin Silva—for providing me with unremitting counsel, advice and encouragement. Furthermore, I would like to thank my collaborators from Mayo Clinic Arizona—Dr. Parminder Singh, Dr. Irbaz Riaz, Dr. Haidar Abdul-Muhsin, Dr. Min Kong and Ryan Tatton—for giving me the opportunity to work with them on exciting medical imaging datasets.

My seniors—Josh Xu, Hyunsoo Yoon, Hope Lancaster, Yinlin Fu, Fei Gao, Nathan Gaw and Xiaonan Liu—worked very hard to get me up to speed on the projects and assignments. My peers and fellow scholars—Firas Al-Hindawi, Jay Shah, Teng Li, Elizabeth White, Jiajing Huang, Abhidnya Pathakar and Adam Cochran—have helped grow as a person and a researcher by pushing me to the limits of my capabilities.

Last but not the least, I would like to acknowledge the role that the ASU-Mayo Center for Innovative Imaging (AMCII) has played in my life over the last few years. This is where it all started, and this will always be a home away from home.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Medical imaging refers to the various techniques that are employed in order to create visual depictions of the human anatomy. Medical imaging helps reveal internal structures and organs that are ofttimes obfuscated by layers of skin, tissue, muscle and bones. Radiology is a field of medicine that makes use of medical imaging for the purpose of diagnosis, monitoring, prognosis and treatment of diseases and disorders.

Some of the most pervasive medical imaging techniques are as follows:

1. Medical x-ray imaging and computed tomography (CT)
2. Magnetic resonance imaging (MRI)
3. Ultrasound imaging (sonography)
4. Positron emission tomography (PET)

The aim of this research is to focus on medical imaging—specifically, MRI and its applications in disease diagnosis.

1.1 Background: Imaging Modalities Review

1.1.1 Medical X-ray Imaging and Computed Tomography

X-rays are a type of high-energy electromagnetic radiation which can pass through solid objects as opposed to visible light. On its way through the body, a fraction of the X-ray beams is absorbed by muscles and tissues, thus attenuating the strength of the emergent stream. Such patterns are captured using an X-ray film, thereby producing a visible representation of the internal structures of the body. Some of the most popular X-ray imaging techniques include computed tomography (CT), radiography, dental cone-beam computed tomography, fluoroscopy, mammography and angiography. Figure 1 contains

the thoracic x-rays of two different patients, in the frontal and side views, as obtained from Cohen et al's "COVID-19 Image Data Collection" lung X-ray dataset [1].

[2] reviewed the current state-of-the-art digital image acquisition and processing frameworks used in medical x-ray imaging. It also discussed in detail the future concepts of digital x-ray image acquisition techniques that shall be used in medicine. Digital x-ray fluoroscopy, and solid-state large area flat dynamic x-ray image detector (FDXD) are two state-of-the-art x-ray imaging techniques that have been described in the paper. Furthermore, the authors discuss digital quantum noise reduction techniques—including but not limited to noise reduction by spectral amplitude estimation—for generating cleaner x-ray images.

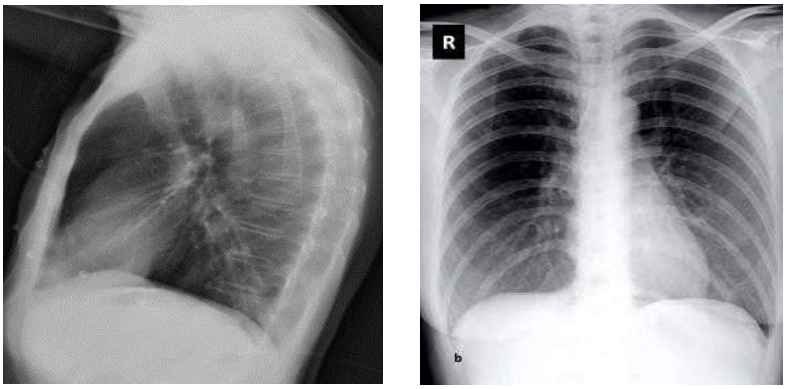


Fig 1. Thoracic X-rays of two different patients, in the frontal and side views, as obtained from Cohen et al's "COVID-19 Image Data Collection" lung X-ray dataset [1]

Computed tomography (CT) scanners comprise a circular, rotating frame. The frames consist of an X-ray tube and a detector on opposite ends. The rotating frame enhances the generation of thousands of sectional views, which help in the reconstruction of the anatomy being imaged. CTs are helpful in generating 3D images, and the quality of the images generated are far better than traditional X-rays [3]. Figure 2 contains the CT scans of four different patients, obtained from Soares et al's "SARS-Cov-2 CT-scan dataset [4]."

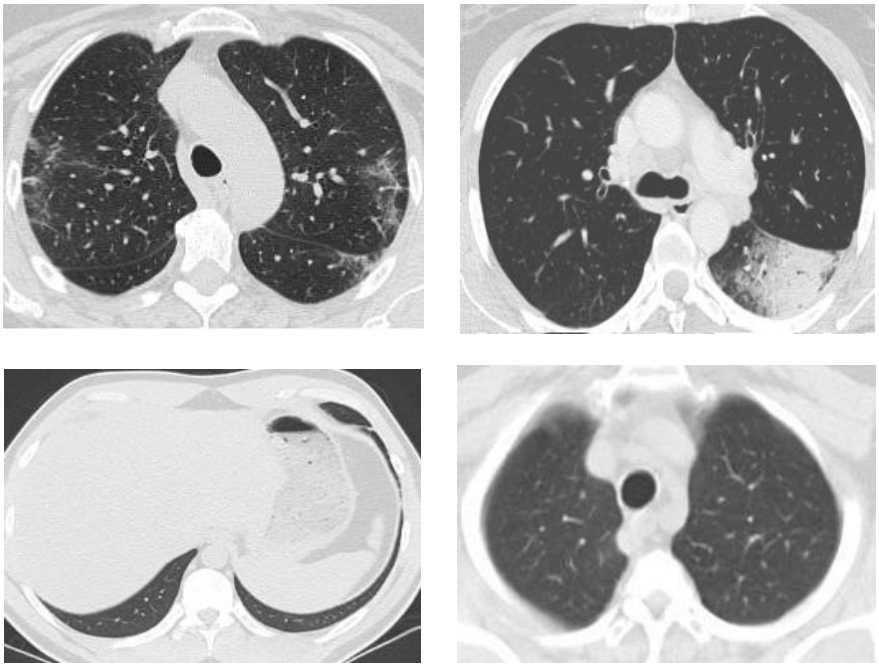


Fig 2. CT scans of four different patients, obtained from Soares et al's "SARS-Cov- 2 CT-scan dataset [4]." The first two patients have been diagnosed with COVID- 19, while the others pertain to non-COVID cases.

CT is a relatively new technology that was developed to improve image acquisition and detection of x-rays [5, 6, 7, 8]. CTs are also especially effective in image reconstruction, as reported by [5]. However, one longstanding concern with CT scans is its increased exposure to radiation [9, 10]. [11] describes some of the basic concepts and principles

behind image acquisition and processing using computed tomography. Although we are concerned primarily with the applicability of CT scans in medical imaging, it must be noted that computed tomography has a plethora of industrial applications [12].

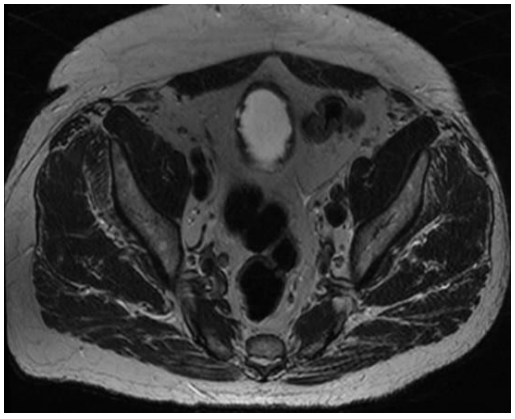
1.1.2 Magnetic Resonance Imaging

Medical resonance imaging (MRI) is a medical imaging technique which employs a magnetic field in combination with computer-generated radio waves in order to generate detailed visual depictions of the internal structures and organs of the human body. Unlike computed tomography (CT) and positron emission tomography (PET), MRI doesn't involve the use of X-rays or any sort of ionizing radiations whatsoever. Figure 3 contains four different MRI contrast types—namely, Apparent Diffusion Coefficient (ADC), T2-weighted Fast Recovery Fast Spin Echo (FRFSE), Pelvis (T2 FatSat) and Gadolinium contrast (GAD)—utilized to capture the prostate scan of the same patient (#2663).

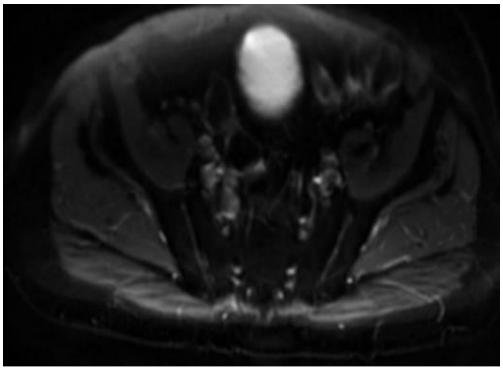
MRI is a pervasively used imaging technique in medicine [13, 14, 15]; it has a wide range of applications and variants including but not limited to: cardiac magnetic resonance imaging (also called cardiovascular magnetic resonance imaging, CMR) [16], functional MRI (fMRI) [17], molecular MRI [18], parallel MRI [19]. Although prevalently used in diagnostic medicine, MRIs have their own unique challenges [18] and side effects (also known as contraindications) [19].

Apparent Diffusion Coefficient:

T2-weighted FRFSE:



Pelvis (T2 FatSat):



Gadolinium contrast (GAD):

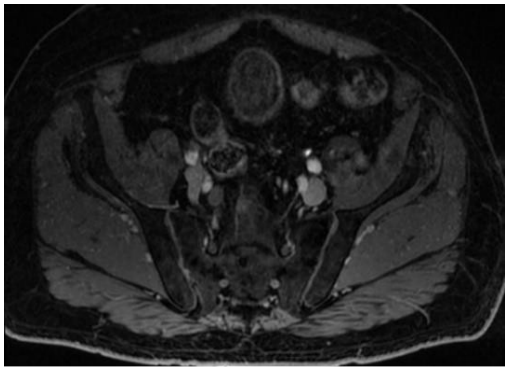


Fig 3. Four different MRI contrast types—namely, Apparent Diffusion Coefficient (ADC), T2-weighted Fast Recovery Fast Spin Echo (FRFSE), Pelvis (T2 FatSat) and Gadolinium contrast (GAD)—utilized to capture the prostate scan of the same patient (#2663). The images have been obtained from a dataset comprising prostate scans of 16 patients, all diagnosed with prostate cancer—which has been provided to us by Mayo Clinic, Arizona.

1.1.3 Ultrasound imaging (Sonography)

Ultrasound imaging (also known as sonography) employs high-frequency sound waves to map the internal structures of the body. Ultrasonograms (USGs) are generated in real-time, therefore depicting motion in the internal tissues and organs as well. They are also good at depicting the blood flow through the organs. Unlike X-rays, this is a relatively safe diagnostic procedure, owing to the fact that the patient does not have to go through any amount of exposure to ionizing radiation. Figure 4 contains four ultrasound scans obtained from the "Point-of-care Ultrasound (POCUS)" recordings, which has been utilized by the first recorded work that automatically detected COVID- 19 from ultrasound alone [20].

Ultrasound imaging depends heavily on the concepts of signal processing [21, 22]; and, as with any of the aforementioned diagnostic imaging techniques, it has evolved considerably over the years [21, 23]. Some of the most important components of ultrasound, as reported by [22], include but are not restricted to—pulse compression, beam forming, transducers, contrast agents, tissue harmonic imaging, tissue motion and blood flow detector, and 3D imaging—with continuous innovations being made on each of these techniques by researchers and medical physicists worldwide.

For a long time, traditional 2D ultrasound techniques struggled with analyzing 3D anatomy; therefore radiologists were tasked with integrating 2D image plates manually. The development of three-dimensional ultrasound was a breakthrough that solved this problem, thus enabling easier visualization and diagnosis [24].

