Prediction and Survival Analysis of Head and Neck Cancer in Patients using Epigenomics Data and Advanced Machine Learning Methods

By

Vikaskumar Chaudhary

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science (MSc) in Computational Sciences

The Office of Graduate Studies

Laurentian University

Sudbury, Ontario, Canada

© Vikaskumar Chaudhary, 2023

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Office of Graduate Studies/Bureau des études supérieures

Title of Thesis Titre de la thèse	Prediction and Survival Analysis of Head and Neck Cancer in Patients using Epigenomics Data and Advanced Machine Learning Methods
Name of Candidate Nom du candidat	Chaudhary, Vikaskumar
Degree Diplôme	Master of Science
Department/Program Département/Programme	Computational Sciences
Date of Defence Date de la soutenance	August 22, 2023

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Kalpdrum Passi
(Supervisor/Directeur(trice) de thèse)

Dr. Ratvinder Grewal
(Committee member/Membre du comité)

Dr. Azam Bidgoli
(Committee member/Membre du comité)

Dr. P. Maulika Patel
(External Examiner/Examinateur externe)

Approved for the Office of Graduate Studies
Approuvé pour le Bureau des études supérieures
Tammy Eger, PhD
Vice-President Research (Office of Graduate Studies)
Vice-rectrice à la recherche (Bureau des études supérieures)
Laurentian University / Université Laurentienne

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

I, **Vikaskumar Chaudhary**, hereby grant to Laurentian University and/or its agents the non-exclusive license to archive and make accessible my thesis, dissertation, or project report in whole or in part in all forms of media, now or for the duration of my copyright ownership. I retain all other ownership rights to the copyright of the thesis, dissertation or project report. I also reserve the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that this copy is being made available in this form by the authority of the copyright owner solely for the purpose of private study and research and may not be copied or reproduced except as permitted by the copyright laws without written authority from the copyright owner.

Abstract

Epigenomics is the field of biology dealing with modifications of the phenotype that do not cause any alteration in the sequence of cell DNA. Epigenomics adds something to the top of DNA to change the properties, which eventually prohibits certain DNA behavior from being performed. Such modifications occur in cancer cells and are the sole cause of cancer. The main objective of this research is to perform prediction and survival analysis of Head and Neck Squamous Cell Carcinoma (HNSCC) which is one of the biggest reasons of death and accounts for more than 650,000 cases and 330,000 deaths annually worldwide. Tobacco use, alcohol consumption, Human Papillomavirus (HPV) infection (for oropharyngeal cancer), and Epstein-Barr Virus (EBV) infection are the main risk factors associated with head and neck cancer (for nasopharyngeal cancer). Males, with a proportion ranging from 2:1 to 4:1, are slightly more affected than females. Four different types of data are used in this research to predict HNSCC in patients. The data includes methylation, histone, human genome and RNA-Sequences. The data is accessed through open-source technologies in R and Python programming languages. The data is processed to create features and with the help of statistical analysis and advanced machine learning techniques, the prediction of HNSCC is obtained from the fine-tuned model. The optimal model was determined to be ResNet50 utilizing the Sobel feature selection method for image data and ReliefF-based feature selection for clinical features, achieving a test accuracy of 97.9%. The model's precision score was 0.929, its recall score was 0.930, and its F1 score was 0.930. Additionally, the ResNet101 model demonstrated the best performance using the Histogram of Gradients feature selection method for image data and mutual information-based feature selection for clinical features, yielding a test accuracy of 96.1%. Its precision score, recall score, and F1 score were identical to the aforementioned ResNet50 model. The research also utilized Kaplan-

Meier survival analysis to investigate the survival rates of patients based on various factors, including age, gender, smoking status, tumor size, and location of site. The results obtained from this analysis yielded the effectiveness of the method in providing valuable insights for risk assessment.

Keywords: Epigenomics, Histone, DNA methylation, Human Genome, RNA

Acknowledgments

I would like to express my deepest gratitude and appreciation to my supervisor, Prof. Kalpdrum Passi, who gave me the opportunity and all the possible support to make this research successful. I believe I have got the finest supervisor who has been standing with me throughout the educational journey and he was always available to clear all my doubts and handled all the queries with patience.

I want to thank my technical coordinator, for providing all the server requirements for training phase. He was always available and supported all the dependencies required in the environments.

