Comparative Analysis of Deep Learning and Machine Learning Techniques in Predicting Radiation Pneumonitis

MSc Research Project
MSc. Data Analytics

Olawaunmi Sunday Anota
Student ID: x19239149

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Olawaunmi Sunday Anota
Student ID:	x19239149
Programme:	MSc. Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Noel Cosgrave
Submission Due Date:	16/08/2021
Project Title:	Comparative Analysis of Deep Learning and Machine Learning Techniques in Predicting Radiation Pneumonitis
Word Count:	5470
Page Count:	22

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

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

Signature:	
Date:	17th September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

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

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

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

Comparative Analysis of Deep Learning and Machine Learning Techniques in Predicting Radiation Pneumonitis

Olawaunmi Sunday Anota
x19239149

Abstract

Radiation pneumonitis (RP) is a form of lung damage induced by an irritant that develops when patients with Non-small Lung Cancer (NSCLC) get radiation therapy. Due to the complexity of computed tomography (CT) scan images, different transformations, augmentation, and normalization approaches are used in the data preparation. The goal of this research is to compare the performance of deep learning techniques with that of machine learning in predicting radiation pneumonitis in patients with Non-Small Cell Lung Cancer. The Cancer Imaging Archive (TCIA) 4D-Lung dataset containing 1699 images was adopted. In this study, three classification models were implemented- VGG16, Capsule Neural Network (CapsuleNet) and Support Vector Machine (SVM) based on deep learning and machine learning with the goal of creating a binary classifier that can predict radiation pneumonitis in Non-Small Cell Lung Cancer patients and the obtained outcome was compared. The Sensitivity and Specificity evaluation metrics of all the implemented classifier models are obtained in this study. To increase models performance, several parameter tuning was employed. From the implementation of models, it is shown that VGG16 had the best performance output of sensitivity 100% and specificity 95%.

1 Introduction

1.1 Background and Motivation

Lung cancer has been noted to be deadliest disease that affects mankind across the globe, with an estimate of 2.21 million cases and accounting for about 1.8 million deaths in 2020 [1]. Lung cancer is named for malignant cells that may be seen under a microscope in traditional oncology. The two major histological subtypes of lung cancer are small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), accounting for 87% of patients. Adenocarcinoma, Squamous cell carcinoma, and large cell carcinoma are the three major subtypes of NSCLC. Due to low survival rates and inter-patient variability, managing these individuals is difficult. These individuals present with a wide range of symptoms, from resectable tumors with microscopic lymph node metastases to bulky, unresectable disease. One of the most prevalent adverse effects of thoracic radiotherapy in patients

[1]https://www.who.int/news-room/fact-sheets/detail/cancer

affected by NSCLC is radiation pneumonitis (RP) which occurs between 1 to 6 months of radiation therapy treatment has been completed. Shen et al. [2019] estimated that approximately 75% of lung cancer patients receive radiation therapy as a treatment alone or in combination with surgery. Furthermore, radiation pneumonitis (RP) is a dose-limiting toxicity of non-small cell lung cancer, it affects about 5 to 20 percent of patients and restricts the highest dose that may be given, lowering tumour control probability (TCP) and perhaps causing dyspnea and poor quality of life [Giuranno et al., 2019]. Machine learning is a subfield of artificial intelligence (AI) that can learn from massive dataset by applying algorithms that can explore and predict future events and outcomes. In the aspect of healthcare, deep learning attempts to enhance the interpretation of medical data, resulting in faster workflow, fewer medical mistakes, lower costs, and better consultation and improved human health. In addition, in recent years, the application of deep learning approaches in radiotherapy has become popular which has aided medical oncologist in reducing the adverse effect from the treatment by way of predicting the outcomes. For the objectives of patient stratification, disease rating, prognosis, and treatment toxicities, machine learning has a substantial influence on the creation of novel prediction models for cancer diagnosis to enhance patient care [Kleppe et al., 2021]. Deep learning approaches have been found to identify well-established predictors of the progression of systemic radiation pneumonitis in NSCLC. Once the data is properly curated, it is possible to detect RP in NSCLC patients using data-driven machine learning techniques. RP has shown to be a significant stumbling block to radiation therapy treatment in lung malignancies, it has become a testing ground for oncologists to experiment with modern machine learning approaches.
VGG16, CapsuleNet and Support Vector Machine models will be implemented comparatively in predicting RP in patients with NSCLC. This study will examine and evaluate the models using sensitivity and specificity metrics to ascertain the true positives and true negatives respectively which are of core importance to the investigation of this study. The outcome of the result will aid radiation oncologist in accurately diagnosing and consultation of NSCLC patients will be less stress.

1.2 Research Question

"Can deep learning techniques such as VGG16, CapsuleNet provide a significant improvement over SVM in predicting pneumonitis beam in NSCLC patients?"

1.3 Research Objective

Table 1: Research Goal

S/N	Discreption	Metrics
1	Critically examine the literature review of the prediction of radiation pneumonitis in NSCLC from 2015 to 2021	
2	Exploratory data analysis about the data to get an in-depth insight about the data	
3	Implementing VGG16 model and evaluating the result	Sensitivity and Specificity
4	Implementing CapsuleNet model and evaluating the result	Sensitivity and Specificity
5	Implementing SVM model and evaluating the result	Sensitivity and Specificity
6	Comparative analysis of the models result and interpret findings	Sensitivity and Specificity

The goal of the research is to apply cutting-edge, comprehensive, and in-depth scientific techniques to predict the RP in NSCLC patients. In line with prior research relevant to the project, only few researchers had utilised an advanced deep learning technique for predicting RP in NSCLC. This research project is structured as follows: Section 2 covers related work in the oncology field in predicting RP in cancer patients, the research methodology is illustrated and explained in section 3, section 4 gives an in-depth illustration of the design architecture, Implementation of the models is explained in section 5, section 6 considers evaluation and results, and section 7 covers the conclusion and future work.

2 Related Work

RP prediction in NSCLC is an intriguing field of study to radiation oncologist where various deep learning techniques and ensemble approaches are being applied on Computed Tomography (CT) scans. RP may have adverse effect on NSCLC patients by way of shortness of breath, chest pain and cough [2], as a result, RP prediction in NSCLC is critical, and deep learning algorithms are used to analyse the NSCLC patients CT scans to figure out which patient is liable of having the side effect from the radiation therapy treatment. This section will focus, critically examine the various deep learning techniques, and machine learning methods of recent literatures in this problem domain.

2.1 Prediction of Radiation Pneumonitis in NSCLC Using Machine Learning

Applying machine learning approaches to the development of systematic radiation pneumonitis can detect known predictors [Luna et al., 2019]. The precision of radiotherapy toxicity prediction has assisted clinicians in choosing the optimal treatment option for patients with radiation pneumonitis who have NSCLC. According to the above study, a unique machine learning technique was utilized to identify 32 continuous and categorical characteristics per patient to detect predictive factors for the development of RP, using optimally trained decision stumps, univariate analysis was utilised to find statistically important features and their related pneumonitis beam thresholds, while Mediboost were used to select features for multivariate analysis.
In another study by Chao et al. [2018], out of the 197 stage 1 NSCLC patients who underwent stereotactic body radiotherapy, 25 patients recorded tumour, dosimetric features and 11 of whom suffered common terminology criteria for adverse events (CTCAE) 4.0 grade 2 chest wall discomfort. Individual feature thresholds for chest wall syndrome (CWS) were determined using decision tree modelling, independent multivariate techniques were used to identify significant characteristics. Out of bag estimation using Random forests (RF) and bootstrapping (100 iterations) with decision trees were adopted. Applying the learning curve experiments, the dataset showed self-consistency and provided a realistic model for chest wall syndrome analysis. Likewise, Yakar et al. [2021] examined 193 stage III lung cancer patients who had radiation therapy and chemotherapy treatment between the year 2014 to 2020. The pneumonitis beam evaluation was conducted using the Common Terminology Criteria for Adverse Events (CTCAE) 5.0 grading system. To produce a balanced data set, a synthetic minority oversampling approach was utilized, following the correlation analysis, a permutation-based technique was used to choose the variables. In 51 of the 193 test subjects, pneumonitis beam was discovered. The machine learning approach employed in this study were logistic regression, artificial neural networks, extreme gradient boosting (XGBoost), support vector machines, random forest, gaussian naive bayes, and light gradient boosting machine (LGBM). The result from the study showed that LGBM algorithm exhibited the greatest accuracy in predicting pneumonitis beam when the clinical and dosimetric data were combined and outperformed other machine learning models with an accuracy of 85 percent, Specificity of 50 percent and Sensitivity of 97percent.
Valdes et al. [2016] developed a patient-specific large data clinical decision approach to predict radiation pneumonitis in stage I NSCLC patients who had stereotactic body radiation treatment (SBRT). A cohort of 201 lung cancer patients was utilized in the study to assess the efficacy of three different algorithms: decision trees, random forest, and random under-sampling (RUS) Boost. The carbon monoxide diffusion potential of the lung and the dose to the heart, trachea, and bronchus were the most important features