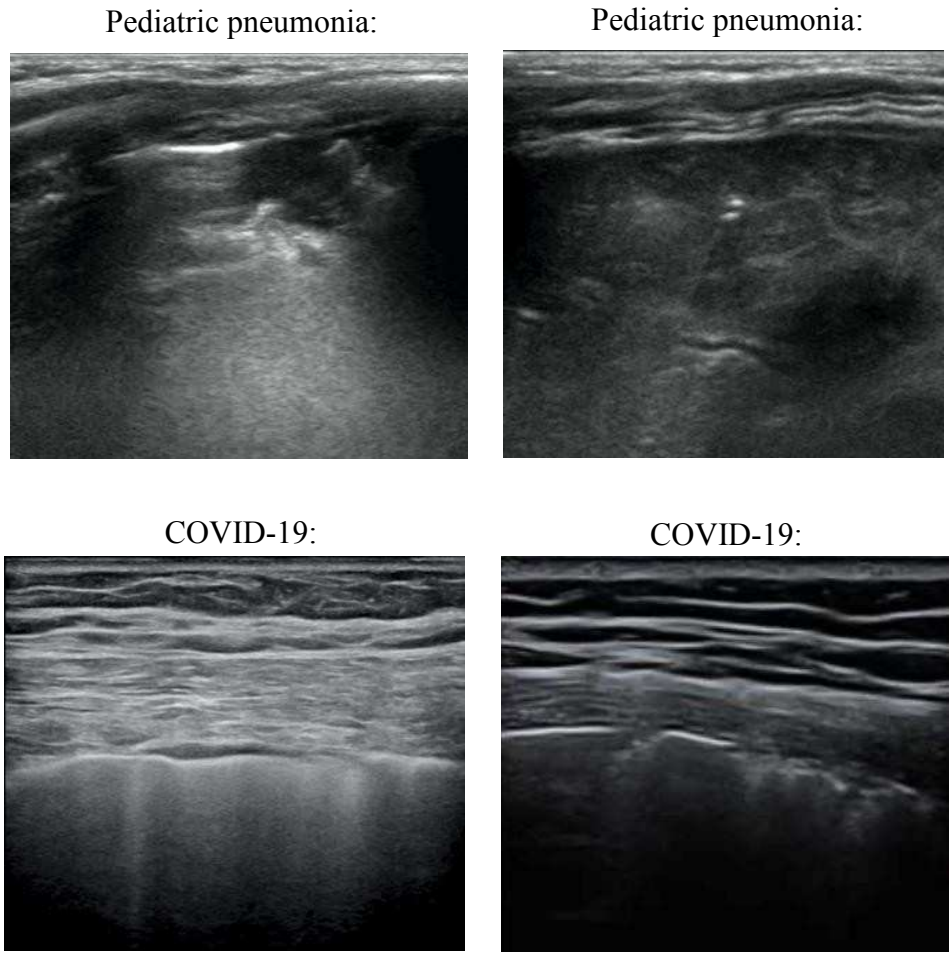


Fig 4. Four ultrasound scans obtained from the "Point-of-care Ultrasound (POCUS)" recordings, which has been utilized by the first recorded work that automatically detected COVID-19 from ultrasound alone [20].

1.1.4 Positron Emission Tomography

Positron emission tomography (PET) is a functional imaging technique. It utilizes radioactive substances—known as radiotracers—to measure and monitor the functioning of tissues and organs. The radiotracer help visualize metabolic processes, physiological activities and progressions of diseases. Based on what part of the body is under review, the

radiotracers could be injected, swallowed or inhaled. Owing to the tendency of radiotracers to accumulate in regions of chemical imbalance, it often manifests as bright spots in diseased areas of the body. PET scans are ofttimes utilized in combination with CTs, MRIs and X-rays for the purpose of diagnosis. Figure 5 contains regions of interest extracted from PET scans of prostate lymph nodes pertaining to four different patients suffering from prostate cancer.

PET is an effective diagnostic imaging technique, and is used in an eclectic range of medical applications [25, 26]. Its role in modern oncological applications, including the detection and staging of a wide variety of tumors, is unparalleled [27, 28, 29, 30, 31]. Some of the neurological applications of PET imaging include but are not restricted to the visualization and diagnosis of movement disorders [32, 33, 34], dementia [35, 36, 37], brain tumors [38, 39, 40], epilepsy [41, 42, 43], neuroplasticity [44, 45], stroke [46, 47, 48], and the study of drug delivery [49, 50, 51] and pharmacodynamics [52, 53, 54].

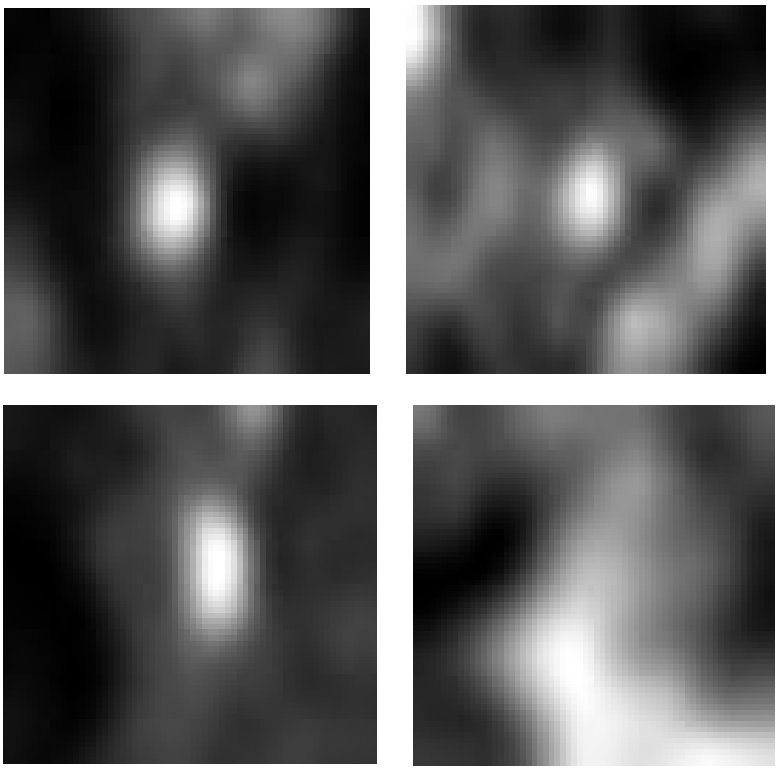


Fig 5. Regions of interest extracted from PET scans of prostate lymph nodes pertaining to four different patients suffering from prostate cancer. The first two images correspond to malignant lymph nodes, whereas the others to healthy ones. The PET images had been obtained from a dataset comprising lymph node scans of 16 patients diagnosed with prostate cancer, provided by Mayo Clinic, Arizona.

1.2 Research Objective

The objective of this research is to design an artificial intelligence (AI) -based framework that will be capable of performing automatic classification on medical imaging datasets

containing a limited number of samples. Manual feature extraction often leads to a lack of representational capabilities of the model, and they often fail to capture the most important distinguishing characteristics of the images which are critical in performing classification. Deep learning, on the other hand, has an innate capability of being able to mine out the most important features from images. However, deep learning has its own drawbacks:

1. Complex deep models are difficult to debug, and suffer from non-interpretability (often referred to as the "black-box problem")
2. Deep models require large volume of data for training, and can easily overfit on small datasets
3. Deep models often require extensive computing time and resources owing to their complex structure

Efficient applicability of neural network-based classifiers on small datasets is an open problem in AI, and the key is to find a balance between manually extracted features and automated deep models. We delve into the challenges faced by texture-based features as well as deep models on medical imaging data, subsequently proposing a model that makes the best of both worlds. We employ a convolutional neural network-based model to extract features; subsequently, those features are used to perform classification on traditional machine learning classifiers. As a result, our model generates effective image representations while at the same time not overfitting on the small data.

CHAPTER 2

LITERATURE REVIEW

2.1 Deep Learning

In this section, we first review four deep models that are commonly used in performing diagnosis on medical images: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Generative Adversarial Network (GAN), and Deep Reinforcement Learning (DRL).

2.1.1 Convolutional Neural Network

Convolutional neural network (CNN) is a category of deep neural networks (DNN), that have been designed specifically for image classification and recognition applications. The design of CNNs is inspired by the visual cortex of the human brain. They are useful for video and audio applications as well. CNNs are a way to make machines capable of perceiving the world in a way that a human being would. It was a great step in the direction of automated computer vision and is becoming an ever-evolving research field. Figure 6 exhibits the layers in a CNN.

Deep neural networks (DNN)—as discussed earlier on in this paper—have made image processing tasks easier for engineers and researchers, owing to the fact that they do not require the model designer to manually extract features. But, in order to optimize performance, appropriate architectures must be chosen—depending on the task at hand, as well as the dataset. [55] attempts at formulating connections between the architecture and depth of a CNN, and its performance— given datasets of varying size and complexity.

Detection, classification and segmentation are important tasks in medical image processing

applications. [56] and [57] implement CNNs and their variants in automatic lesion detection, and multiple abnormality detection, respectively—from medical images. [58] implements a deconvolutional CNN for classification of acute lymphoblastic leukemia (ALL). [59] designs a multi-network feature extraction model using pre-trained deep CNNs, which aids in breast cancer classification. [60] gives a concise introduction to multiscale CNNs, and their applicability in the classification of cells from medical images. [61] makes use of multiscale all convolutional neural network (MA-CNN) for classification of mammograms. [62] designs DCNN ensembles for the purpose of segmentation in infant brain MRI images. [63] segments anomalies in abdominal CT images by CNN, and then classifies them using fuzzy SVM.

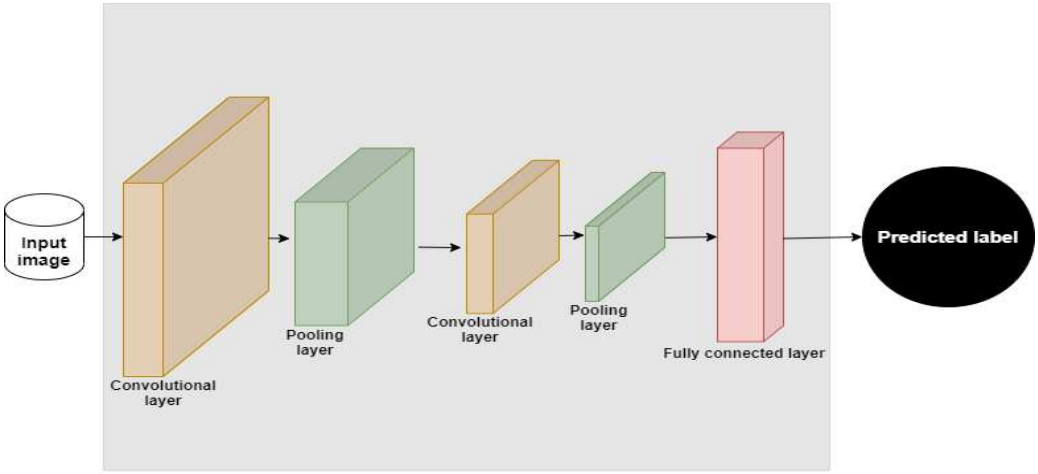


Fig 6. Layers in a convolutional neural network (CNN)

2.1.2 Recurrent Neural Network

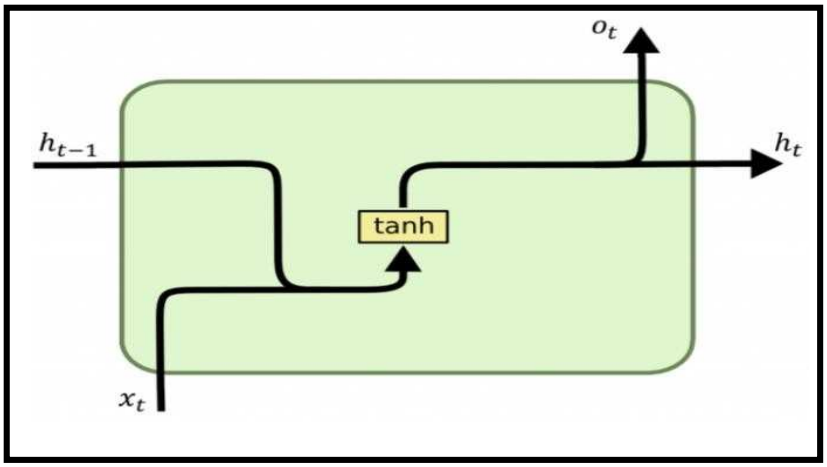
Recurrent Neural Network (RNN) is a category of artificial neural networks that are designed to handle sequential data using directed, cyclic connections between layers of nodes. Unlike feedforward neural networks, they are capable of processing variable-length data. RNNs make use of a notion of state or memory, to process subsequent sequences of

incoming data. The output generated at a time step is used as input for the next time step. Long short-term memory (LSTM) networks and gated recurrent units (GRU) are two of the most commonly used types of RNNs. Figure 7 depicts the general structure of an RNN.