I also want to thank my parents my brother and all my friends who made my research experience enjoyable. I want to give a shout out to all my classmates for their help and support for the testing phase.

Finally, and most significantly, I offer special thanks to my parents for support and patience and believing in me and trusting my caliber.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Cancer

Cancer is a disease where a set of undesirable cells are formed due to abnormal cell division. Normal cells are responsive to the signal of cell division or cell death. However, changes to the cell's genome take place in cancerous cells. It interferes with how cells normally divide and die. Cancer is brought on by the unchecked cell division that occurs from it. Once malignant cells have developed, they can pose a health risk due to metastasis. Cancer is a multigene, multistep condition that starts with a single aberrant cell. As a result of modifications to the DNA sequence, the cell steadily advances to the stage of moderate aberration. The creation of a mass of aberrant cells is brought on by natural selection and mutations and a few tumor cells have undergone mutations, which result in the development of cancerous cells that metastasize [1].

1.1.1 Cancer risk factors

There are various cancer risk factors which are discussed below.

- Only 5 to 10 percent of malignancies are brought on by inherited gene mutations while the remaining 90 to 95 percent of cancer cases are caused by environmental factors and a variety of lifestyle choices, among others.
- Environmental contaminants, tobacco use, alcoholism, diet (over consumption of fast food and red meat), obesity, inactivity, sun exposure etc. are other risk factors.

1.2. Head and Neck cancer

Head and Neck Cancer, also known as HNSCC (Head and neck squamous cell carcinoma), is a growing problem worldwide, especially in Europe. It affects the mouth, lips, and tongue. Oropharyngeal and oral cancers are ranked as the sixth most prevalent cancers worldwide. Oral cancer is the leading cause of death worldwide, and since there are no distinct early signs, the early-stage lesions that manifest are not often recognized as malignant lesions. The likelihood of the patient surviving is increased, and extensive surgery can be avoided if it is discovered at a young age and treated promptly. Head and neck squamous cell carcinoma (HNSCC) is a widespread heterogeneous malignancy that accounts for 500,000 new cases globally per year and involves cancers of the oral cavity, oropharynx, nasopharynx, hypopharynx, and larynx [1,2]. HNSCC traditional treatment involves surgery, radiotherapy and chemotherapy, used alone or in combination, depending on the level of the tumor and the main site [3]. Clinical treatment responses differ greatly among patients with HNSCC and remain disappointing, especially in advanced stage disease [4]. Chemotherapy and radiation, in comparison, also confer major toxicity [3]. The most prevalent cancers of the head and neck are head and neck squamous cell carcinomas (HNSCCs), which emerge from the mucosal epithelium of the oral cavity, throat, and larynx. The prevalence of HNSCC varies between nations/regions and has often been linked to either excessive alcohol intake, exposure to carcinogens produced from cigarettes, or both. Infection with oncogenic strains of the human papillomavirus (HPV), especially HPV-16 and, to a lesser extent, HPV-18 and other strains is becoming more and more associated with the development of oropharyngeal tumors. It is possible that HPV-positive HNSCC might be avoided by effective vaccination campaigns across the world because the two most prevalent oncogenic HPVs, HPV-16 and HPV-18, are covered by FDA-approved HPV vaccines. Smoking is still the main risk factor

for HNSCCs of the oral cavity and larynx, which are now referred to as HPV-negative HNSCCs. Careful physical examination continues to be the key method for early detection because no screening tool has demonstrated to be useful. The majority of patients with advanced-stage HNSCC come without a clinical history of a premalignancy, despite the fact that certain oral premalignant lesions, which first manifest as leukoplakia (white patches) or erythroplakia (red patches), proceed to invasive cancer. Depending on the disease stage, HNSCC of the oral cavity is often treated with surgical resection, followed by adjuvant radiotherapy or chemotherapy with radiation (also known as chemoradiation or CRT). Cancers that develop in the pharynx or larynx are often treated with CRT. A current study is evaluating the effectiveness of therapeutic dose reduction (of radiation and chemotherapy) in the treatment of HPV-positive illness. HPV-positive HNSCC often has a better prognosis than HPV-negative HNSCC. With the exception of early-stage oral cavity cancers, which may be treated with surgery alone, and larynx cancers, which can be treated with radiation or surgery alone, the bulk of HNSCC patients require multimodality treatments and, consequently, multidisciplinary care. The FDA has authorised the monoclonal anti-epidermal growth factor receptor (EGFR; also known as HER1) antibody cetuximab as a radiation sensitizer for recurrent or metastatic illness, either alone or in conjunction with chemotherapy. Cetuximab is frequently used in individuals who are ineligible for cisplatin, despite being less effective than cisplatin as a radio sensitizer in HPV-associated illness. Pembrolizumab is licensed as first-line therapy for patients who present with metastatic cancer, and the immune checkpoint inhibitors nivolumab and pembrolizumab are approved for the treatment of cisplatin-refractory recurrent or metastatic HNSCC. Comprehensive molecular characterization and immunological profile of HNSCC point to the potential for overcoming barriers to targeted treatments and extending longevity through the integration of prognostic and predictive