[2]https://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/managing-side-effects/radiation-pneumonitis/?region=bc

for radiation pneumonitis prediction, according to the feature collection. The data set's quality is crucial because the machine learning algorithm would learn the parameters from the available data. The study's drawback is that if the training data set is sparse, the model is unlikely to acquire a representative set of parameters that can be used in scenarios outside of the dataset. Furthermore, Moran et al. [2017] looked at the feasibility of utilizing computed tomography (CT)-based radiomic features to define post-SBRT lung damage, as well as using dose-response curves to examine the relationship between improvements in radiomic feature values and cumulative dosage, using just Gray Level Co-occurrence Matrix (GLCM) functions, the researcher were able to get AUC values in the range of 0.64-0.75, indicating that eight out of nine features showed a significant dose-response association, implying that post-SBRT lung damage could be objectively evaluated.

Luo et al. [2017] adopted the Bayesian Network (BN) methodology to analysis the biophysical signalling pathways affecting pneumonitis beam grade 2 from a heterogeneous dataset that included single nucleotide polymorphisms, microRNAs, cytokines, clinical data, and radiation treatment plans before and during radiotherapy. The BN methodology mainly depended on a large-scale Markov blanket (MB) method to pick significant predictors, and K-fold cross-validation was used to minimize overfitting.

Krafft et al. [2019] used computed tomography to extract radiomic features from 192 NSCLC patients, as well as clinical and dosimetric parameters, to develop a predictive model for radiation pneumonitis. Eighty percent of the 192 patients received intensity-modulated radiation therapy (IMRT), while the rest received 3D-cathode-ray tube radiation therapy (CRT), a LASSO logistic regression classifier generated an average AUC of 0.68 when compared to models without image features.

Du et al. [2019] analysed a total of 118 lung cancer patients who had radiation therapy, generalized linear models through Lasso and ElasticNet regularization (GLMNET) were applied to evaluate 700,000 single-nucleotide polymorphism (SNP) sites to check whether they had any synergistic impact on pneumonitis beam risk prediction. A multiple linear regression model known as Radiation Pneumonitis Index (RPI) were developed based on the outcome of the research for the evaluation of grade 2 pneumonitis beam risk. The outcome from the analysis showed 92% sensitivity and 100% specificity which can correctly differentiate the pneumonitis beam population. In addition, Yu et al. [2019] applied statistical analysis to determine predictive cytokines from the evaluated 131 NSCLC patients out of which 17% had pneumonitis beam grade 2, a generalized linear model (GLM) for predicting pneumonitis beam grade 2 risk was developed using a machine learning technique, in a fully independent test set, the model prediction ability was confirmed with an accuracy of 80%, specificity 77% and sensitivity 100%. This research work established and validated a comprehensive model for predicting pneumonitis beam grade 2 before radiotherapy by combining inflammatory cytokines with clinical factors.

Yan and Wang [2020] analysed fifty NSCLC patients CT scans treated with radiotherapy using various machine models to predict the tumour responses, from the analysis it showed that lower-order features inside the tumour have a stronger predictive capacity that higher-order feature.

Yu et al. [2021] developed and validated a weighted-support vector machine classifier that incorporates circulating Chemokine (C-C motif) ligands CCL4 levels with important dosimetric and clinical factors to predict radiation pneumonitis . The analysis had an accuracy of 75%, one limiting factor from the author's research sample size data was small to justify the result of the findings.

2.2 Prediction of Radiation Pneumonitis in Cancer Patients using Deep Learning Approaches

Liang et al. [2020], developed a prediction model using a Convolutional 3D neural network, the neural network was pre-trained using UCF101 video dataset. From the analysis, C3D prediction model performance was compared with 3 multivariate Logistic regression on the data sample containing 70 NSCLC patients treated with volumetric modulated arc treatment. The outcome showed C3D neural network outperforming all other models. The goal was to develop a dose distribution-based prediction model and investigate the relationship between pneumonitis beam incidence and high order dose distribution features. They were few drawbacks from the research, data sample size was very low which made training the complex convolutional neural network from the scratch difficult and clinical significance of the observation was unknown. However, Chang et al. [2020] carried out a comparative analysis of the efficiency of a deep learning approach to a static doismetric model and general linear model in predicting stereotactic body radiation therapy toxicity in the 351 lung cancer patients used in the test. The result showed deep learning technique outperforming the other models.

Huang et al. [2021] applied a hybrid model that included a Fuzzy Clustering Means and Neural Network to predict pneumonitis beam using 4-dimensional computed tomography ventilation image based dosimetric parameters in NSCLC patients, from the analysis, the combination of the two models gave a better prediction when compared to conventional neural network. There were some notable setbacks to this study, the clusters of samples are mapped onto a two-dimensional space with a reduced gap between them and the sample data size used was small.

Xu et al. [2019] analysed the time series CT scans of NSCLC patients using transfer learning of CNN with Recurrent Neural Network (RNN) to validate the assumption that deep learning networks would accurately predict clinical outcomes, the main goal of the study was to classify NSLC patients into two categories based on their mortality risk. A major limitation of this research is with limited sample data size, it was proven that machine learning based on engineered features outperformed the deep learning model.

To identify lung nodules from heterogeneous in CT images, Wang et al. [2017] presented a data-based approach termed the Central Focused Convolutional Neural Network (CF-CNN). The proposed model captures a diverse range of nodule-sensitive features from both 3D and 2D CT scans at the same time, which is the basis for this study's technique. To achieve this aim, a central integration layer and a multi-scale patch learning approach were utilized, and the study revealed a sensitivity of 82 percent.

Jiang et al. [2019] devised a multi-scale CNN technique for volumetric cross-section pulmonary tumours that allows for precise, automated tumour volume diagnosis and quantification, tumour localisation has been found to have a sensitivity of 85%.

Velec et al. [2017] applied the Principal Component Analysis (PCA) to extract features from entire dose-volume histograms (DVHs) for the estimate of radiation exposure in the liver. According to the author, a single architecture of Actuarial Deep Learning Neural Network (ADNN) model can predict various endpoints of radiation pneumonitis and local control. Heng Yu, Zhou, and Wang (2020) implemented the Generative Adversarial Networks (GANs) to synthesize additional radiotherapy-like information and overcome the issue of datasets with limited sample sizes.

3 Methodology

The Knowledge Discovery in Databases (KDD) paradigm will be used in the research project's implementation. The goal of the KDD method is to extract information from massive data sets. This paradigm divides the execution of a data analysis project into four stages: data selection and extraction, data preparation, data transformation, data mining, and evaluation/interpretation. The in-depth summarise of the steps is explained in later section.