[64] designs a combined CNN-RNN model for the purpose of blood cell classification from a dataset comprising a total of 12,444 blood cell images. This model performs better than some of the popular CNN models such as ResNet and Inception V3. The authors of [65] design a novel, recursive generative adversarial network (GAN)—named RNN-GAN—for performing efficient image segmentation on medical image datasets where the region of interest is very small in size compared to the background. The proposed combined model, when trained on medical images of different types and sizes, yield consistently good performance. Adequate image acquisition is essential when it comes to diagnosis using medical images. [66] makes use of RNN to design a model, in real time, to predict ultrasound scan adequacy—for diagnosing pediatric, developmental dysplasia of the hip (DDH). [67] makes use of a combination of multilayer perceptron and RNN for the purpose of classification in MRIs, based on the longitudinal characteristics of the data. The authors report an accuracy of 89.7% using the proposed model for AD classification. In [68], the authors—for the very first time—design a way to automatically segment the frame-wise glottal areas, along with the vocal fold tissues directly from the video. A CNN model, which makes use of LSTM cells, is implemented. This model is picked from a total of eighteen different CNN configurations which had been trained and evaluated on 13,000 HSV frames acquired from 56 healthy and 74 unhealthy patients. The authors in [69] propose a model—called the 2D- ResNet18+LSTM—for brain age prediction from 3D pediatric brain MRIs. The proposed model comprises four parts: 2D ResNet18 for feature

extraction on 2D images; a pooling layer for feature reduction over sequences; an LSTM layer; and a regression layer—in that order. The authors of [70] propose a technique that combines optimal wavelet statistical texture features and RNN, for tumor detection in brain MRIs. The system comprises four phases: feature extraction, feature selection, classification, and segmentation. After the feature selection process, selected features are fed into RNN for brain tumor classification. After classification, the images with tumors are selected for segmentation—which, in turn, was done using a customized version of the region growing algorithm. The authors of [71] propose a fully automated technique for denoising task- based fMRI data, without assuming any particular noise model. The designed system uses a DNN made up of: one temporal convolutional layer, one LSTM layer, one time-distributed fully- connected layer, and one unconventional selection layer—in that order.

Fig 7. The structure of a recurrent neural network (RNN); where h_t is the hidden layer vector, x_t is the input vector and o_t is the output vector [2].



2.1.3 Generative Adversarial Network

Generative adversarial networks (GAN) are a class of deep learning models that comprise two competing neural networks. One of those networks is known as the generator network, while the other the discriminator network. The task of the generator is to generate fake data, while the task of the discriminator network is to try to discriminate between the fake data and the real one. Figure 8 depicts the general structure of a GAN. [72] devises an unsupervised pre-training method on a GAN, before performing classification on a deep neural network for brain tumor classification. [73] does pulmonary nodule classification on lung CT images (which in turn aids in the early detection of lung cancer), using deep convolutional neural network (DCNN) and GAN. [74] employs U-Net and GAN for unbalanced medical image segmentation. [75] performs medical image segmentation using adversarial image-to-image networks. In [76], task-driven GANs aid in the unsupervised segmentation of x-ray images. Conditional GANs are deployed in [77] to denoise low dose chest images, with encouraging results.

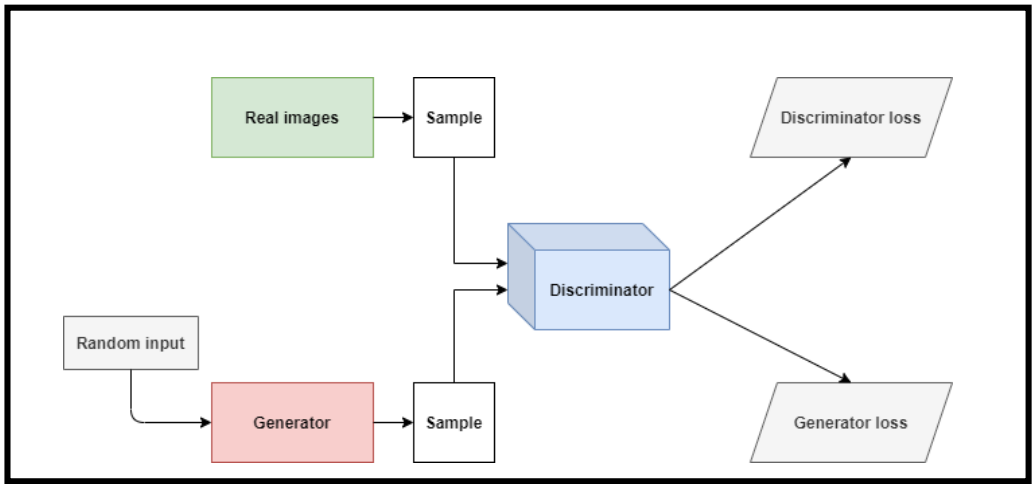


Fig 8. General structure of a generative adversarial network (GAN)

2.1.4 Deep reinforcement learning

Reinforcement learning (RL) is a type of machine learning where an agent tries to maximize some predefined measure of "reward" by interacting with the environment. Deep reinforcement learning (DRL) entails a combination of deep learning and reinforcement learning techniques, meant to enable agents perform a wide variety of AI-related tasks. Figure 9 depicts the general structure of a reinforcement learning model.

The authors of [78] design an automatic data augmentation module using DRL. This augmentation module is cascaded with an image segmentation module such that the resulting combination can act as one complete, end-to-end training model with a consistent loss. [79] proposes a DRL model for the classification of pulmonary nodules from thoracic CT images. On feeding as input to the system a raw thoracic CT image, it predicts whether or not a nodule is present in that image. The training accuracy is reported as 99.1%; while the test accuracy is 64.6% (sensitivity 58.9%, specificity 55.3%, PPV 54.2%, NPV 60.0%). [80] makes use of RL for the purpose of detecting breast lesions from dynamic contrast-enhanced magnetic resonance volume (DCE-MRI). The research aims at maintaining state-of-the-art accuracy while also speeding up the time required for lesion detection. This study makes use of the deep Q-network approach (DQN) for lesion detection. [81] presents a comparative study of popular, representative deep reinforcement learning models in lung tumor detection. Deep Q-network (DQN) and its variants—such as the deep dueling Q-network, deep recurrent Q-network, and hierarchical deep Q-network (h-DQN) have been explained.

The authors of [82] evaluate novel deep reinforcement learning strategies which are aimed at training agents to localize landmarks in medical images. Fixed- and multi-scale search

strategies have been evaluated, with novel action steps in a coarse-to-fine manner. DQN-, double DQN (DDQN)-, duel DQN-, and duel DDQN-based architectures have been compared. [83] makes use of multi-scale image analysis and deep reinforcement learning for real-time 3D landmark detection in CT images. The performance and runtime of the proposed method has been compared with five reference methods, and the results seem promising.

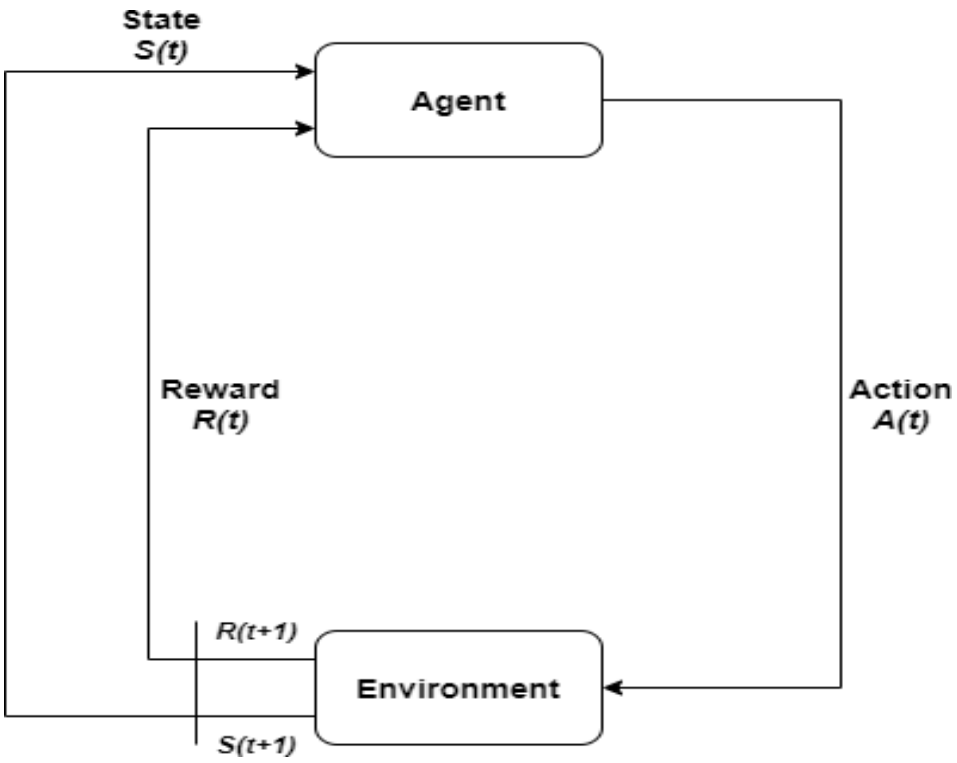


Fig 9. The general structure of a reinforcement learning model

2.2 Transfer Learning

While the success of deep models attracts great attention in medical research, it is not without limitation. As indicated in [84], clinical studies often have limited samples which posed great challenges to CNN model. One solution is transfer learning which is a

technique prevalently used to train deep networks on small datasets. Transfer learning (TL) refers to the migration of knowledge between applications. Owing to restrictions on sample quality, data availability, lack of domain knowledge et cetera, it is ofttimes challenging to develop robust models based only on the resources available for the purpose of the application at hand. In such scenarios, researchers would train models that had previously acquired some knowledge from similar tasks, or data sets and transferred the pre- trained model to the dataset of interest. For example, [85] selects color optic disc-centered fundus images using active learning; subsequently identifying glaucoma using transfer learning on a deep CNN. In [86], a problem-based architecture of DCNN called ChestNet is proposed. This variant of DCNN, ChestNet, is pre-trained on a set of relevant and irrelevant data sets; before finally being trained on the Pediatric Chest X-ray dataset for detection of pulmonary consolidation. [87] gives a concise yet informative description about ChestNet, and its applicability in the detection of thoracic diseases on chest images. The authors of [88] develop an ensemble of five of the most commonly used deep CNN models (AlexNet, DenseNet121, InceptionV3, ResNet18 and GoogLeNet) pre-trained on ImageNet, for the purpose of pneumonia detection in the Guangzhou Women and Children's Medical Center dataset of chest X-rays.

In many a case, it is extremely difficult for humans or DL methodologies alone to extract the most important set of features from medical image data sets. Hence researchers oftentimes need to go with a combination of machine learning and deep learning approaches in order to utilize the representational capabilities of deep models, while at the same time not overfitting the data. [89] has followed one such approach. The researchers have access to 58 in-house brain MR images, and 128 MR images from The Cancer

Genome Atlas—all of patients with high-grade glioma. For each patient, the researchers calculate 348 hand-crafted radiomics features; and extract 8192 using a pre-trained deep CNN. Next, they perform feature selection and Elastic Net-Cox modeling to classify patients into long- and short- term survivors. [90] is a detailed study of ROI-based opacity classification of diffuse lung diseases in chest CT images. It makes use of the Cifar-10 and Cifar-100 data sets for pre-training deep CNNs (VGG16 used in this study); subsequently, a CT image data set of diffuse lung diseases for parameter tuning, and classification. It delves into the structure of CNN used, and how to implement pre-training and parameter tuning. This paper gives the reader an intuition about the relation between the type and characteristics of the data sets used for pre-training and parameter-tuning, and the effectiveness of the TL model. From the experimental results presented here, it is apparent that pre-training data sets with a higher number of class divisions; and fine-tuning data sets with a compact structure, or compact FC layer—yield best results. One of the many challenges of working with medical image data sets is the fact that they often contain too few samples to effectively train a deep neural network. Apart from under-training due to a lack of samples, deep models trained on small data sets also often suffer from: accuracy issues, unnecessarily complex structure of the deep network, and high computational time. In [91], the researchers implement transfer learning on CNN; and extreme learning machine to classify between malignant and benign pulmonary nodules on CT images. The deep CNN, which has been pre- trained with the ImageNet data set, is used to extract high-level features of pulmonary nodules. These features help in the classification of benign and malignant pulmonary nodes, using the extreme learning machine (ELM) model that has been developed by the researchers. The classification performance of this combined, deep

transfer CNN and extreme learning machine model is then evaluated using two data sets: the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) public data set; and a private data set from the First Affiliated Hospital of Guangzhou Medical University (FAH-GMU) in China.

The authors of [92] make use of cosine loss after the softmax activation layer, instead of the pervasively used cross-entropy loss, to achieve better training of CNN models on small datasets. [93] utilizes transfer learning and deep CNNs to classify COVID-19, pneumonia and normal patients from a small chest X-ray dataset. The authors of [94] make use of pre-trained VGG19, MobileNet v2, Inception, Xception and Inception ResNet v2 models to train CNNs on small datasets of lung images, which would ultimately achieve a classification accuracy, sensitivity and specificity of 96.78%, 98.66% and 96.46% respectively. The authors of [95] design a deep model called AppendiXNet for the purpose of detecting appendicitis from a dataset of less than 500 CT images. The model was pre-trained on 500,000 video clips, where each video was annotated for one of 600 human actions. The pre-trained model weights are then fine-tuned using 438 CT scans of the appendix. The study has exhibited how pre-training significantly improves results, over training the model directly on the target data. Figure 10 depicts the general structure of traditional machine learning and transfer learning tasks.