biomarkers into therapeutic care. The kinds of HNSCC, their epidemiology, and the pathophysiology of each type, as well as how this affects the management strategy, are all covered in this primer.

1.2.1. Epidemiology

Epidemiology is the study of how diseases are distributed and spread within populations, and how different factors can influence this distribution. It is a critical field in public health and plays a crucial role in understanding the causes and outcomes of various health issues. Epidemiologists use a range of methods to investigate the distribution of diseases, including descriptive epidemiology, which examines patterns of disease occurrence, and analytical epidemiology, which seeks to identify risk factors that contribute to the development of diseases. They also use statistical methods to analyze data and draw conclusions about the relationships between different factors and disease outcomes.

One important aspect of epidemiology is the identification of outbreaks and pandemics. Epidemiologists work to identify the causes of outbreaks, including the transmission modes, and develop strategies to prevent further spread of the disease. They also work to monitor the prevalence of diseases within communities and track changes over time.

In addition to infectious diseases, epidemiologists also study chronic diseases such as heart disease, cancer, and diabetes. These diseases often have complex causes and require careful analysis to identify risk factors and develop effective prevention and treatment strategies.

Overall, epidemiology plays a vital role in understanding and addressing public health issues. By identifying the causes of diseases and how they are spread, epidemiologists can help

develop strategies to prevent and treat these diseases, improving the health and wellbeing of populations around the world [2].

1.2.2. Risk factors of oral cancer

Many factors, such as the age, stage of the tumor, behavioral and cultural factors, exposure to the risks, approach to medical care services, occupation of the patient, time period between the disease diagnosis, prognosis and the start of treatment, types of treatment etc. all have influence on the survival rate of oral cancer.

1.3. Epigenomics and cancer

Epigenomics is the study of how changes in gene expression are controlled by modifications to the structure of DNA and its associated proteins, known as histones [3]. These modifications can be influenced by environmental factors such as diet and exposure to toxins, as well as genetic factors.

In recent years, epigenomics has emerged as a promising field in cancer research. Researchers have found that changes in DNA methylation and histone modifications can contribute to the development of cancer by altering gene expression patterns. These changes can also be used as biomarkers for early cancer detection and to predict the likelihood of cancer recurrence.

By analyzing epigenetic changes in cancer cells, researchers can identify specific patterns that are associated with different types of cancer. For example, some cancers have a characteristic pattern of DNA methylation that can be used to distinguish them from normal cells. This information can be used to develop new diagnostic tests and treatments that are tailored to the specific epigenetic changes associated with each type of cancer.

In addition to cancer diagnosis and treatment, epigenomics also has the potential to improve cancer prevention. By identifying the epigenetic changes that are associated with increased cancer risk, researchers can develop new strategies for cancer prevention, such as lifestyle modifications or targeted drug therapies.

Overall, epigenomics is a rapidly evolving field with enormous potential for improving cancer diagnosis, treatment, and prevention. As our understanding of the epigenetic mechanisms that underlie cancer continues to expand, we can expect to see new advances in cancer research and treatment that will benefit patients around the world [3].