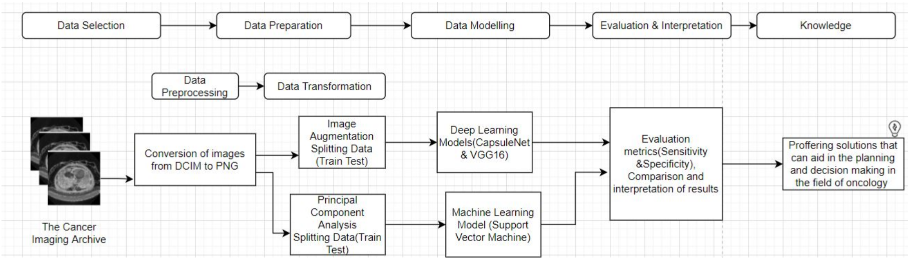

Figure 1: KDD Methodology

3.1 Data Selection

The 4D-Lung Image dataset used for this research project was published by Hugo et al. [2017] on The Cancer Imaging Archive (TCIA) website. The dataset consists of 1,699 images CT scans. The data consist of Computed Tomography (CT) scans images of chemoradiotherapy.

3.2 Data Preparation

This step comprises of data pre-processing and data transformation stage. the data is prepared for the upcoming phase of modelling. In the data pre-processing stage, the image is of DCIM (Digital Camera Images) format which makes it hard to work in a python environment, thus the DCIM images were manually converted to PNG format without losing the data. The pre-processing stage was done in two phases (Deep Learning phase and Machine Learning phase). The deep learning phase, hyper-parameter was set as random seed, image size, batch size, training steps, validation steps and epoch. The image folder containing two folders of the classes were combined before data shuffling and implementing image augmentation. Image augmentation, the images were re-scaled, rotation range, width shift range, height shift range and shear range, brightness, vertical, horizontal, fill mode was set, and image data generator created in other split data into train and valid generators. For the machine learning phase, the images were flattened by reshaping the images from 4D to 2D images. Image feature extract was done using principal component analysis (PCA) and the image maximum value was derived. The data transformation data was split into 80:20 ratio for training and test.

3.3 Data Modelling

The processed data will then be run through the classification models. The models adopted for the prediction of RP in NSCLC are VGG16, CapsuleNet and SVM.The image

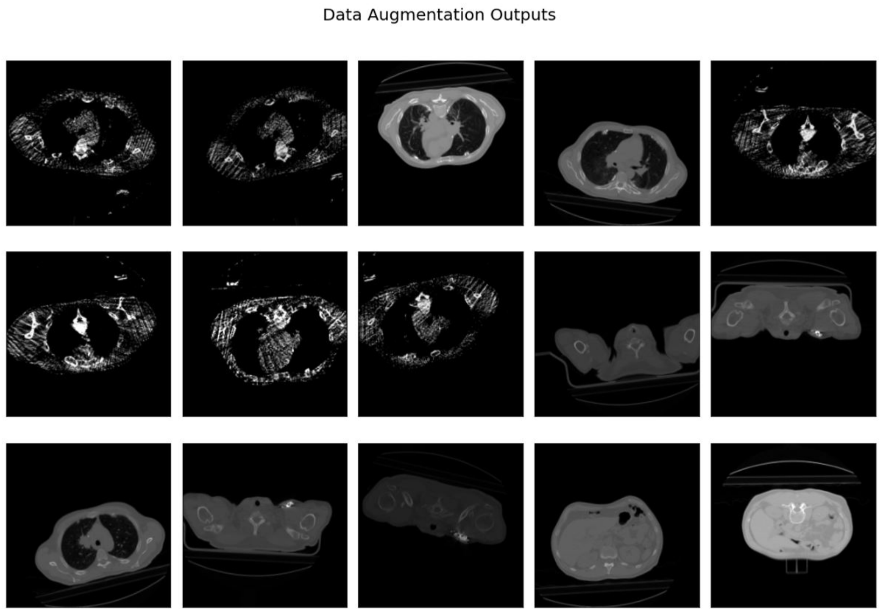

Figure 2: Image Augmentation

augmentation data is run through VGG16 and CapsuleNet and PCA decomposition was applied on SVM model, this is done to facilitate the model execution.

3.3.1 VGG16

VGG16 is a convolutional neural network model which has large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. The model has 16 deep layers which load the ImageNet data. The data images in the research needs to be uploaded as ImageNet. The model can classify up to 1000 classes. This model has won 1st and 2nd place in the categories of object localization and image classification respectively in the year 2014 at the ILSVRC challenge. The implementation and efficiency of VGG16 model will be explained in section 5.

3.3.2 Capsule Neural Networks (CapsuleNet)

Capsule Neural Networks (CapsNet) are networks that can retrieve spatial information and other critical aspects to overcome the data loss that occurs during pooling operations. Four main components are present in the CapsNet that are listed below[3]:

- Matrix Multiplication: It is used to transform the image that is supplied as an input to the network into vector values so that the spatial portion can be understood.
- Scalar Weighting of the Input: It figures out which higher level capsule should get the present capsule's output.
- Dynamic routing algorithm: It enables these many components to exchange data with one another. Lower-level capsules provide input to higher level capsules.

[3]https://analyticsindiamag.com/understanding-capsule-net-with-its-implementation-in-computer-vision/

- Squashing Function: It is the final component that summarizes the data. The squashing function takes all the data and turns it into a vector that is smaller than or equal to 1, while keeping the vector's orientation.

The implementation of the model will be further explained in section 5.

3.3.3 Support Vector Machine (SVM)

SVM is a classification model that divided the two groups by a hyperplane which makes the SVM identity a non-probabilistic binary classifier. The training data point which is closest to the nearest classifier is also known as Support Vector. To achieve a better separation between classes, support vector machines with a radial basis function (RBF) kernel modify the original feature space. The implementation of the model will be further explained in section 5. By increasing the margin between the two classes after translating the training data, x into a higher dimensional space using a mapping function $\phi(x)$. As a result, there is a decision-making function as shown below:

$$f(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle + b, \tag{1}$$

3.4 Performance Evaluation

After the models have been trained on train test sets of data, evaluation metrics (Specificity and Sensitivity) were adopted to evaluate the model's efficiency and to confirm hypothesis predictions. The evaluation metrics are further explained in Section 5.

4 Design Specification

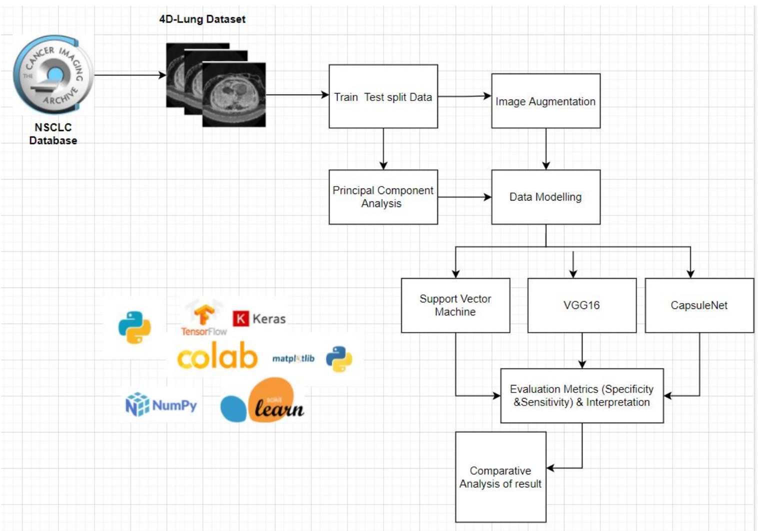

Figure 3: Process Flow Diagram

The process flow diagram in Figure 4 clearly explains the steps taken in answering the research question. Python programming was selected as the primary tool for development due to its ease of use and the availability of several libraries such as Keras, TensorFlow, OpenCv, Sklearn, and Matplotlib that can be used to quickly achieve the capabilities and create models shown in Figure 4. Pre-processing and data transformation was executed using Google Colab notebook, which functioned as the Integrated Development Environment (IDE). Following that, models like VGG16, SVM, and CapsuleNet were implemented. The collected findings are then depicted using the python matplotlib tools, allowing us to readily compare and evaluate the models implemented.