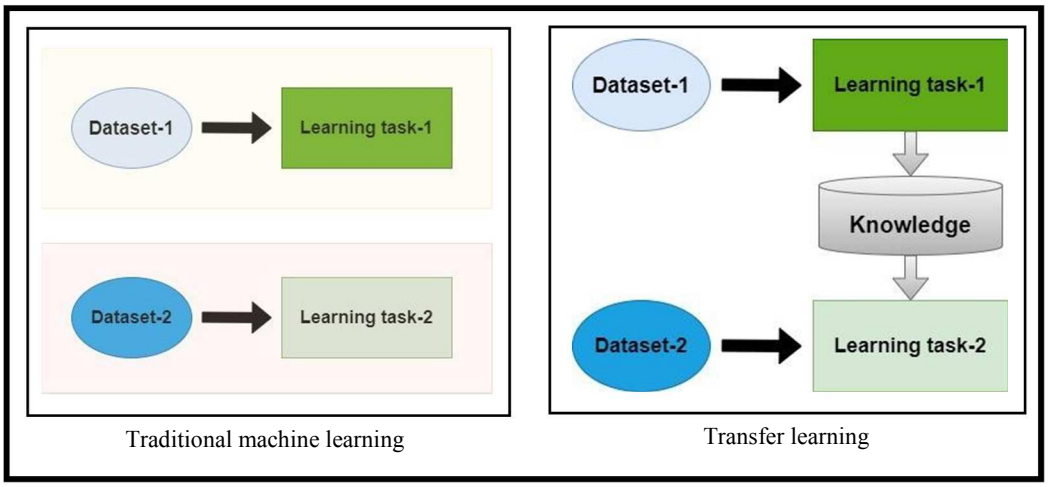


Fig 10. The general structure of traditional machine learning and transfer learning tasks

2.3 Summary of Literature Review

Convolutional neural networks (CNNs) are particularly effective in performing classification on medical imaging datasets. However, like all deep neural network-based architectures, they require lots of samples for training owing to their complex structure. In this study, we were working with a small medical imaging dataset of 126 samples, where the task was to perform classification of lymph node metastasis on prostate cancer patients. Therefore, in order to utilize the representational capabilities of deep models, while at the same time alleviating overfitting, we made use of a combined deep learning-machine learning approach. The ResNet18 model (which is a 71-layer deep CNN-based network) was utilized for the purpose of feature extraction from the gray-scale images. The decision tree-based machine learning classifier was subsequently used for performing classification. Also, in order to remove unimportant features, we used a combination of unsupervised (statistical) and supervised feature selection techniques. The

proposed model outperformed two texture feature-based approaches, namely GLCM and Gabor. And it was comparable in performance to some of the state-of-the-art neural network-based classification models. The proposed deep learning-machine learning framework had a classification accuracy of 76.19%, sensitivity of 79.76%, specificity of 69.05%, precision of 83.75% and F1-score of 81.71%.

CHAPTER 3

DETECTION OF LYMPH NODE METASTASIS IN PROSTATE CANCER PATIENTS: A CASE STUDY

3.1 Introduction

In 2017 in the United States, prostate cancer was the second most common new cancer diagnosis and the third most common cause of cancer-related death [96]. Often prostate cancer is slow growing and initially confined to the prostate gland [97]. In these instances, these patients may need no treatment, opting instead for active surveillance. Other patients need surgery, chemotherapy, immunotherapy, radiation therapy, or often a combination of these. The decision to intervene and the best intervention hinges on the cancer's stage. Staging cancer depends on primary tumor growth, for example, growth into adjacent organs such as the seminal vesicles or urinary bladder. Staging also depends on secondary, metastatic extent. Prostate cancer has a predilection for spreading to pelvic/retroperitoneal lymph nodes and bones.

As overall prostate cancer tumor burden often determines the treatments offered, reliable staging is important. In this respect, imaging can play a key role. Magnetic Resonance Imaging (MRI) has emerged as an important imaging modality for assessment of tumor invasion and pelvic lymph node metastases [98]. However, determination of lymph node metastatic status can be challenging because abnormal (cancerous) and normal lymph nodes often appear similar. As such, the sensitivity of imaging for lymph node metastasis in prostate cancer is low [99]. In this section, we will discuss some existing literature pertinent with feature extraction on medical imaging datasets—starting from the conventional approach of manual texture-based features, to the more recent developments on automated feature embedding generations using convolutional neural network (CNN)-

based deep models.

Texture analysis refers to the segregation of the different regions in an image, based on their physical characteristics or intensity distribution. Traditionally, researchers focused on texture analysis to extract features from images, which were subsequently utilized for the purpose of classification by machine learning algorithms. There are many different types of texture analyses, some of the most common being gray-level co-occurrence matrix (GLCM), local binary patterns (LBP) and Gabor filters. For extensive and complete analyses, researchers often utilized different types of texture algorithms in combination with features that are representative of the imaging modality in question [100]. [101] designed a multifractal feature descriptor to classifying non-neoplastic tissues and tumors, as well as grade hepatocellular carcinoma tissues into five stages. For both of those tasks, the proposed feature descriptor outperformed Gabor filter, Leung-Malik filter, local binary patterns (LBP) and Haralick. [102] utilized three texture-based features—Haralick, Gabor and LBP—to classify between benign and malignant pulmonary nodules using the Support Vector Machine (SVM) classifier. The results showed a similar performance accuracy of 90% for all three of these features, but a highest area under the curve (AUC) of 92.70% for Haralick. [103] made use of a dataset of 22 patients with glioblastoma to perform classification between true progression and pseudoprogression from T2-weighted MRI images. Five GLCM features, namely homogeneity, entropy, energy, correlation and contrast were extracted, out of which correlation generated the best classification performance with an accuracy of 86.40%, AUC of 89.20%, sensitivity of 75% and specificity of 100%. [104], [105], [106], [107] and [108] all utilize texture-based features to perform breast cancer classification. [104] makes use of LBP, statistical features,

discrete wavelet transform (DWT) and contourlet transform (CT) to classify breast abnormalities on the publicly available DDSM database of mammograms, using the SVM classifier. Statistical-based and LBP features return an accuracy of 98.43% while DWT returns an accuracy of 96.93%, all of which outperform CT-based features. Moreover, the results indicate that a combination of LBP and statistical features performs the best, with an accuracy of 98.63%. [105] designed a computer-aided detection (CAD) system for breast cancer diagnosis on 56 patients from multi-parametric ultrasounds. A combination of LBP and gray-scale intensity features were extracted, resulting in a highest classification accuracy of 90% on the 8th week, highest sensitivity of 95% on the 4th week, and highest specificity of 91% on the 8th week. [106] proposed a feature descriptor for efficient breast cancer classification on ultrasound images using a combination of phase congruency-based and LBP features, from a small database containing 138 cases. The combined feature descriptor is reported to perform better than either of the individual approaches, with a highest AUC of 89.40%. [107] extracts a huge feature set comprising morphological, Gabor and other textural features to incorporate imaging modality-based knowledge into the model. Subsequently, feature selection is performed to identify the most relevant features, followed by classification. The proposed model was used in the detection of breast cancer from a dataset containing 438 lesion and 1898 control tissue scans, in which regard an AUC of 96.17% was achieved. [108] performs classification between normal and cancerous tissues from breast thermograms, using Gabor wavelet features to measure the amount of asymmetry between right and left breast tissues for patients with and without cancer. A Gabor filter bank was used to extract features from 20 frontal breast thermograms, out of which the two most important features were energy and amplitude.

The results reported significant difference in the variance of energy and amplitude values between right and left breast tissues, for patients with and without cancer—which led the authors to hypothesize that Gabor wavelet-based feature descriptors are efficient in classifying between normal and malignant breast thermograms.

Conventional machine learning classifiers depend heavily on manually extracted features. CNNs, on the other hand, have an innate capability of automatically extracting consequential feature vectors from images, even though the extracted features might not be the most appealing or meaningful to the human eyes. Such drawbacks of texture-based features, paired with the computational capabilities associated with most modern computers, as well as the availability of enormous quantities of data, deep learning is in pervasive use in medical imaging applications. The most common type of deep neural network is the convolutional neural network (CNN). CNNs are modelled on the visual cortex of the human brain. They are extensively used for imaging, video and audio applications. The authors of [109] have conducted a detailed survey pertaining to the efficacy of CNNs in object detection applications. CNNs have been studied for detection, classification and segmentation tasks in medical research. For example, [110] proposed a CNN-based double-branched model wherein one branch was utilized for feature extraction, the other for segmentation for multiple abnormality detection from medical images. [111] proposed the CemrgApp, a CNN model, to classify cardiovascular properties from cardiovascular magnetic imaging (CMRI) scans of different cardiac patients, for efficient diagnosis and treatment. [112] and [113] implemented CNNs and their variants in automatic lesion detection and multiple abnormality detection from medical images.

In the context of this paper, it is critical not only to review the applicability of deep neural

networks in the field of radiology in general, but also to delve into how pervasive it is in handling lymph node metastases. [114] utilized three different CNN architectures to detect axillary lymph node metastasis from primary breast cancer patients. The three models used were Inception V3, Inception-ResNet V2, and ResNet-101. The implementation yielded better prediction results than those performed manually by the radiologists. Inception V3, the best performing model, reported an area under curve (AUC) of 0.89, 85% sensitivity and 73% specificity. The radiologists achieved 73% sensitivity and 63% specificity. [115] made use of eight different pre-trained CNN models to diagnose cervical lymph node metastasis from CT scans of patients with thyroid cancer. The image dataset, containing 995 axial CT scans, had roughly twice as many benign lymph nodes as malignant ones. ResNet50 is the best performing model with an AUC of 0.953 and an accuracy of 90.4%. Focusing the discussion further to the use of deep models in analyzing lymph node metastasis, we see how [116] utilizes ten CNN-based architectures on a small dataset of 218 patients to perform classification. They reported a mean AUC of 0.68, accuracy 61.37%, sensitivity 53.09% and specificity 69.65%.

The effective detection of lymph node metastasis is highly dependent on imaging modality. For example, the authors of [117] designed a radiomics signature by manually extracting a total of nine features from the lymph node CT scans of 118 patients. A radiomic nomogram was subsequently generated using the logistic regression model, which achieved an AUC of 0.8986. [118] utilizes a total of 103 T2-weighted MRIs to perform classification of lymph node metastasis on bladder cancer patients. 718 features were manually extracted from the bladder scans, and a nomogram was generated using logistic regression. Feature selection was performed, which reduced the number of important features from 718 to 9.

Classification was subsequently performed on these 9 important features, generating a validation AUC of 0.8447.

As reviewed above, existing research shows some success of using deep learning model (e.g., pre-trained model taking advantage of transfer learning) to extract features used in machine learning models. Motivated by this, we propose an integrated deep learning-machine learning pipeline to utilize the representational capabilities of CNNs, yet at the same time not overfit the model on small prostate imaging dataset. We utilized the CNN-based ResNet18 model which was pre-trained on the ImageNet dataset, to extract features from our own prostate image dataset for classifying between normal and metastatic lymph nodes. These features subsequently underwent a detailed feature selection framework comprising supervised and unsupervised feature selection techniques. The most important features as obtained from the feature selection step were finally used for classification using the decision tree classifier. The ResNet18-based features yield a highest classification accuracy of 76.19%, sensitivity of 79.76%, specificity of 69.05%, precision of 83.75% and F1-score of 81.71% on the 10-fold decision tree (DT) classifier. Our ResNet-based feature extraction framework was compared against two texture feature-based classification models namely, GLCM and Gabor. The decision tree classifier trained on GLCM features achieved a best classification accuracy of 61.90%, precision of 71.43% and F1-score of 72.73%. The same decision tree when trained on Gabor features achieved a best classification accuracy of 65.08%, precision of 76.25% and F1-score of 74.84%. These set of experiments have shown that the proposed, combined deep learning-machine learning architecture is not only comparable in performance to other state-of-the-art deep learning-based classifiers; but also outperforms some of the most pervasively utilized

texture feature-based approaches, especially when trained on medical imaging data containing limited samples.

3.2 Methodology

We have designed a deep learning-machine learning combined model wherein the deep learning module is employed for the purpose of feature extraction from the raw images; and the features are subsequently classified using a machine learning classifier. In this section, we present the various components, as well as the classification results of the deep learning-machine learning combined framework. Figure 11 provides a schematic representation of the overall workflow, starting from the raw images of lymph nodes, to their automatic categorization as malignant/ non-malignant. The methodology includes feature extraction, feature selection and classification. In the proposed model, we first extract a 512-element feature vectors per image from the average pooling layer of the ResNet18 pre-trained model. (For the purpose of comparison, we also extract features using two texture analysis-based approaches, namely GLCM and Gabor.) Then we perform our feature selection algorithm on this feature vector. That leaves us with the most important set of features, which are subsequently classified utilizing a machine learning classifier. By employing such a combination of deep learning (for feature extraction) and machine learning (for classification) models, we can ensure that we have a model with a nuanced representation of the images, however at the same time assuaging the problem of overfitting on the small dataset.