1.4 Background

The most frequent cancers of the head and neck are head and neck squamous cell carcinomas (HNSCCs), which emerge from the mucosal epithelium of the oral cavity, throat, and larynx. Infection with oncogenic strains of the human papillomavirus (HPV), especially HPV-16 and, to a lesser extent, HPV-18 and other strains, is becoming more and more associated with the development of oropharyngeal tumours [1] [2] [3]. It is possible that HPV-positive HNSCC might be avoided by effective vaccination campaigns across the world because the two most prevalent oncogenic HPVs, HPV-16 and HPV-18, are covered by FDA-approved HPV vaccines [4]. Smoking is still the main risk factor for HNSCCs of the oral cavity and larynx, which are now referred to as HPV-negative HNSCCs. Many patients with advanced-stage HNSCC appear without a clinical history of a pre-malignancy, despite the fact that certain oral pre-malignant lesions (OPLs), which manifest as leukoplakia (white patches) or erythroplakia (red patches), proceed to invasive cancer. Most patients with HNSCC require multimodality techniques and, consequently, multidisciplinary care, with the exception of early-stage oral cavity cancers, which may be treated

with surgery alone, and larynx tumours, which can be treated with radiation or surgery alone [5] [6, 7, 8, 9]. Detailed molecular characterization as well as immune profiling of HNSCC suggests that incorporation of prognostic and predictive biomarkers into clinical management may overcome obstacles to targeted therapies and enable prolonged survival. In this Primer, we provide an overview of the types of HNSCC and their epidemiology, as well as the pathogenesis of each type and how this influences the management approach [7] [8] [9].

In recent years, advances in epigenomic technologies and machine learning algorithms have made it possible to analyze large amounts of data quickly and accurately. This has led to the development of new tools and techniques that can be used to predict the likelihood of head and neck cancer in patients based on their epigenomic profiles.

By combining epigenomic and machine learning methods, researchers hope to develop a predictive model that can accurately predict the likelihood of head and neck cancer in individual patients. This would allow doctors to provide earlier and more targeted treatments, potentially improving patient outcomes and survival rates.

Overall, the use of epigenomics and machine learning in cancer research represents a promising new approach to improving cancer diagnosis, prognosis, and treatment. As researchers continue to develop and refine these methods, we can expect to see significant advances in our ability to predict, diagnose, and treat head and neck cancer, improving the lives of patients around the world.

1.5 Incidence, Prevalence and Mortality

With 890,000 new cases and 450,000 fatalities from HNSCC in 2018, it is the sixth most prevalent cancer globally (Fig. 1) [10] [11] [12]. Figure 1 shows the prevalence of HNSCC in the world.