5 Implementation

This section discusses the system's overall implementation. This section also discusses how all the activities were completed to accurately predict the radiation pneumonitis in NSCLC patients. The initial step of this stage is environment setup, which provides information on the tools utilized in this research as well as the necessary setting. The data is then pre-processed and transformed before comprehensive information on the model design and operation is provided.

5.1 ENVIRONMENT SETUP

The research was carried out using a 64-bit Windows 10 operating system with 8GB of RAM, and Python was utilized as the programming language. The research is implemented in the Google Colab environment. It is a cloud platform that implements the python environment and has Jupyter notebook pre-setup so the data analysis can get started easily. For image preparation and execution of CapsuleNet, VGG16, and SVM models, this notebook uses Python 3.6.9 and libraries like OpenCV, Keras, TensorFlow, and Sklearn were employed. The environment also gives support to GPUs and TPUs which leads to faster development of the image processing and image classification algorithms. All data are saved on google drive and accessible in the notebook by mounting Google Drive with the aid of a python library. The environment has certain python libraries pre-installed, even though any required library can be installed easily if required.

5.2 Data Handling

From Google Drive, all CT scan images are read and loaded into the directories. Each of the directories were created belonging to pneumonitis and no pneumonitis classes. The data exploration was done by creating the path of each class directories and to check how many images are in each directory, data wrangling and visuals about the data was done. The radiation pneumonitis contains 1062 images, and no radiation pneumonitis images contains 637 images. The images were augmented with the cv2 python package, the Keras library's ImageDataGenerator class is adopted to offer real-time data augmentation. tf.keras.preprocessing.image.ImageDataGenerator() function in the ImageDataGenerator class accepts the loading of the data into train and test data generators with the right parameters.

5.3 Classification Techniques

5.3.1 Vision Geometry Group 16 (VGG16)

The initial model is the VGG16, which has been shown to be an effective and accurate Convolutional Neural Network (CNN) for image dataset classification [Liu and Deng, 2015]. The architecture was pretrained on the ImageNet. For each image, this component generated a tensor with 7x7x512 values. The first two dimensions (7x7) correspond to the original image dimensions. The network perception of the 4D-Lung image dataset is described by 512 features. A total of 25,088 values represents the number of features extracted by the network from the image. The Flatten layer follows, which transforms the input to one-dimensional space without changing the batch size. A dropout layer was added to sets input unit to 0 with a rate frequency preventing overfitting of the model. To accomplish the binary classification, the Softmax activation function has been implemented. The model was built using Adam optimiser, which has a learning rate of 0.001.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Functional)           (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
dropout (Dropout)            (None, 25088)             0
_________________________________________________________________
dense (Dense)                (None, 2)                 50178
=================================================================
Total params: 14,764,866
Trainable params: 50,178
Non-trainable params: 14,714,688
```

Figure 4: VGG16 model Architecture

Table 2: VGG16 Model Parameters

Parameter	**Value**
Batch Size	224, 224
Epoch	25
Optimiser	Adam
Learning Rate	0.001
Early Stopping Monitor	Validation Accuracy
Early Stopping Patience	8

5.3.2 Capsule Neural Network (CapsuleNet)

Capsules are equivariant networks of neurons that receive and output vectors [Kwabena Patrick et al., 2019]. The CapsuleNet architecture is made of two convolutional layers. The first layer Conv1 is designed with 256 channels with 9 x 9 filters, ReLU activation function and stride of one. The second convolutional layer known as the primary layer with 8 convolutional layer unit, kernel of 9 x 9 and Squash activation function. The third layer known as Digitcaps which consist of a fully connected (FC) layer with ten 16D capsules that receive input from all capsules in the layer below to conduct classification based on two classifications. The last layer, known as Decoder, is critical in identifying the real length of each capsule in the preceding layer, which is required to determine the

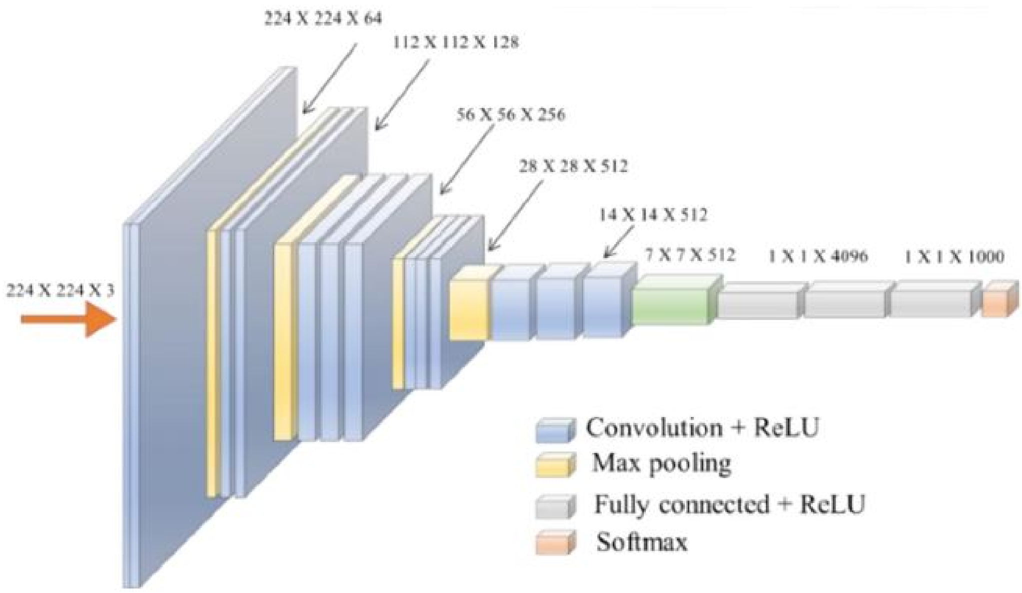

Figure 5: VGG16 Architecture Pandiyan et al. [2019]

possibility of radiation pneumonitis in 4D-lung CT scans. Adam optimiser learning rate of 0.001.

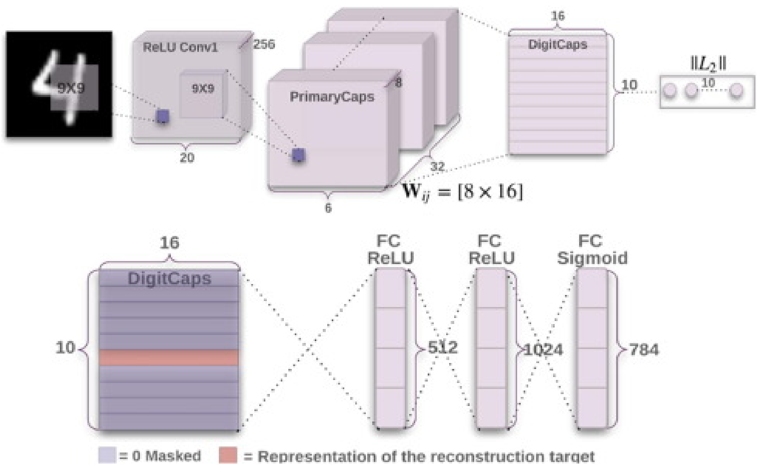

Figure 6: CapsuleNet Architecture [Kwabena Patrick et al., 2019]

Table 3: CapsuleNet Model Parameters

Parameter	**Value**
Batch Size	224, 224
Epoch	25
Optimiser	Adam
Learning Rate	0.001
Early Stopping Monitor	Validation Accuracy 1
Early Stopping Patience	8

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 224, 224, 64)      1792
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 224, 224, 128)     73856
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 112, 112, 128)     0
_________________________________________________________________
flatten_1 (Flatten)          (None, 1605632)           0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 3211266
=================================================================
Total params: 3,286,914
Trainable params: 3,286,914
Non-trainable params: 0
```

Figure 7: CapsuleNet Model Architecture

5.3.3 Support Vector Machine (SVM)

SVM is a machine learning model that uses the SVM python package sklearn to classify radiation pneumonitis in NSCLC CT scans. It features numerous hyper-tuneable parameters such as Radial Basis Function (RBF) kernel, gamma, and C (regularisation parameter) for better performance outcome. The svm.SVC () function from the sklearn package was used to run the SVM classification model. The clf.predict() function was used in predicting RP in NSCLC from the classifier.