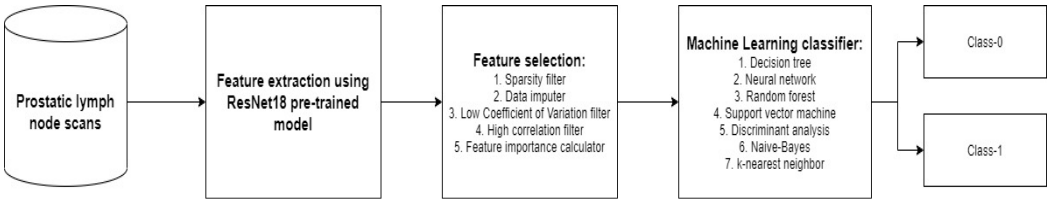


Fig 11. A schematic representation of the overall workflow—starting from the raw images of lymph nodes, to their automatic categorization as malignant/ non-malignant. The methodology includes feature extraction, feature selection and classification.

3.2.1 Feature Extraction

3.2.1.1 Texture-Based Feature Extraction

This section delves into discussions about the texture-based features that were extracted, namely GLCM and Gabor.

3.2.1.1.1 Gray Level Co-Occurrence Matrix (GLCM)

11 GLCM features, namely contrast, correlation, energy, entropy, homogeneity, variance, sum of average, sum of variance, sum of entropy, difference of variance and difference of entropy were extracted from each of the ROI scans. The scalar distance was selected as $\mathbf{s = 2}$; and the orientations were selected as $\boldsymbol{\theta = 0^o, 45^o, 90^o}$ and $\mathbf{135^o}$ corresponding to 4 different GLCM matrices. Subsequently, the corresponding features from each orientation were averaged to generate the final 11 GLCM features.

3.2.1.1.2 Gabor

Further, to illustrate the efficacy of the proposed approach over simple texture analysis, Gabor and GLCM features were extracted from the ROIs. The Gabor feature extraction procedure was inspired by [119]—where a filter bank of 40 filters (5 different scales, with 8 orientations per scale) was utilized to extract features. A uniform input image size

of $\mathbf{28 \times 28}$ was used, resulting in a total of $\mathbf{28 \times 28 \times 40 = 31360}$ features extracted per ROI scan. Subsequently—in order to remove redundancy of features extracted [119, 120, 121]—we decided to downsample the features by a scale of 4, thereby reducing the feature vector size to $\mathbf{31360/(4 \times 4) = 1960}$ features per scan.

3.2.1.2 ResNet18-Based Feature Extraction

After having performed feature extraction using a wide variety of CNN-based architectures, it was determined that ResNet18 generated the best vector representations of the images based entirely on classification accuracy. A 71-layer ResNet18 model pre-trained on the ImageNet dataset was used to extract features from the raw scans. Weights from the fifth and last pooling layer were extracted and used as weights for classification between malignant and non-malignant scans. Note that the weights obtained using this approach are concatenated to generate a 512-element feature vector corresponding to each lymph node scan. Figure 12 provides a pictorial representation of the ResNet18 model that has been utilized in this work. The model comprises a total of 71 layers, out of which the trained weights from the "average pooling" layers are used for the purpose of classification.

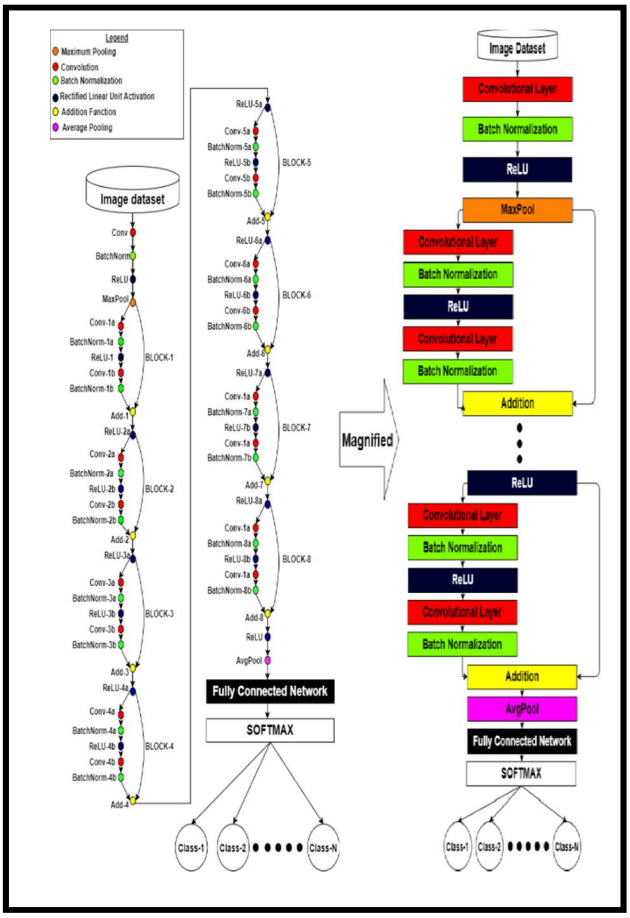


Fig 12. A pictorial representation of the ResNet18 model that has been utilized in this work.

3.2.2 Feature Selection Mechanism

We employed an ensemble technique to mine out the most consequential features from the feature vector obtained using the pre-trained model, in order to perform classification of malignant and non-malignant lymph nodes. Please note that the feature matrix has been arranged in such a manner that the rows depict the samples, and the columns are representative of the features. The feature selection process is as depicted in table 1. We employ the Random Forest algorithm, one of the most widely used algorithms in machine learning for the purpose of classification as well as feature selection. In the Python programming language, the Random Forest library comes with an in-built importance

generation function—which, based on the feature set and the dependent variable, returns the most important features along with their percentage of importance. Figure 13 provides a schematic depiction of the feature selection algorithm.

Table 1. Feature selection algorithm

Step 1. Sparsity filter: First we need to remove any missing values that might be present in the feature matrix. Through a procedure of trial and error, we determined that we should drop all features having at least 20% of the values missing (or zero). ***Step 2. Data imputer:*** Impute missing values of the remaining features by the average value of the respective feature. ***Step 3. Low coefficient of variation (CV) filter:*** This particular step is based on the idea that features with higher variance have more information contained within them. Coefficient of variation is standard deviation normalized by mean. This metric is very effective in taking care of the characteristic differences in the range of values amongst the features, thereby minimizing any bias that might have arisen by the utilization of raw feature values. Coefficient of variation is standard deviation normalized by mean: $\mathrm{CV} = \frac{\sigma}{\mu}$, where σ is the standard deviation, and μ is the mean. ***Step 4. High correlation filter:*** This step drops columns with a correlation value greater than 95%. The way that this is achieved is by calculating the correlation matrix of all of the feature pairs, and subsequently dropping one of the features with pairwise correlation value greater than 95% (the feature that has a lower correlation with the dependent variable, in this case the malignant/ non-malignant labels that have been provided along with the data). ***Step 5. Feature importance calculator:*** This is the most important step in the dimensionality reduction process. We make use of machine learning to mine out the features that are most consequential in forecasting the output variable (in this case, the malignant/ non-malignant labels).

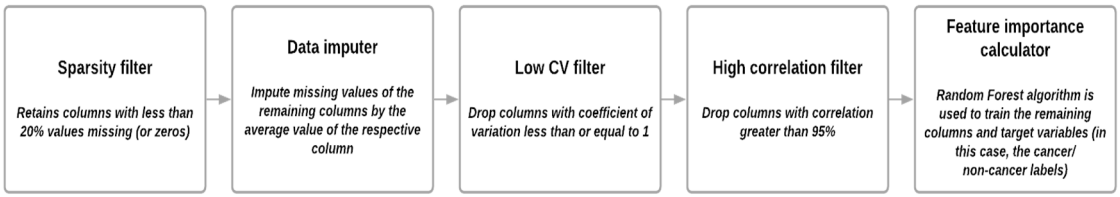


Fig 13. A flowchart of the feature selection algorithm

3.2.3 Machine Learning Classification

The salient features obtained from the feature selection algorithm (see section 4.2.2) were fed into a machine learning classifier to differentiate the malignant versus non-malignant lymph nodes. In order to effectively investigate the classification performance of the proposed model, 10-fold Cross Validation (CV) was utilized. The dataset was randomly shuffled and split evenly into 10 different groups. In each iteration, 9 of these groups were used for training and the remaining 1 for testing. The iteration process was repeated 10 times, each time the test set being a different group. The results obtained across the 10 iterations were finally averaged and reported. It must be noted that the groups were generated in such a manner as to ensure that the training and test sets were mutually exclusive of one another (that is, no overlap of samples). The 10-fold Decision Tree (DT) model was utilized for the purpose of classification.

3.3 Dataset

A radiological dataset was provided by Mayo Clinic, Arizona. The dataset comprised multiple de-identified gray-level MRI scans of lymph nodes, obtained using a varied range of image contrast types from a prospective clinical trial of 15 high-risk (Gleason $\geq$ 8) prostate cancer patients. The patients underwent prostate MRI before prostatectomy and pelvic lymph node dissection as part of a trial. Tissue was submitted for pathologist review. The location of each lymph node was confirmed, and labels of the pre-operative MRI data were generated. The labels were “positive” (meaning harboring metastatic cancer cells) and “negative” (no cancer metastases were found). There was a total of 126 lymph node images: 41 positives and 85 negatives.

The four MRI sequences, each with different tissue contrast characteristics, are as follows:

1. Apparent Diffusion Coefficient (ADC)

2. T2-weighted Fast Recovery Fast Spin Echo (FRFSE)

3. Pelvis (T2 FatSat)

4. MRI with Gadolinium Contrast (Water-GAD)

Figure 14 comprises a prostate MRI from the same patient; four MRI sequences shown.

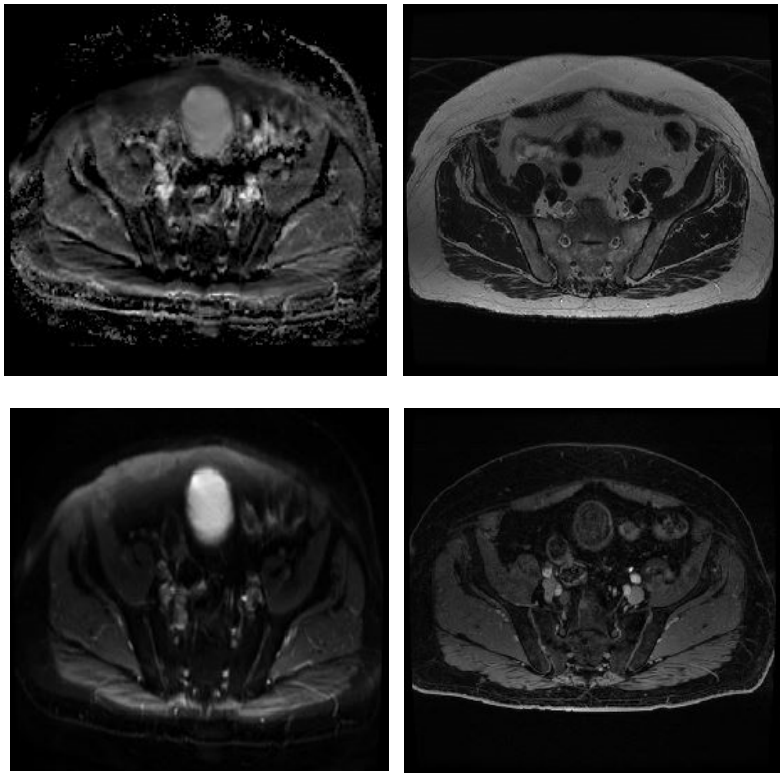


Fig 14. Prostate MRI from the same patient (patient number: 2663); four MRI sequences— Apparent Diffusion Coefficient (ADC), Fast Recovery Fast Spin Echo (FRFSE), Pelvis and MRI with Gadolinium contrast (Water-GAD).

3.4 Results

The ResNet18-based approach described before was utilized for feature extraction. For the purpose of comparison, two texture-based features were also extracted, namely Gabor and GLCM. Subsequently, feature selection was performed on each of these feature sets. The most important features were then classified using the 10-fold Decision Tree (DT) classifier.

The top 9 most important ResNet18-based features yield a highest classification accuracy of 76.19%, sensitivity of 79.76%, specificity of 69.05%, precision of 83.75% and F1-score of 81.71% on the 10-fold decision tree (DT) classifier. Our ResNet-based feature extraction framework was compared against two texture feature-based classification models namely, GLCM and Gabor. The decision tree classifier trained on 11 GLCM features (contrast, correlation, energy, entropy, homogeneity, variance, sum of average, sum of variance, sum of entropy, difference of variance, difference of entropy), after feature selection using the aforementioned algorithm, achieved a best classification performance on the top 5 most important features (contrast, energy, homogeneity, sum of variance and difference of entropy in that order)—with an accuracy of 61.90%, precision of 71.43% and F1-score of 72.73%. The same decision tree when trained on a set of 1960 Gabor features achieved a best performance on the top 15 features—with an accuracy of 65.08%, precision of 76.25% and F1-score of 74.84%.