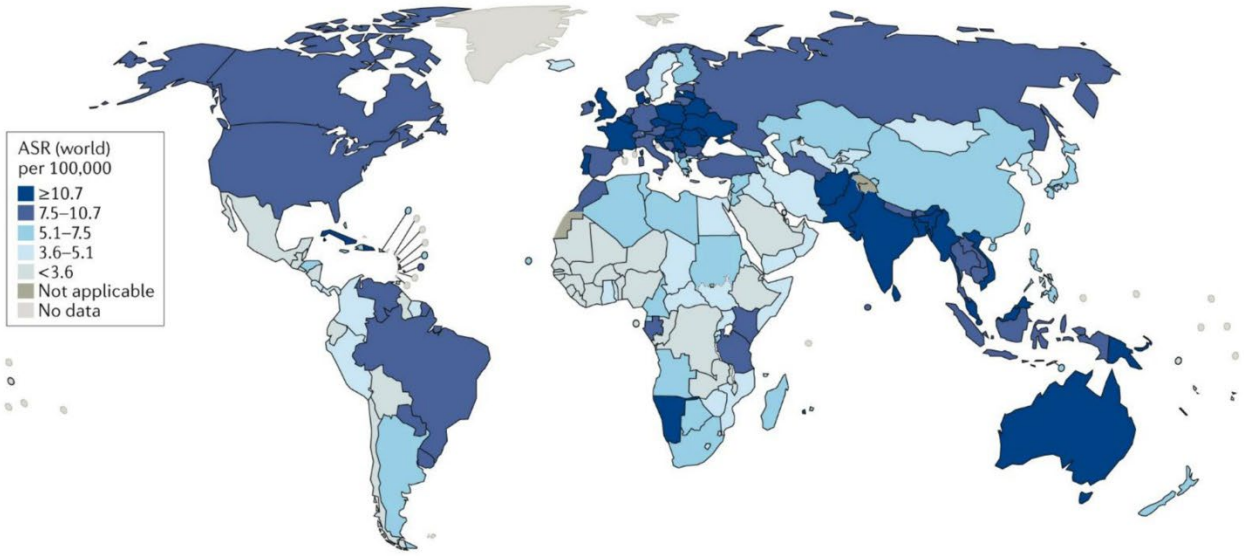

Figure 1.1 Global incidence of head and neck squamous cell carcinoma

By 2030, 1.08 million additional cases of HNSCC, or a 30% increase from the current incidence rate, are expected (GLOBOCAN; gco.iarc.fr/today) [10] [11] [12]. Consumption of certain carcinogen-containing goods is linked to high rates of HNSCC in places like Southeast Asia and Australia but rising HPV oropharyngeal infection rates are linked to high rates of HNSCC in the USA and Western Europe. [13] [14] [15]. Males are 2-4 times more likely than women to acquire HNSCC. In contrast to Epstein-Barr virus (EBV)-related nasopharyngeal cancer and HPV-associated oropharyngeal cancer, which have median diagnosis ages of 53 years and 50 years, respectively, non-virally associated HNSCC has a median diagnostic age of 66 years [16] [17]. Over the past three decades, there has been a little improvement in the survival rate for HNSCC.

For instance, the 5-year survival rate increased from 55% from 1992 to 1996 to 66 % from 2002 to 2006 when examined across all age groups and anatomic locations in the SEER registry [18]. With the exception of elderly patients (>75 years) and all anatomic locations save the larynx, subgroup analysis showed better survival in all age categories. Rather than advancements in multimodality therapy per se, the development of HPV-associated HNSCC is largely responsible for the increase in survival rate. A subsequent SEER analysis that included tissue evaluation for HPV showed better survival in HPV-positive but not HPV-negative HNSCC [19].

1.6. Objectives of the study and Outline of the thesis

The objective of this study is to develop a predictive model for head and neck cancer in patients using epigenomics data and advanced machine learning methods. Specifically, we aim to identify epigenetic changes that are associated with the development of head and neck cancer, as well as changes that are associated with cancer progression and treatment response.

To achieve this objective, we used a TCGA dataset of epigenomic data from patients with head and neck cancer, which also consists of data from healthy individuals. Predictive models are developed that can accurately predict the likelihood of head and neck cancer in patients based on their epigenomic profiles. The models are validated using the hold out method.

In addition to developing a predictive model for head and neck cancer, we also aim to identify specific epigenetic changes that are associated with cancer progression and treatment response by performing Survival Analysis. This information could be used to develop more targeted and effective treatments for head and neck cancer patients.

Overall, the objective of this study is to use epigenomics data and advanced machine learning methods to improve our understanding of the epigenetic mechanisms underlying head and neck cancer and to develop new diagnostic model and to assess the survival rates of patients based on factors such as age, gender, smoking status, tumor size, and location of site. By achieving this objective, we hope to improve patient outcomes for head and neck cancer.

The rest of the thesis is organized as below:

In Chapter 2, we discuss literature review of related works towards HNSCC prediction and Survival Analysis using machine learning and deep learning methods. In Chapter 3, the data set, data processing and data analysis are discussed in detail. In Chapter 4, the feature extraction with different feature selection methods explained in details. Chapter 5, we have provided comprehensive explanations for each method we employed. Chapter 6, the results of our experimentation are discussed and compared with some related papers. Chapter 7 concludes the study and future work is discussed.

Chapter 2: Literature Review

Head and neck squamous cell carcinoma (HNSCC), which causes 330,000 deaths and 650,000 new cases annually globally [20][21], is the sixth most common kind of cancer. Nearly all HNSCC patients have oral squamous cell carcinomas (OSCC) [22]. More than 90% of HNSCC cases are associated to patients with OSCC [23]. South Asian countries including India [24], Bangladesh (Collaboration, 2019), and Pakistan [25] have higher OSCC incidence rates than other parts of the world on average (Collaboration, 2019). Oncogenic viruses such as the Human Papillomavirus (HPV) [26], excessive alcohol use, chewing tobacco usage, and cigarette smoking are some of the known risk factors for HNSCC [27]. Additionally, epigenetic regulation, mutation, copy number variation (CNV), and immunological host response have a substantial impact on the development of cancer [28]. Despite recent breakthroughs in cancer detection and treatment, the overall 5-year survival rate for HNSCC is less than 50% due to a lack of suitable diagnostic markers and targeted treatments [29]. In many cancers, including HNSCC, early finding is known to improve survival rates in comparison to late discovery. According to the American Joint Committee on Cancer Stages (TNM), an early-stage primary tumor is one with a diameter of 2-4 cm and no lymph node growth or metastases (TNM stage I and II). A tumor is considered advanced (late stage), (https://www.cancer.org/treatment/understanding-yourdiagnosis/staging.html), if it is bigger (>5 cm) and has either grown into just nearby lymph nodes (TNM stage III) or has metastasized to other areas of the body (TNM stage IV). Over the past few decades, there has been a lot of new study in HNSCC, but no breakthroughs that are clinically important. Even if certain biomarkers exist (such as HPV +ve and -ve), they don't have important qualities like high specificity and sensitivity, a low cost, or a rapid turnaround. A timely and accurate diagnosis would