6 Evaluation and Results

This section of the study looks at the models and all of the parameters that were fine-tuned to get the best outcome. For this study project on 4D - NSCLC CT scan images, three models were adopted and in order to make the study novel models like VGG16 and CapsuleNet are implemented and outcome is compared with that of SVM. The training and validation accuracies, as well as the losses determined for each epoch, are first used to assess the models performance and efficiency. This may be shown in graphs, which shows how the accuracy and loss of the training and validation data change over time with each epoch. In addition, Sensitivity and specificity metrics are utilized in order to assess the model prediction efficiency.

6.1 Experiment 1: VGG16 model built on ImageNet

This model was created with the help of a pre-trained model (ImageNet) and without the top layer. The model was run for 25 epochs, however it was halted early after 14 epochs since the validation accuracy was no longer rising. It yielded a 98.64 percent accuracy and a 98.64 percent validation accuracy. The model variance was tested on 5 folds cross validation. The model accuracy and loss plots are shown to show the variance between the training and validation datasets. The model is further tested using the validation_generator and predict_ generator functions to assess how well it performs on unknown test data. In addition, the model sensitivity and specificity metrics generated an outcome of 100 percent and 95percent respectively on the third fold cross validation. The result shows a highly sensitivity outcome which means that are no false positive result, while the specificity result signifies there are fewer sensitivity outcome. The model took 1400 seconds to train on the training dataset.

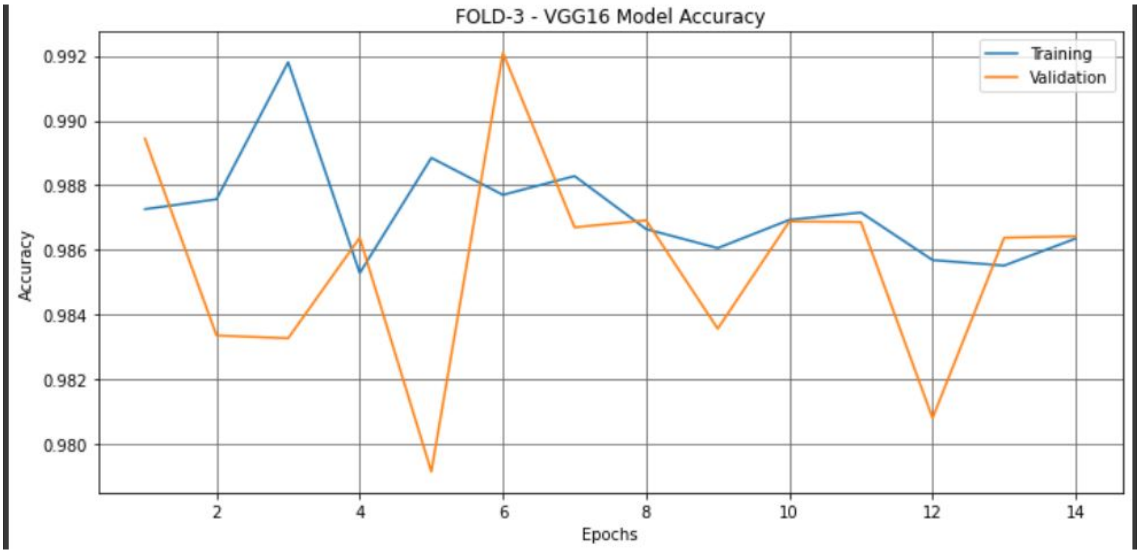

Figure 8: VGG16 Model Accuracy

Figure 9: VGG16 Model Loss

6.2 Experiment 2 : Capsule Neural Network (CapsuleNet) built from scratch

The CapsuleNet is built from scratch with 2 convolutional 2D layers. The first convolutional 2D layer designed with 256 channels and 9 x 9 filter size. The model was executed with 25 epochs, however it halted early after 14 epochs since the validation accuracy was no longer rising and resulted in training accuracy of 98.48% and validation accuracy of 96.59%. The model variance was tested on 5 folds cross validation. The accuracy and loss plots for the model are plotted to see the variations for training and validation dataset.

The model is evaluated using validation_generator function to check the model performance on unseen validation data. The evaluation metrics shows sensitivity and specificity outcome of 98.15% and 90.52% on the third fold cross validation. The sensitivity outcome depicts that there are 98% of NSCLC patients with radiation pneumonitis, while the specificity result shows there are 90% of NSCLC patients without the RP.The model took 1300 seconds to train on the training dataset.