Table 2 provides the classification performance of the 10-fold decision tree classifier obtained using the original ResNet18-based features, the GLCM features, and the Gabor-based features. Table 3 provides the classification performance of the 10-fold decision tree classifier obtained using the feature selected ResNet18-based features, the GLCM features, and the Gabor-based features. Further, to illustrate the efficacy of the feature selection

algorithm, figure 15 has been presented, which includes the receiver operating characteristics (ROC) curves of each of the three sets of features—namely, the ResNet18-based features, GLCM features and Gabor features—after classification using the 10-fold decision tree classifier. The ResNet18-based features show a marked improvement from an area under curve (AUC) of 57.82% on the 512 original features to an AUC of 94.59 on the 9 selected features. GLCM shows a slight improvement from an AUC of 92.11 on the 11 original features to an AUC of 95.12 on the 5 selected features. Gabor feature performances remain comparable between the original feature set of 1960 features which had an AUC of 99.20%, and the selected feature set of 15 features which had an AUC of 98.98%.

Classifier	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**F1-score (%)**
ResNet18-based features	57.14	62.35	46.34	70.67	66.25
GLCM	57.94	72.84	33.33	67.82	70.24
5 × 8 Gabor filter bank	59.52	67.47	47.50	72.73	70.00

Table 2. The performance of the 10-fold decision tree classifier obtained using the original ResNet18-based features, the GLCM features, and the Gabor-based features.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)
ResNet18-based features	76.19	79.76	69.05	83.75	81.71
GLCM	61.90	74.07	42.86	71.43	72.73
5 × 8 Gabor filter bank	65.08	73.49	52.50	76.25	74.84

Table 3. The performance of the 10-fold decision tree classifier obtained using the feature selected ResNet18-based features, the GLCM features, and the Gabor-based features.

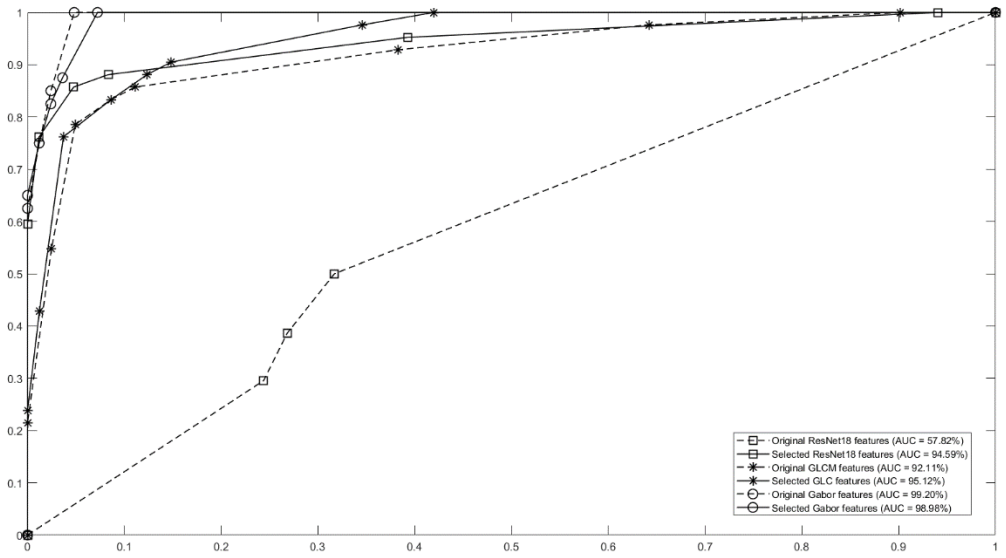


Fig 15. The receiver operating characteristics (ROC) curves of each of the three sets of features—namely, the ResNet18-based features, GLCM features and Gabor features—after classification using the 10-fold decision tree classifier.

3.5 Conclusion and Future Work

3.5.1 Conclusion

This case study proposes an integrated pipeline to help detect malignant lymph nodes in patients with prostate cancer. The accuracy obtained by our approach had a higher sensitivity and comparable specificity compared to sensitivity and specificity reported in prior studies relying on imaging alone. For example, [4] states that the area under receiver operating curve (AUC) varies between 0.69 and 0.81 for prostate cancer detection over multi-parametric MRIs, which includes diffusion-weighted imaging (DWI). The authors of [122] have designed an automatic deep CNN-based architecture to detect prostate cancer on diffusion-weighted magnetic resonance imaging (DWI). The database comprised DWI images of 427 patients (175 prostate cancer, 252 healthy patients). The model yielded an area under the receiver operating curve (AUC) of 0.87. [123], the runners up of the 2017 PROSTATEx challenge, designed a deep learning model called the XmasNet which was based on deep CNNs, for the purpose of performing classification on prostate cancer lesions utilizing 3D multiparametric MRIs. They achieve an AUC of 0.84. The database comprises 341 patients, each having Diffusion Weighted Images (DWI), Apparent Diffusion Coefficient (ADC) scans, Ktrans, and T2-weighted images (T2WI).

In medical imaging research, one of the primary challenges is the fact that interpretability varies greatly between radiologists. The PI-RADS architecture is pervasively utilized for the purpose of image interpretation. However, [124] exhibits the many impending issues of inter-observer interpretability associated with the PI-RADS model. Owing to the innate difficulties associated with identifying anomalies in prostate MRIs, there seems to be varying consensus among researchers and radiologists alike, with respect to determining the best identification methodologies. Sample size—as in the case with our study as well—

often proves to be an important factor that determines the selection of the classification model. While we know that machine learning performs better on datasets with limited samples, we also acknowledge the capability of deep models to extract more meaningful features. In [125], the authors demonstrated that on a dataset of multiparametric MRIs obtained from 52 prostate cancer patients, hierarchical clustering performed better than deep models in differentiating between normal and tumor prostate tissues.

While the statistical measures that we have employed to evaluate our model seem to work well, we do acknowledge that there might be more medically sound metrics for measuring performance which—to a certain degree—predict the chance of survival as well. However, in order to evaluate such parameters, we need to have an in-depth understanding of associated biomedical processes, as well as access to a wide variety of radiological features. The authors of [126] have identified certain measures and classifiers that outperform others when it comes to prostate cancer with lymph node metastasis. A sample size of 1400 patients—all with metastatic prostate cancer to lymph nodes—have been employed in this study. Univariate analysis revealed that age, Gleason score, radiotherapy history, T stage, log odds of metastatic lymph node (LODDS) classifier, lymph node ratio (LNR) classifier, and number of metastatic lymph node (NMLN) classifier except total number of lymph nodes examined (TNLE), are some of the most consequential predictors of patient survival. Multivariate analysis suggested that LODDS, LNR, and NMLN except TNLE classifiers are some of the most important parameters for measuring survival rate.

In addition, it is our intention to extend our analysis to subsequently be able to localize the lesion region in lymph node images where cancer has been detected. From previous experience with working on similar medical imaging datasets, we have noticed that having

an additional set of "difference features" helps in not only localizing the lesion region, but also monitoring the progression of lesion over time. This basically involves subtracting time-variant images from a fixed baseline data. In this case, since we have a control image of an unaffected portion of lymph node provided, corresponding to every cancerous region of the lymph node, we could use the control as a baseline image to then perform localization. However, for that to work effectively we need more positively labeled (or cancerous) samples; at the moment, we have too few lesion lymph node images to train a full-fledged model to perform localization. However, on acquiring more samples, if this sees fruition, then we should be able to use the classifier in cascade with the localization framework.

3.5.2 Future Work

While our initial analyses yield promising results, we acknowledge that there are several areas in which we could improve—both in terms of model design, as well as ways of generating more meaningful feature representations from datasets having limited samples and class imbalance. In this section, I will transcribe some of the difficulties that we have been able to identify, and how we plan on overcoming them in the future.

3.5.2.1 Feature Extraction

We plan on exploring the feature extraction domain in greater detail. We have extracted two texture-based features namely GLCM and Gabor. We believe that if we can devise an effective combination of manually extracted and automatically generated features, that will improve the performance of the proposed model significantly.

3.5.2.2 Data Augmentation

We intend to conduct an in-depth augmentation process that might help alleviate the

problem of small, imbalanced sample size. We implemented simple physical transformations on the data (rotation, flipping et cetera), but those methods have proven ineffective. For this particular task, we will most likely need to use Generative Adversarial Networks (GANs).

CHAPTER 4
CONCLUSION AND FUTURE WORK

4.1 Conclusion

In summary, we designed a deep learning-machine learning model—paired with a feature selection framework consisting of statistical (or unsupervised) and supervised techniques—to detect lymph node metastasis effectively from a small prostate cancer dataset. The pre-trained deep model was utilized for feature extraction, while the decision tree classifier was used for classification. The results obtained were compared against features generated using GLCM and Gabor-based texture features. The proposed ResNet18-based features outperformed the GLCM and Gabor features during classification. The ResNet18 features had a classification accuracy of 76.19%, sensitivity of 79.76%, specificity of 69.05%, precision of 83.75% and F1-score of 81.71%. In comparison, the GLCM features had an accuracy of 61.90%, sensitivity of 74.07%, specificity of 42.86%, precision of 71.43% and F1-score of 72.73%. The Gabor features had an accuracy of 65.08%, sensitivity of 73.49%, specificity of 52.50%, precision of 76.25% and F1-score of 74.84%.

Despite the efficacy of the proposed deep learning-based features in classifying lymph node metastasis from prostate cancer patients, there are several drawbacks of this model. Next we will discuss some of the shortcomings of the designed classification model, and also discuss ways to make it more effective in the future.

4.2 Future Work

4.2.1 Feature Extraction

4.2.1.1 Combination of Manual and Deep Model-Generated Features

GLCM and Gabor-based feature extraction techniques have been employed in this study. However, features generated by deep models performed better—both in terms of classification performance, and also when feature selection was done. However, employing a combination of deep as well as manually extracted texture-based features, is expected to significantly improve classification performance. The texture-based features are efficient in incorporating domain knowledge in the system, whereas the deep CNN-based features are efficient in generating effective image representations. This is an improvement that we expect to include in later versions of the classification model.

4.2.1.2 Generating a New Set of Difference Features

For each patient, we were provided a set of lesion as well as control scans. If we could devise a way to generate a set of difference features, that might not only help us segment the ROIs better, but could potentially also help generate a feature pertaining to progression of the lesion over time. However, in order to do that we must have one of the following:

1. A set of time varying scans
2. More lymph node scans with lesion ROIs

4.2.2 Data Augmentation

Simple physical transformations (such as rotating, flipping et cetera) proved to be ineffective in enhancing performance. Data augmentation, however, is critical in alleviating problems associated with small sample size and imbalanced classes—both of which being relevant problems in our prostate cancer dataset. Therefore, an effective data

augmentation technique would prevent these problems in the future, and help improve classification performance. Generative adversarial networks (GANs) are an effective and widely used technique for data augmentation [127, 128, 129, 130, 131], which we plan on implementing in the future.

4.2.3 Segmentation

While the proposed model performs well on classification tasks, it is not capable of localizing (segmenting) regions of interest on images. Therefore, manual ROI extraction is required before classification is performed—which is a cumbersome task for the radiology team, considering the small size of the lesion regions compared to the entire MRI scans. Therefore, we intend to incorporate an automatic segmentation step before classification is performed by the proposed deep learning-machine learning approach.

4.2.4 Model Generalizability

The proposed model performs well on MRI and CT scans, but poorly on PET imaging contrasts. Also, the thresholds and hyperparameters being used in the model are based simply on classification performance. Consequently, it requires considerable manual tweaking to train different datasets, even pertaining to the same imaging and disease types. We intend to update the model, thereby making it more robust to different image types and input datasets.

4.2.5 Model Performance Improvement

In this study, we have employed a very straightforward implementation of the machine learning classifier, namely 10-fold decision tree. For future works, we would like to implement more nuanced classifiers, and classifiers with regularization and hinge loss. That will not only help improve classification performance, but also make the model

more robust against overfitting.

Another critical area that we have not addressed is the class-imbalanced problem. In this dataset, we have 85 control and 41 lesion samples—which enables the model to learn the negative cases better than the positive ones, ultimately incorporating a bias towards the control cases. Sampling (over-sampling and under-sampling), boosting, bagging and noise filtering methods are used pervasively to counter the class-imbalance problem [132, 133, 134, 135]. The Synthetic Minority Over-Sampling Technique (SMOTE), along with its variants, is a common type of over-sampling technique that is used to generate balanced classes before performing classification [136, 137].

REFERENCES

1. Cohen, Joseph Paul, et al. "Covid-19 image data collection: Prospective predictions are the future." *arXiv preprint arXiv:2006.11988* (2020). DOI: https://arxiv.org/abs/2006.11988.

2. Aach, Til, Ulrich W. Schiebel, and Gerhard Spekowius. "Digital image acquisition and processing in medical x-ray imaging." Journal of Electronic Imaging 8.1 (1999): 7-22.