have several benefits for the patients, such as proper therapy lowering morbidity and enhancing treatment results.

Unfortunately, no commonly used biomarker for HNSCC has been authorized for use in clinical trials. There is a critical need for clinically relevant biomarkers and innovative, effective treatments for HNSCC patients. Micro-RNAs (miRNAs), also known as non-coding RNAs, range in length from 18 to 25 nucleotides. They can regulate mRNA expression by triggering mRNA breakdown by binding with the 30 UTR. Because they regulate the production of mRNA, these miRNAs have important regulatory roles in processes such as cell division, maturation, angiogenesis, proliferation, migration, invasion, metastasis, autophagy, and apoptosis [30]. These miRNAs themselves, nevertheless, have the potential to exhibit dysregulation in a number of diseases, including cancer [31]. Many miRNAs have been very well documented due to their biological significance in cancer and their ability to regulate the expression of numerous cancer pathways [32]. It is also well known that changes in the miRNA expression profile in certain malignancies can be detected even before the development of clinical symptoms [33]. Due to their stability and ease of detection (in tissues and body fluids), these miRNAs offer a logical way for the development of excellent biomarkers [34]. Analyzing miRNA expression levels may be able to provide new therapeutic targets or shed light on the processes behind tumor progression. Mathematical modelling has long been used extensively in the modelling, classification, and prediction of disease molecular function. Machine learning now has new chances to comprehend and keep track of disease processes because to the advancement of next-generation sequencing (NGS) technology and the accessibility of enormous volumes of sequencing data. Machine learning techniques are often used for risk categorization, mutational frequency prediction, CNV,

and the identification of new targets. The use of machine learning (ML) techniques as a tool for cancer diagnosis and prognosis has been suggested [35]. In order to uncover Rat Sarcoma (RAS) activation pathways in malignancies, for instance, expression data, SNPs (single nucleotide polymorphisms), and CNVs (copy number variations) have all been used in combination with machine learning algorithms [36]. A multi-parametric Decision Support System (DSS) has been used to estimate the OSCC progression and likely relapse (local or metastatic) utilizing a range of heterogenic data (clinical information, genomes, and imaging data) [37]. In research by Avissar to determine the miRNAs that indicate the occurrence of HNSCC, miR-221 and miR375 were demonstrated to be predictors [38]. However, the study did not include grades or developmental stages. Neural networks were used to predict tongue cancer recurrences in a recent machine learning study [39]. For those with oral cancer, Kim developed algorithms for survival prediction [40]. MiRNA expression patterns have not yet been employed in any research to categories the clinical stages of HNSCC patients. Given that detecting late phases typically necessitates time-consuming examination and intrusive procedures, such study may be highly beneficial in reaching improved diagnostic and therapeutic outcomes. Such a strategy can be used to difficult-to-identify and-reach conditions like pancreatic cancer. The clinical stage must be kept in mind since it is known to have a substantial impact on the patients' overall survival. In addition to patient risk stratification that may help in the selection of treatment regimens, such as whether it is effective to use less aggressive treatment than highly toxic therapies, these studies have identified signatures that can effectively classify HNSCC patients into early and late clinical stages.

In this review, we will highlight novel diagnostic biomarkers that are in research as well as known diagnostic biomarkers that are clinically used in HNSCC. The prognosis of HNSCC

patients is frequently discussed in HNSCC literature [41, 42]. However, other organizations agree that further integrative research is required to identify possible new biomarkers [43–45]. Unsupervised molecular data transformations (such as RNA sequencing, DNA methylation, and miRNA sequencing) are known to aid machine learning-based survival prediction in various malignancies [46, 47]. With HNSCC, not much study has been done in this way. Similar to this, there is no literature on machine learning imputation of sparse clinical data.