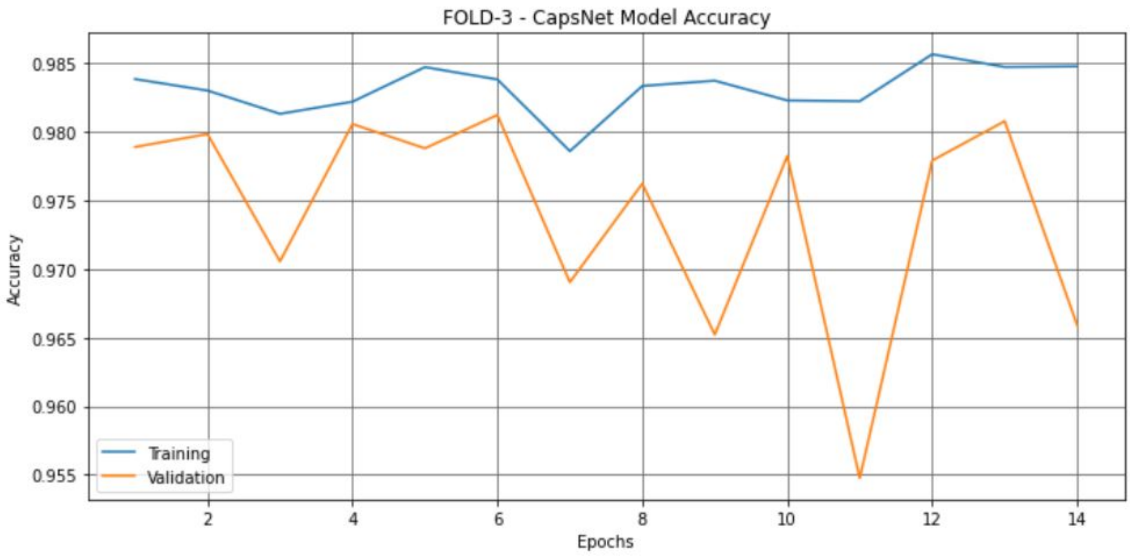

Figure 10: CapsuleNet Model Accuracy

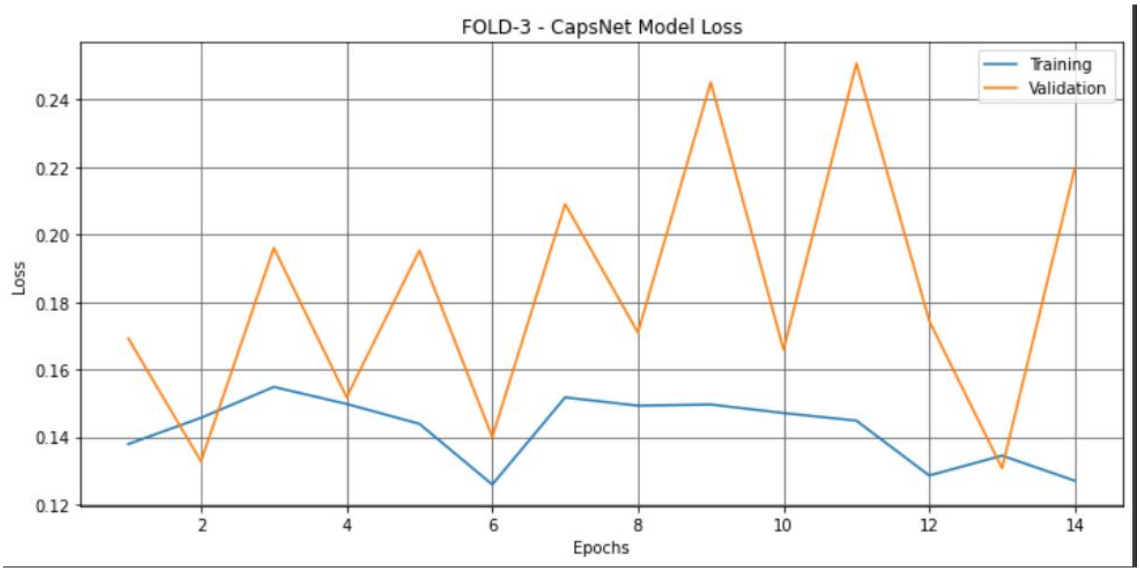

Figure 11: CapsuleNet Model Loss

6.3 Experiment 3: Support Vector Machine

The last model executed in this project report is SVM. It has been recommended because SVM model has proven to produce a favourable outcome in the prediction of radiation pneumonitis in NSCLC CT scans [Oh et al., 2009]. The data points are classified into classes in this model in order to obtain a maximum marginal hyperplane that acutely classifies the data points. The Sklearn library was imported to apply SVM. SVM model was split into 80:20 train test dataset. The model variance was tested on 5 folds cross validation, the evaluation metrics showed sensitivity and specificity results of 99.18% and 96.79%, which signifies that there are 99% of NSCLC patients with radiation-induced lung injury while specificity denotes that there 96% of NSCLC patients without radiation-induced lung injury. The model took less than 200 seconds to train on the training dataset.

6.4 Experiment 4 : Execution Time of Models

Following the completion of the various implementation of models, a critical analysis of the training times required by the various models was conducted. The execution time is the time it takes for a function to execute and is expressed in seconds. When the computational time is extremely long, Graphics Processing Unit (GPU) support is critical since it drastically decreases the execution time. The computing time of several models is compared, and the SVM model is shown to be the quickest of all the implemented techniques with less than 200 seconds of training time.

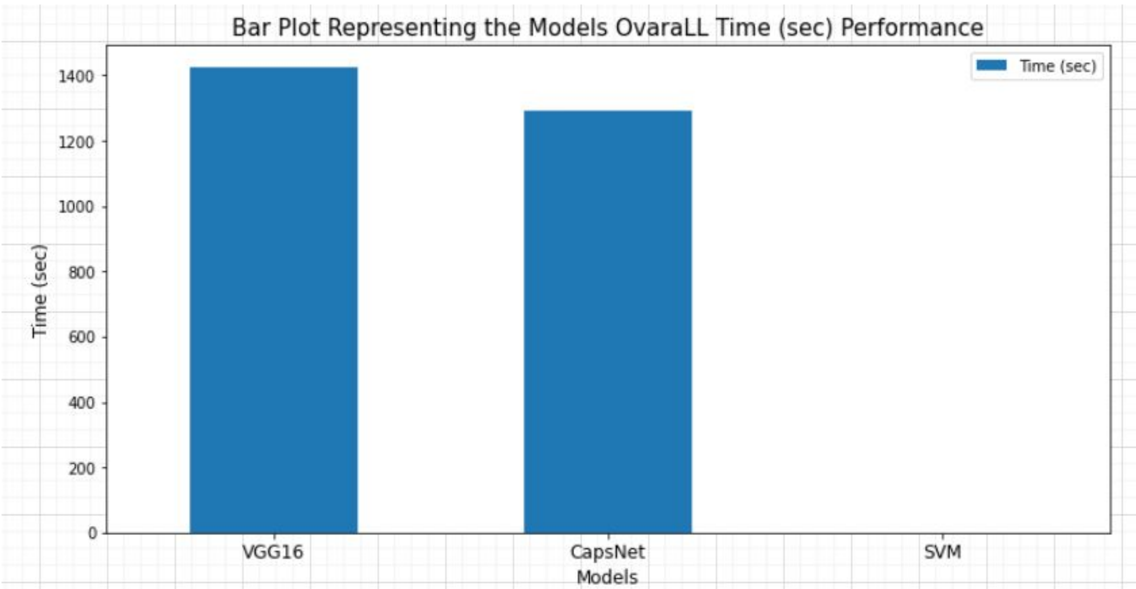

Figure 12: Execution Time Comparison

6.5 Discussion

The main goal of the research project was to carry out a comparative analysis between deep learning and machine learning algorithm based on Sensitivity and Specificity, K-fold cross validation and execution time. The image pre-processing and augmentation process was very crucial to the performance of all the models implemented. The experiments commenced with the application of model based on evaluation metrics. The VGG16 model attained Sensitivity and Specificity of 100% and 95% respectively under the 3 fold cross validation. However, the execution time of VGG16 was very high, but it is understandable as cross validation was applied on the training data. The model was trained on 25 epochs but halted after 14 epochs. The next experiment applied was CapsuleNet built from the scratch to improve model performance by using hyper-parameter tuning on 25 epochs but halted after 14 epochs. The sensitivity and specificity outcome achieved was 98.15% and 90.52% respectively under the thrid fold cross validation, the execution time was lesser compared to that of VGG16 model. The last model applied was SVM, its performance was measured also by sensitivity and specificity with an outcome of 96.03% and 97.20% respectively. The execution time for the model to be trained was less than 200 seconds. The reason for a very execution time may be as result of the processes involved in image classification with a machine learning model. The K-fold cross validation of the model is shown below.

Table 4: Comparison based on K-fold cross validation

K-Folds	**Model**	**Sensitivity**	**Specificity**
1	VGG16	98.33%	95.19%
	CapsuleNet	92.37%	97.13%
	SVM	96.15%	99.05%
2	VGG16	96.71%	96.76%
	CapsuleNet	96.64%	94.57%
	SVM	88.06%	95.63%
3	VGG16	100%	95%
	CapsuleNet	98.15%	90.52%
	SVM	99.18%	96.79%
4	VGG16	99.18%	96.79%
	CapsuleNet	93.13%	97.13%
	SVM	97.60%	97.21%
5	VGG16	95.35%	98.10%
	CapsuleNet	91.11%	98.04%
	SVM	96.00%	96.73%

The Bar plot representing the models performance score is illustrated in figure 12 below.

However, all models performed exceedingly well in terms of sensitivity and specificity in predicting RP in NSCLC patients CT scans but it is worth noting that VGG16 achieved overall best result compared to other models. In terms of execution time to train the models, SVM outperformed the deep learning techniques.