3. "Computed Tomography (CT) – Medical Radiations."" [Online]. Available: http://www.medicalradiation.com/types-of-medical-imaging/imaging-using-x-rays/computed-tomography-ct/ [Accessed: 11-Mar-2021]

4. Soares, Eduardo, et al. "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CTscans for SARS-CoV-2 identification." *medRxiv* (2020). Dataset available online: https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset

5. Buzug T.M. (2011) Computed Tomography. In: Kramme R., Hoffmann KP., Pozos R.S. (eds) Springer Handbook of Medical Technology. Springer Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74658-4_16

6. Kalender, Willi A., 2006, Phys. Med. Biol., Volume 51, Number 13.

7. Fleischmann, D., Boas, F.E. Computed tomography—old ideas and new technology. Eur Radiol 21, 510–517 (2011). https://doi.org/10.1007/s00330-011-2056-z

8. Brooks, Sharon L. "Computed tomography." Dental Clinics of North America 37.4 (1993): 575-590.

9. Brenner, David J., and Eric J. Hall. "Computed tomography—an increasing source of radiation exposure." New England Journal of Medicine 357.22 (2007): 2277-2284.

10. Smith-Bindman, Rebecca. "Is computed tomography safe." N Engl j Med 363.1 (2010): 1-4.

11. Hathcock, John T., and Russ L. Stickle. "Principles and concepts of computed tomography." Veterinary Clinics of North America: Small Animal Practice 23.2 (1993): 399-415.

12. De Chiffre, Leonardo, et al. "Industrial applications of computed tomography." CIRP annals 63.2 (2014): 655-677.

13. Vlaardingerbroek, Marinus T., and Jacques A. Boer. Magnetic resonance imaging: theory and practice. Springer Science & Business Media, 2013.

14. Young, Stuart W. "Magnetic resonance imaging: basic principles." (1987).

15. Geva, Tal. "Magnetic resonance imaging: historical perspective." Journal of

cardiovascular magnetic resonance 8.4 (2006): 573-580.

16. Morris, Shaine A., and Timothy C. Slesnick. "Magnetic resonance imaging." Visual Guide to Neonatal Cardiology (2018): 104-108.

17. Huettel, Scott A., Allen W. Song, and Gregory McCarthy. Functional magnetic resonance imaging. Vol. 1. Sunderland, MA: Sinauer Associates, 2004.

18. Terreno, Enzo, et al. "Challenges for molecular magnetic resonance imaging." Chemical reviews 110.5 (2010): 3019-3042.

19. Dill, T. "Contraindications to magnetic resonance imaging." Heart 94.7 (2008): 943-948.

20. Born, Jannis, et al. "POCOVID-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS)." *arXiv preprint arXiv:2004.12084* (2020). Dataset available online: https://github.com/jannisborn/covid19_pocus_ultrasound

21. Szabo, Thomas L. Diagnostic ultrasound imaging: inside out. Academic Press, 2004.

22. Wells, Peter NT. "Ultrasound imaging." *Physics in Medicine & Biology* 51.13 (2006): R83.

23. Chan, Vincent, and Anahi Perlas. "Basics of ultrasound imaging." Atlas of ultrasound-guided procedures in interventional pain management. Springer, New York, NY, 2011. 13-19.

24. Fenster, Aaron, Donal B. Downey, and H. Neale Cardinal. "Three-dimensional ultrasound imaging." Physics in medicine & biology 46.5 (2001): R67.

25. Tai, Y. F., and P. Piccini. "Applications of positron emission tomography (PET) in neurology." Journal of Neurology, Neurosurgery & Psychiatry 75.5 (2004): 669-676.

26. Bailey, Dale L., et al. Positron emission tomography. Vol. 2. London: Springer, 2005.

27. Weber, Wolfgang A., Norbert Avril, and Markus Schwaiger. "Relevance of positron emission tomography (PET) in oncology." Strahlentherapie und Onkologie 175.8 (1999): 356-373.

28. Gallamini, Andrea, Colette Zwarthoed, and Anna Borra. "Positron emission tomography (PET) in oncology." Cancers 6.4 (2014): 1821-1889.

29. Petersen, Anncatrine Luisa. "Liposomal delivery of radionuclides for cancer diagnostics and radiotherapy: From material development to in vivo applications using positron emission tomography (PET) imaging." (2012).

30. Kubota, Kazuo. "From tumor biology to clinical PET: a review of positron emission

tomography (PET) in oncology." Annals of nuclear medicine 15.6 (2001): 471-486.

31. Hillner, Bruce E., et al. "Impact of positron emission tomography/computed tomography and positron emission tomography (PET) alone on expected management of patients with cancer: initial results from the National Oncologic PET Registry." Journal of Clinical Oncology 26.13 (2008): 2155-2161.

32. Berti, Valentina, Alberto Pupi, and Lisa Mosconi. "PET/CT in diagnosis of movement disorders." Annals of the New York Academy of Sciences 1228 (2011): 93.

33. Maiti, Baijayanta, and Joel S. Perlmutter. "PET imaging in movement disorders." Seminars in nuclear medicine. Vol. 48. No. 6. WB Saunders, 2018.

34. Gerasimou, George P. "Molecular imaging (SPECT and PET) in the evaluation of patients with movement disorders." Nuclear Medicine Review 9.2 (2006): 147-153.

35. Zhang, Jing. "Mapping neuroinflammation in frontotemporal dementia with molecular PET imaging." Journal of neuroinflammation 12.1 (2015): 1-7.

36. Ishii, K. "PET approaches for diagnosis of dementia." American Journal of Neuroradiology 35.11 (2014): 2030-2038.

37. Quigley, Harriet, Sean J. Colloby, and John T. O'Brien. "PET imaging of brain amyloid in dementia: a review." International journal of geriatric psychiatry 26.10 (2011): 991-999.

38. Chen, Wei, et al. "18F-FDOPA PET imaging of brain tumors: comparison study with 18F-FDG PET and evaluation of diagnostic accuracy." Journal of Nuclear Medicine 47.6 (2006): 904-911.

39. Wong, Terence Z., Gert J. van der Westhuizen, and R. Edward Coleman. "Positron emission tomography imaging of brain tumors." Neuroimaging Clinics 12.4 (2002): 615-626.

40. Kratochwil, Clemens, et al. "Intra-individual comparison of 18F-FET and 18F-DOPA in PET imaging of recurrent brain tumors." Neuro-oncology 16.3 (2014): 434-440.

41. Toczek, M. T., et al. "PET imaging of 5-HT1A receptor binding in patients with temporal lobe epilepsy." Neurology 60.5 (2003): 749-756.

42. Gok, Beril, et al. "The evaluation of FDG-PET imaging for epileptogenic focus localization in patients with MRI positive and MRI negative temporal lobe epilepsy." Neuroradiology 55.5 (2013): 541-550.

43. Sarikaya, Ismet. "PET studies in epilepsy." American journal of nuclear medicine and molecular imaging 5.5 (2015): 416.

44. Lee, Jae Sung, et al. "PET evidence of neuroplasticity in adult auditory cortex of postlingual deafness." Journal of Nuclear Medicine 44.9 (2003): 1435-1439.

45. Wey, Hsiao-Ying, et al. "Insights into neuroepigenetics through human histone deacetylase PET imaging." Science translational medicine 8.351 (2016): 351ra106-351ra106.

46. Marchal, Gilles, et al. "PET imaging of cerebral perfusion and oxygen consumption in acute ischaemic stroke: relation to outcome." The Lancet 341.8850 (1993): 925-927.

47. Mayberg, Helen S., et al. "PET imaging of cortical S2 serotonin receptors after stroke: lateralized changes and relationship to depression." Am J Psychiatry 145.8 (1988): 937-943.

48. Evans, Nicholas R., et al. "PET imaging of atherosclerotic disease: advancing plaque assessment from anatomy to pathophysiology." Current atherosclerosis reports 18.6 (2016): 30.

49. Xiao, Yuling, et al. "Gold nanorods conjugated with doxorubicin and cRGD for combined anticancer drug delivery and PET imaging." Theranostics 2.8 (2012): 757.

50. Dong, Xinghua, et al. "Intelligent MoS2 nanotheranostic for targeted and enzyme-/pH-/NIR-responsive drug delivery to overcome cancer chemotherapy resistance guided by PET imaging." ACS applied materials & interfaces 10.4 (2018): 4271-4284.

51. Chakravarty, Rubel, et al. "Hollow mesoporous silica nanoparticles for tumor vasculature targeting and PET image-guided drug delivery." Nanomedicine 10.8 (2015): 1233-1246.

52. Gupta, N., P. M. Price, and E. O. Aboagye. "PET for in vivo pharmacokinetic and pharmacodynamic measurements." European Journal of Cancer 38.16 (2002): 2094-2107.

53. Scott, Andrew M., et al. "Pharmacodynamic analysis of tumour perfusion assessed by 15 O-water-PET imaging during treatment with sunitinib malate in patients with advanced malignancies." EJNMMI research 2.1 (2012): 1-10.

54. Wang, Y., et al. "Sodium fluoride PET imaging as a quantitative pharmacodynamic biomarker for bone homeostasis during anti-DKK1 therapy for multiple myeloma." Blood cancer journal 7.10 (2017): e615-e615.

55. Shabbeer Basha, S. H., et al. "Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification." *arXiv preprint arXiv:1902.02771* (2019).

56. Dou, Qi, et al. "Automatic lesion detection with three-dimensional convolutional neural

networks." *Biomedical Information Technology*. Academic Press, 2020. 265-293. DOI: https://www.sciencedirect.com/science/article/pii/B9780128160343000092

57. Hajabdollahi, Mohsen, et al. "Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network." *Biomedical Signal Processing and Control* 57 (2020): 101792.

58. Gehlot, Shiv, Anubha Gupta, and Ritu Gupta. "SDCT-AuxNetθ: DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis." *Medical Image Analysis* 61 (2020): 101661.

59. Y. Wang *et al*., "Breast Cancer Image Classification via Multi-Network Features and Dual- Network Orthogonal Low-Rank Learning," in *IEEE Access*, vol. 8, pp. 27779-27792, 2020, DOI: https://ieeexplore.ieee.org/abstract/document/8950452

60. Buyssens, Pierre, Abderrahim Elmoataz, and Olivier Lézoray. "Multiscale convolutional neural networks for vision–based classification of cells." *Asian Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2012.

61. Agnes, S.A., Anitha, J., Pandian, S.I.A. et al. Classification of Mammogram Images Using Multiscale all Convolutional Neural Network (MA-CNN). J Med Syst 44, 30 (2020). DOI: https://doi.org/10.1007/s10916-019-1494-z

62. Dolz, Jose, et al. "Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation." *Computerized Medical Imaging and Graphics* 79 (2020): 101660.

63. Kumar S.N., Lenin Fred A., Ajay Kumar H., Sebastin Varghese P., Jacob S.A. (2020) Segmentation of Anomalies in Abdomen CT Images by Convolution Neural Network and Classification by Fuzzy Support Vector Machine. In: Bhattacharyya S., Konar D., PlatosJ., Kar C., Sharma K. (eds) Hybrid Machine Intelligence for Medical Image Analysis.Studies in Computational Intelligence, vol 841. Springer, Singapore, DOI: https://doi.org/10.1007/978-981-13-8930-6_7, Print ISBN: 978-981-13-8929-0, Online ISBN: 978-981-13-8930-6

64. Liang, Gaobo, et al. "Combining convolutional neural network with recursive neural network for blood cell image classification." *IEEE Access* 6 (2018): 36188-36197. DOI: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8402091

65. Rezaei, Mina, Haojin Yang, and Christoph Meinel. "Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation." *Multimedia Tools and Applications* (2019): 1-20. DOI: https://doi.org/10.1007/s11042-019-7305-1

66. Paserin O., Mulpuri K., Cooper A., Hodgson A.J., Garbi R. (2018) Real Time RNN Based 3D Ultrasound Scan Adequacy for Developmental Dysplasia of the Hip. In:

Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol 11070. Springer, Cham. DOI: https://doi.org/10.1007/978- 3-030-00928-1_42

67. Cui, Ruoxuan, Manhua Liu, and Gang Li. "Longitudinal analysis for Alzheimer's disease diagnosis using RNN." *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018.

68. Fehling, Mona Kirstin, et al. "Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network." *Plos one* 15.2 (2020): e0227791. DOI: https://doi.org/10.1371/journal.pone.0227791

69. He, Sheng, et al. "Brain Age Estimation Using LSTM on Children's Brain MRI." *arXiv preprint arXiv:2002.09045* (2020). DOI: https://arxiv.org/abs/2002.09045

70. Begum, S. Salma, and D. Rajya Lakshmi. "Combining optimal wavelet statistical texture and recurrent neural network for tumour detection and classification over MRI." *Multimedia Tools and Applications* (2020): 1-22. DOI: https://doi.org/10.1007/s11042-020-08643-w

71. Yang, Zhengshi, et al. "A robust deep neural network for denoising task-based fMRI data: An application to working memory and episodic memory." *Medical Image Analysis* 60 (2020):101622. DOI: https://doi.org/10.1016/j.media.2019.101622

72. Ghassemi, Navid, Afshin Shoeibi, and Modjtaba Rouhani. "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images."