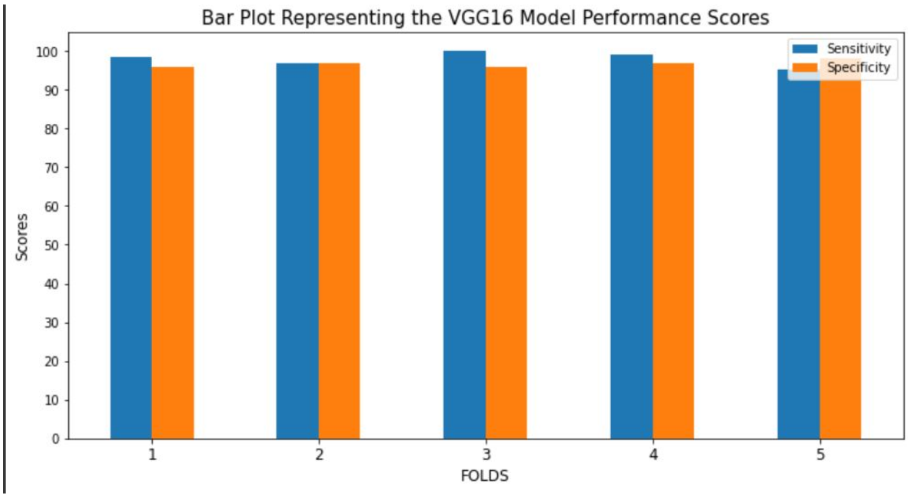

Figure 13: Bar plot representing VGG16 performance Scores

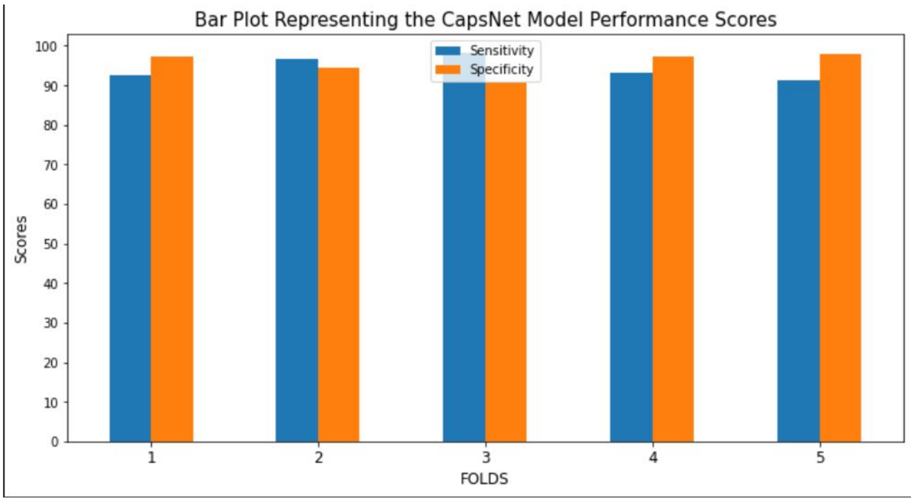

Figure 14: Bar plot representing CapsNet performance Scores

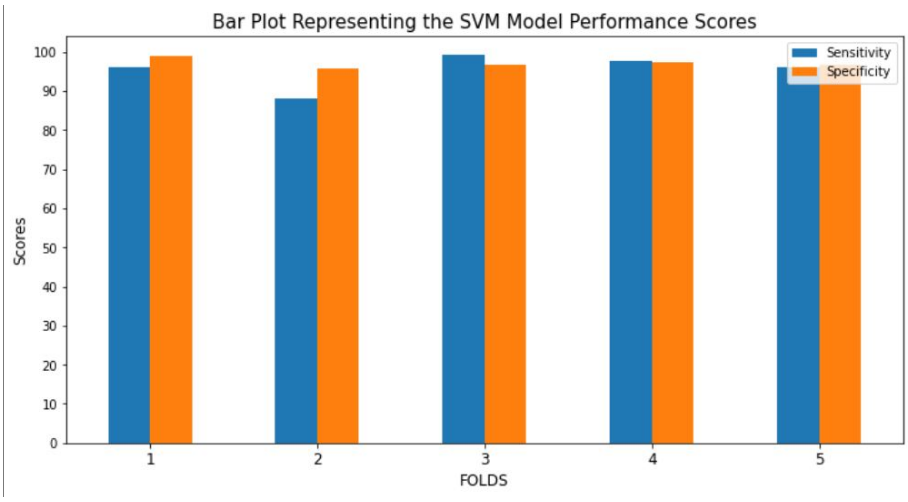

Figure 15: Bar plot representing SVM performance Scores

7 Conclusion and Future Work

Radiation pneumonitis prediction in NSCLC has been a persisting issue that oncologists have attempted to address. However, a significant amount of medical research has gone into establishing a reliable predictive machine learning model for categorizing NSCLC radiation CT images. Different sorts of models based on traditional machine learning and deep learning approaches are implemented in this study. The models and architectures are chosen based on knowledge about their performance and execution time gleaned from a thorough literature review. The dataset selection was first subjected to an exploratory data analysis. After confirming that the data was valid, image pre-processing, image augmentation, principal component analysis procedures were employed to transform and normalise the image data. SVM using the Sklearn python package, CapsuleNet built from scratch, and VGG16 with pre-trained weights were the three models executed. The models were then trained on 1699 NSCLC radiotherapy CT scan images from the Cancer Imaging Archive database. The performance of the models was compared using the 5-fold cross-validation approach. The deep learning model VGG16 had the greatest performance in the 3-fold cross validation, with a sensitivity of 100% and a specificity of 95%, while the machine learning model SVM had the fastest computational time in training the model. While the models achieved the study aim, there were several constraints in identifying the right hyperparameters to train the CapsuleNet model, as well as the Google Colab restricted memory space which made the GPU runtime slower.
For future work, it is recommended that a high performing system be used with Tensor Processing Unit (TPU)/ Graphics processing Unit (GPU) for improving the execution time for the models. This is required due to the use of images, which takes a lot of time to process. Also, the images were of DCIM format which needed to be converted to PNG file due to the limitation of the python DCIM file library. During the conversion some data is usually lost, for further work, different images formats and libraries could be used. In addition, the aim is to indulge in further study of novel deep learning models like Deep Belief Network and Convolutional extreme Gradient Boosting on huge dataset to add to the body of knowledge in the field of oncology.

8 Acknowledgment

Mr. Noel Cosgrave, my supervisor deserves my deepest appreciation for his invaluable advice, supervision, and regular input during the project. I will also want to express my gratitude to my fiance and friends for their unending support.

References

Enoch Chang, Justin Du, Roy Decker, James Yu, Henry Park, and Sanjay Aneja. Development and validation of a deep learning algorithm to predict radiation pneumonitis and esophagitis following lung stereotactic body radiation therapy. *International Journal of Radiation Oncology*Biology*Physics*, 108:E14, 10 2020. doi:10.1016/j.ijrobp.2020.02.496.

Hann-Hsiang Chao, Gilmer Valdes, Jose Luna, Marina Heskel, Abigail Berman, Timothy Solberg, and Charles Simone. Exploratory analysis using machine learning to predict for chest wall pain in patients with stage i non-small-cell lung cancer treated with stereotactic body radiation therapy. *Journal of Applied Clinical Medical Physics*, 19, 07 2018. doi:10.1002/acm2.12415.

Lehui Du, Baolin Qu, Fang Liu, Na Ma, Shouping Xu, Wei Yu, Xiangkun Dai, and Xiang Huang. Precise prediction of the radiation pneumonitis with rpi: An explorative preliminary mathematical model using genotype information. *Journal of Clinical Oncology*, 37(15_suppl):e14569–e14569, 2019. doi:10.1200/JCO.2019.37.15_suppl.e14569. URL https://doi.org/10.1200/JCO.2019.37.15_suppl.e14569.

Lorena Giuranno, Jonathan Ient, Dirk De Ruysscher, and Marc A. Vooijs. Radiation-induced lung injury (rili). *Frontiers in Oncology*, 9:877, 2019. ISSN 2234-943X. doi:10.3389/fonc.2019.00877. URL https://www.frontiersin.org/article/10.3389/fonc.2019.00877.

Peng Huang, Hui Yan, Zhihui Hu, Zhiqiang Liu, Yuan Tian, and Jianrong Dai. Predicting radiation pneumonitis with fuzzy clustering neural network using 4dct ventilation image based dosimetric parameters. *Quantitative Imaging in Medicine and Surgery*, 0(0), 2021. ISSN 2223-4306. URL https://qims.amegroups.com/article/view/67939.

G D Hugo, Weiss Elisabeth, S Balik, W C Sleeman, P J Keall, J Lu, and J F Williamson. A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer, Feb 2017. URL https://pubmed.ncbi.nlm.nih.gov/27991677/.

Jue Jiang, Yu-Chi Hu, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann, Joseph O. Deasy, Gig Mageras, and Harini Veeraraghavan. Multiple resolution residually connected feature streams for automatic lung tumor segmentation from ct images. *IEEE Transactions on Medical Imaging*, 38(1):134–144, 2019. doi:10.1109/TMI.2018.2857800.

Andreas Kleppe, Ole-Johan Skrede, Sepp De Raedt, Knut Liestøl, David J Kerr, and Håvard E Danielsen. Designing deep learning studies in cancer diagnostics. *Nature reviews. Cancer*, 21(3):199—211, March 2021. ISSN 1474-175X. doi:10.1038/s41568-020-00327-9. URL https://doi.org/10.1038/s41568-020-00327-9.

Shane P. Krafft, Arvind Rao, Francesco Stingo, Tina Marie Briere, Laurence E. Court, Zhongxing Liao, and Mary K. Martel. Erratum: The utility of quantitative ct radiomics features for improved prediction of radiation pneumonitis (medical physics, (2018), 45, 11, (5317-5324), 10.1002/mp.13150). *Medical Physics*, 46(2):1079, feb 2019. ISSN 0094-2405. doi:10.1002/mp.13327. Funding Information: SPK was funded, in part, through a fellowship from Cancer Prevention Research Institute of Texas Research Training Program RP170067. Funding Information: The authors would like to add the following statement to the Acknowledgements section in the published article: SPK was funded, in part, through a fellowship from Cancer Prevention Research Institute of Texas Research Training Program RP170067. Publisher Copyright: © 2019 American Association of Physicists in Medicine.

Mensah Kwabena Patrick, Adebayo Felix Adekoya, Ayidzoe Abra Mighty, and Baagyire Y. Edward. Capsule networks – a survey. *Journal of King Saud University - Computer and Information Sciences*, 2019. ISSN 1319-1578.

doi:https://doi.org/10.1016/j.jksuci.2019.09.014. URL https://www.sciencedirect.com/science/article/pii/S1319157819309322.

Bin Liang, Yuan Tian, Xinyuan Chen, Hui Yan, Lingling Yan, Tao Zhang, Zongmei Zhou, Lvhua Wang, and Jianrong Dai. Prediction of radiation pneumonitis with dose distribution: A convolutional neural network (cnn) based model. *Frontiers in Oncology*, 9:1500, 2020. ISSN 2234-943X. doi:10.3389/fonc.2019.01500. URL https://www.frontiersin.org/article/10.3389/fonc.2019.01500.

Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734, 2015. doi:10.1109/ACPR.2015.7486599.

Jose Luna, Hann-Hsiang Chao, Eric Diffenderfer, Gilmer Valdes, Chidambaram Chinniah, Grace Ma, Keith Cengel, Timothy Solberg, Abigail Berman, and Charles Simone. Predicting radiation pneumonitis in locally advanced stage ii–iii non-small cell lung cancer using machine learning. *Radiotherapy and Oncology*, 133:106–112, 04 2019. doi:10.1016/j.radonc.2019.01.003.

Yi Luo, Issam El Naqa, Daniel L. McShan, Dipankar Ray, Ines Lohse, Martha M. Matuszak, Dawn Owen, Shruti Jolly, Theodore S. Lawrence, Feng-Ming (Spring) Kong, and Randall K. Ten Haken. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via bayesian network analysis. *Radiotherapy and Oncology*, 123(1):85–92, 2017. ISSN 0167-8140. doi:https://doi.org/10.1016/j.radonc.2017.02.004. URL https://www.sciencedirect.com/science/article/pii/S0167814017300634.

Angel Moran, Megan E. Daly, Stephen S.f. Yip, and Tokihiro Yamamoto. Radiomics-based assessment of radiation-induced lung injury after stereotactic body radiotherapy. *Clinical Lung Cancer*, 18(6), 2017. doi:10.1016/j.cllc.2017.05.014. URL https://www.sciencedirect.com/science/article/pii/S1525730417301493.

Jung Hun Oh, Rawan Al-Lozi, and Issam El Naqa. Application of machine learning techniques for prediction of radiation pneumonitis in lung cancer patients. In *2009 International Conference on Machine Learning and Applications*, pages 478–483, 2009. doi:10.1109/ICMLA.2009.118.

Vigneashwara Pandiyan, Tegoeh Tjahjowidodo, Wahyu Caesarendra, and Pushparaja Murugan. In-process virtual verification of weld seam removal in robotic abrasive belt grinding process using deep learning. *Robotics and Computer-Integrated Manufacturing*, 57:477–487, 01 2019. doi:10.1016/j.rcim.2019.01.006.

Tianle Shen, Liming Sheng, Ying Chen, Lei Cheng, and Xianghui Du. High incidence of radiation pneumonitis in lung cancer patients with chronic silicosis treated with radiotherapy. *Journal of radiation research*, 61, 12 2019. doi:10.1093/jrr/rrz084.

Gilmer Valdes, Timothy D Solberg, Marina Heskel, Lyle Ungar, and Charles B Simone. Using machine learning to predict radiation pneumonitis in patients with stage i non-small cell lung cancer treated with stereotactic body radiation therapy. *Physics in Medicine and Biology*, 61(16):6105–6120, jul 2016. doi:10.1088/0031-9155/61/16/6105. URL https://doi.org/10.1088/0031-9155/61/16/6105.

Michael Velec, Carol R. Haddad, Tim Craig, Lisa Wang, Patricia Lindsay, James Brierley, Anthony Brade, Jolie Ringash, Rebecca Wong, John Kim, and Laura A. Dawson. Predictors of liver toxicity following stereotactic body radiation therapy for hepatocellular carcinoma. *International Journal of Radiation Oncology*Biology*Physics*, 97(5):939–946, 2017. ISSN 0360-3016. doi:https://doi.org/10.1016/j.ijrobp.2017.01.221. URL https://www.sciencedirect.com/science/article/pii/S0360301617302845.

Shuo Wang, Mu Zhou, Zaiyi Liu, Zhenyu Liu, Dongsheng Gu, Yali Zang, Di Dong, Olivier Gevaert, and Jie Tian. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Medical Image Analysis*, 40:172–183, 2017. ISSN 1361-8415. doi:https://doi.org/10.1016/j.media.2017.06.014. URL https://www.sciencedirect.com/science/article/pii/S1361841517301019.

Yiwen Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R. Mak, and H. Aerts. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clinical Cancer Research*, 25:3266 – 3275, 2019.

Melek Yakar, Durmus Etiz, Muzaffer Metintas, Guntulu Ak, and Ozer Celik. Prediction of radiation pneumonitis with machine learning in stage iii lung cancer: A pilot study. *Technology in Cancer Research & Treatment*, 20:15330338211016373, 2021. doi:10.1177/15330338211016373. URL https://doi.org/10.1177/15330338211016373. PMID: 33969761.

Mengmeng Yan and Weidong Wang. Radiomic analysis of ct predicts tumor response in human lung cancer with radiotherapy. *Journal of Digital Imaging*, 33:1–3, 10 2020. doi:10.1007/s10278-020-00385-3.

Hao Yu, Huanmei Wu, Weili Wang, Shruti Jolly, Jian-Yue Jin, Chen Hu, and Feng-Ming (Spring) Kong. Machine learning to build and validate a model for radiation pneumonitis prediction in patients with non–small cell lung cancer. *Clinical Cancer Research*, 25(14):4343–4350, 2019. ISSN 1078-0432. doi:10.1158/1078-0432.CCR-18-1084. URL https://clincancerres.aacrjournals.org/content/25/14/4343.

Hao Yu, Ka-On Lam, Huanmei Wu, Michael Green, Weili Wang, Jian-Yue Jin, Chen Hu, Shruti Jolly, Yang Wang, and Feng-Ming Spring Kong. Weighted-support vector machine learning classifier of circulating cytokine biomarkers to predict radiation-induced lung fibrosis in non-small-cell lung cancer patients. *Frontiers in Oncology*, 10:3191, 2021. ISSN 2234-943X. doi:10.3389/fonc.2020.601979. URL https://www.frontiersin.org/article/10.3389/fonc.2020.601979.