73. Onishi, Yuya, et al. "Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks." *International journal of computer assisted radiology and surgery* 15.1 (2020): 173-178. DOI: https://doi.org/10.1007/s11548-019-02092-z

74. Murugesan, Balamurali, et al. "A context based deep learning approach for unbalanced medical image segmentation." *arXiv preprint arXiv:2001.02387* (2020). DOI: https://arxiv.org/abs/2001.02387

75. Yang, Dong, et al. "Segmentation using adversarial image-to-image networks." *Handbookof Medical Image Computing and Computer Assisted Intervention*. Academic Press, 2020. 165-182. DOI: https://doi.org/10.1016/B978-0-12-816176-0.00012-0

76. Zhang, Yue, et al. "Unsupervised X-ray Image Segmentation with Task Driven Generative Adversarial Networks." *Medical Image Analysis* (2020): 101664. DOI: https://doi.org/10.1016/j.media.2020.101664

77. Kim, Hee-Joung, and Donghoon Lee. "Image denoising with conditional generative adversarial networks (CGAN) in low dose chest images." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 954 (2020): 161914. DOI: https://doi.org/10.1016/j.nima.2019.02.041

78. Qin, Tiexin, et al. "Automatic Data Augmentation via Deep Reinforcement Learning for Effective Kidney Tumor Segmentation." *arXiv preprint arXiv:2002.09703* (2020). DOI: https://arxiv.org/abs/2002.09703

79. Ali, Issa, et al. "Lung nodule detection via deep reinforcement learning." *Frontiers in oncology* 8 (2018): 108. DOI: https://doi.org/10.3389/fonc.2018.00108

80. Maicas G., Carneiro G., Bradley A.P., Nascimento J.C., Reid I. (2017) Deep Reinforcement Learning for Active Breast Lesion Detection from DCE-MRI. In: Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (eds) Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science, vol 10435. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-66179-7_76

81. Liu, Zhuo, et al. "Deep reinforcement learning with its application for lung cancer detection in medical internet of things." *Future Generation Computer Systems* 97 (2019): 1-9. DOI: https://www.sciencedirect.com/science/article/pii/S0167739X19303772

82. Kainz, Bernhard. "Evaluating Reinforcement Learning Agents for Anatomical Landmark Detection." DOI: https://www.sciencedirect.com/science/article/pii/S1361841518306121

83. F. Ghesu et al., "Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 1, pp. 176-189, 1 Jan. 2019.

84. Thomas, Rajat M., et al. "Dealing with missing data, small sample sizes, and heterogeneity in machine learning studies of brain disorders." Machine Learning. Academic Press, 2020. 249-266.

85. Hemelings, Ruben, et al. "Accurate prediction of glaucoma from colour fundus imageswith a convolutional neural network that relies on active and transfer learning." Actaophthalmologica (2019).

86. Behzadi-khormouji, Hamed, et al. "Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images." Computer methods and programs in biomedicine 185 (2020): 105162.

87. Wang, Hongyu, and Yong Xia. "Chestnet: A deep neural network for classification of thoracic diseases on chest radiography." arXiv preprint arXiv:1807.03058 (2018). DOI: https://arxiv.org/abs/1807.03058

88. Chouhan, Vikash, et al. "A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images." Applied Sciences 10.2 (2020): 559. DOI: https://doi.org/10.3390/app10020559

89. Han, W., et al. "Deep Transfer Learning and Radiomics Feature Prediction of Survival of Patients with High-Grade Gliomas." American Journal of Neuroradiology 41.1 (2020): 40-48. DOI: http://www.ajnr.org/content/41/1/40

90. Mabu, Shingo, et al. "Investigating the Effects of Transfer Learning on ROI-based Classification of Chest CT Images: A Case Study on Diffuse Lung Diseases." Journal of Signal Processing Systems (2020): 1-7. DOI: https://doi.org/10.1007/s11265-019-01499-w

91. Huang, Xufeng, et al. "Deep Transfer Convolutional Neural Network and Extreme Learning Machine for Lung Nodule Diagnosis on CT images." arXiv preprint arXiv:2001.01279 (2020). DOI: https://arxiv.org/abs/2001.01279

92. Barz, Bjorn, and Joachim Denzler. "Deep learning on small datasets without pre-training using cosine loss." The IEEE Winter Conference on Applications of Computer Vision. 2020.

93. Bassi, Pedro RAS, and Romis Attux. "A Deep Convolutional Neural Network for COVID-19 Detection Using Chest X-Rays." arXiv preprint arXiv:2005.01578 (2020).

94. Apostolopoulos, Ioannis D., and Tzani A. Mpesiana. "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks." Physical and Engineering Sciences in Medicine (2020).

95. Rajpurkar, P., Park, A., Irvin, J. et al. AppendiXNet: Deep Learning for Diagnosis of Appendicitis from A Small Dataset of CT Exams Using Video Pretraining. Sci Rep 10, 3958 (2020). DOI: https://doi.org/10.1038/s41598-020-61055-6

96. “USCS Data Visualizations - CDC.” [Online]. Available: https://gis.cdc.gov/Cancer/USCS/DataViz.html. [Accessed: 01-Oct-2020].

97. “Prostate cancer - Symptoms and causes - Mayo Clinic.” [Online]. Available: https://www.mayoclinic.org/diseases-conditions/prostate-cancer/symptoms-causes/syc-20353087. [Accessed: 01-Oct-2020].

98. G. Murphy, M. Haider, S. Ghai, and B. Sreeharsha, “The Expanding Role of MRI in Prostate Cancer,” Am. J. Roentgenol., vol. 201, no. 6, pp. 1229–1238, Dec. 2013.

99. A. M. Hövels et al., “The diagnostic accuracy of CT and MRI in the staging of pelvic

lymph nodes in patients with prostate cancer: a meta-analysis," Clin. Radiol., vol. 63, no. 4, pp. 387–395, 2008.

100. Wu, Teresa, et al. "Quantitative Imaging System for Cancer Diagnosis and Treatment Planning: An Interdisciplinary Approach." Leading Developments from INFORMS Communities (2017): 153-177.

101. Atupelage, Chamidu, et al. "Computational grading of hepatocellular carcinoma using multifractal feature description." Computerized Medical Imaging and Graphics 37.1 (2013): 61-71.

102. Han, Fangfang, et al. "Texture feature analysis for computer-aided diagnosis on pulmonary nodules." Journal of digital imaging 28.1 (2015): 99-115.

103. Chen, Xin, et al. "Differentiation of true-progression from pseudoprogression in glioblastoma treated with radiation therapy and concomitant temozolomide by GLCM texture analysis of conventional MRI." Clinical imaging 39.5 (2015): 775-780.

104. Reyad, Yasser A., Mohamed A. Berbar, and Muhammad Hussain. "Comparison of statistical, LBP, and multi-resolution analysis features for breast mass classification." Journal of medical systems 38.9 (2014): 1-15.

105. Gangeh, Mehrdad J., et al. "Computer aided theragnosis using quantitative ultrasound spectroscopy and maximum mean discrepancy in locally advanced breast cancer." IEEE transactions on medical imaging 35.3 (2015): 778-790.

106. Cai, Lingyun, et al. "Robust phase-based texture descriptor for classification of breast ultrasound images." Biomedical engineering online 14.1 (2015): 1-21.

107. Lu, Wei, Zhe Li, and Jinghui Chu. "A novel computer-aided diagnosis system for breast MRI based on feature selection and ensemble learning." Computers in biology and medicine 83 (2017): 157-165.

108. Suganthi, S. S., and S. Ramakrishnan. "Analysis of breast thermograms using gabor wavelet anisotropy index." Journal of medical systems 38.9 (2014): 1-7.

109. Dhillon, Anamika, and Gyanendra K. Verma. "Convolutional neural network: a review of models, methodologies and applications to object detection." Progress in Artificial Intelligence 9.2 (2020): 85-112.

110. Hajabdollahi, Mohsen, et al. "Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network." Biomedical Signal Processing and Control 57 (2020): 101792.

111. Razeghi, Orod, et al. "CemrgApp: An interactive medical imaging application with image processing, computer vision, and machine learning toolkits for cardiovascular research." SoftwareX 12 (2020): 100570.

112. Dou, Qi, et al. "Automatic lesion detection with three-dimensional convolutional neural networks." Biomedical Information Technology. Academic Press, 2020. 265-293. Online link: https://www.sciencedirect.com/science/article/pii/B9780128160343000092

113. Hajabdollahi, Mohsen, et al. "Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network." Biomedical Signal Processing and Control 57 (2020): 101792. Online link: https://www.sciencedirect.com/science/article/pii/S1746809419303738

114. Zhou, Li-Qiang, et al. "Lymph node metastasis prediction from primary breast cancer US images using deep learning." Radiology 294.1 (2020): 19-28.

115. Lee, J.H., Ha, E.J. & Kim, J.H. Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. Eur Radiol 29, 5452–5457 (2019).

116. Wessels, Frederik, et al. "Deep learning approach to predict lymph node metastasis directly from primary tumor histology in prostate cancer." BJU international (2021).

117. Wu, Shaoxu, et al. "A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer." Clinical Cancer Research 23.22 (2017): 6904-6911.

118. Wu, Shaoxu, et al. "Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer." EBioMedicine 34 (2018): 76-84.

119. Haghighat, Mohammad, Saman Zonouz, and Mohamed Abdel-Mottaleb. "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification." Expert Systems with Applications 42.21 (2015): 7905-7916.

120. Shen, LinLin, Li Bai, and Michael Fairhurst. "Gabor wavelets and general discriminant analysis for face identification and verification." Image and Vision Computing 25.5 (2007): 553-563.

121. Liu, Chengjun, and Harry Wechsler. "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition." IEEE Transactions on Image processing 11.4 (2002): 467-476.

122. Yoo, S., Gujrathi, I., Haider, M.A. et al. Prostate Cancer Detection using Deep Convolutional Neural Networks. Sci Rep 9, 19518 (2019). https://doi.org/10.1038/s41598-019-55972-4

123. Liu, Saifeng, et al. "Prostate cancer diagnosis using deep learning with 3D multiparametric MRI." Medical imaging 2017: computer-aided diagnosis. Vol. 10134. International Society for Optics and Photonics, 2017.

124. Rosenkrantz, Andrew B., et al. "Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists."

Radiology 280.3 (2016): 793-804.

125. Akamine, Yuta, et al. "Application of hierarchical clustering to multi-parametric MR in prostate: Differentiation of tumor and normal tissue with high accuracy." Magnetic Resonance Imaging 74 (2020): 90-95.

126. Jin, Shengming, et al. "Comparison of different lymph node staging schemes in prostate cancer patients with lymph node metastasis." International Urology and Nephrology 52.1 (2020): 87-95.

127. Antoniou, Antreas, Amos Storkey, and Harrison Edwards. "Data augmentation generative adversarial networks." arXiv preprint arXiv:1711.04340 (2017).

128. Frid-Adar, Maayan, et al. "Synthetic data augmentation using GAN for improved liver lesion classification." 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018.

129. Waheed, Abdul, et al. "Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection." Ieee Access 8 (2020): 91916-91923.

130. Frid-Adar, Maayan, et al. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification." Neurocomputing 321 (2018): 321-331.

131. Lim, Swee Kiat, et al. "Doping: Generative data augmentation for unsupervised anomaly detection with gan." 2018 IEEE International Conference on Data Mining (ICDM). IEEE, 2018.

132. Rekha, Gillala, Amit Kumar Tyagi, and V. Krishna Reddy. "Solving class imbalance problem using bagging, boosting techniques, with and without using noise filtering method." International Journal of Hybrid Intelligent Systems 15.2 (2019): 67-76.

133. Ling C.X., Sheng V.S. (2011) Class Imbalance Problem. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA.

134. D. Devi, S. K. Biswas and B. Purkayastha, "A Review on Solution to Class Imbalance Problem: Undersampling Approaches," 2020 International Conference on Computational Performance Evaluation (ComPE), 2020, pp. 626-631, doi: 10.1109/ComPE49325.2020.9200087.

135. Leevy, J.L., Khoshgoftaar, T.M., Bauder, R.A. et al. A survey on addressing high-class imbalance in big data. J Big Data 5, 42 (2018). https://doi.org/10.1186/s40537-018-0151-6

136. Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

137. Han, Hui, Wen-Yuan Wang, and Bing-Huan Mao. "Borderline-SMOTE: a new

over-sampling method in imbalanced data sets learning." International conference on intelligent computing. Springer, Berlin, Heidelberg, 2005.

BIOGRAPHICAL SKETCH

Suryadipto Sarkar was born in Kolkata, India. He finished his secondary education from St. Xavier's Collegiate School, India. After that, he did his Bachelor of Technology (B.Tech.) from Manipal Institute of Technology, India. Subsequently, he worked for two years at TEKsystems Global Services, India as a Software Developer. He will receive his Master of Science (MS) in Computer Engineering from Arizona State University in the Summer of 2021.