A coordinated initiative to collect, exchange, and analyze the data from next-generation molecular sequencing is the Cancer Genome Atlas (TCGA) Research Network [48], which aims to advance our knowledge of cancer processes at the molecular level [49]. Data used in our study came from the Genomic Data Centre Portal [50] and included information on 528 TCGA HNSC patients, including genotyping, whole exome sequencing, solid-tumor RNA expression, methylation data, and clinical details. Only the clinical data and RNA expression variables are taken into account in this work. Clinical data comprises details about the tumor grade, patient demographics, histories of alcohol and/or drug use, as well as a number of characteristics that are linked to disease progression, such lympho-vascular invasion and margin status. Since the presence of the human papillomavirus (HPV) has a significant impact on prognosis and tumor progression [51, 52], the HPV status (based on ISH and P16 tests) was also included. Large, multi-institutional datasets pose a few difficulties for the creation of clinical decision support tools and methodologies. In particular, the TCGA-clinical HNSC's data fields have concerns with sparsity and consistency. No one patient had any of the 15 specified clinical features that were important to the treatment plan filled out. More specifically, the proportion of cases (from of a total of 564 potential cases in TCGA) in which 520 patients lacked tumors. In addition to the issue with missing

data, some fields had inconsistently filled in data, and the replies varied as a result of human error. Due to these problems, it was necessary to perform considerable pre-processing and build an expert system employing domain-specific knowledge to determine whether each patient had received a specific type of therapy. Even after this pre-processing and condensing of treatment fields, issues of missing data persisted. Whether a patient had received radiotherapy and/or chemotherapy was unclear for 47% and 27% of cases, respectively. Excluding instances or variables with missing data is one approach that might be used to address such issues, as was previously done with this dataset [53]. It is tried to make the most of the available data by imputing missing values because these features are important to our decision support objectives and there aren't enough examples to draw from. Molecule datatypes frequently have a very high dimensionality. In order to make the greatest use of the computing resources available while using such data, feature selection and dimensionality reduction approaches are required. Feature filtering, feature transformations, and wrapper methods such sequential selection are some of the techniques for selection and dimensionality reduction [54]. With respect to the TCGA-HNSC, feature filtering and an unsupervised sparse PCA feature space transformation of over 20,000 solid-tumor RNA expression variables were used in this study. Using techniques like pre-processing, balancing the imbalance class, and missing data imputation, a considerably more condensed and useful dataset was produced.

Radiomics characteristics are calculated for quantitative features that identify an illness in medical pictures. From the larger viewpoints of neural networks and machine learning algorithms, the relevance of radiomic characteristics in tumor classification has been investigated. A more well-liked method of categorization involves employing machine learning algorithms to study a

collection of attributes that are helpful in diagnosis and then applying classifiers. In this regard, Wu et al. [55] applied Random Forests, Naive Bayes, and K-nearest neighbors to the radiomic data in order to study the link between radiomic features and the tumor histology. In the paper by Chen et al. [56], the authors developed a radiomics signature for predicting lymph node metastasis in patients with early-stage cervical cancer. The signature was based on features extracted from MRI images using a texture analysis approach.

The radiomics signature consisted of 15 texture features, including the four Laws characteristics, minimum, energy, skewness, and uniformity. The Laws features are based on the Laws texture energy measures, which are derived from a set of pre-defined filters. The minimum feature is a measure of the minimum pixel intensity value within a region of interest, while energy measures the overall amount of texture in the image. Skewness is a measure of the asymmetry of the texture distribution, while uniformity is a measure of the homogeneity of the texture distribution.

The authors used a support vector machine (SVM) classifier to predict lymph node metastasis based on the radiomics signature. They reported that the signature achieved an area under the receiver operating characteristic curve (AUC) of 0.77, which was significantly higher than the AUC of 0.66 achieved by clinical features alone. He also used SFS and SVM classifiers. Choi et al. [57] utilized a hierarchical clustering technique to determine that the bounding box anterior-posterior dimension and the standard deviation of the inverse difference moment were the two most important variables for diagnosing lung cancer.

The creation of effective neural networks based on radiomics is another forward-thinking strategy for classifying tumors. To categorize the lung nodules in CT scans, Liu et al. [58],[59] suggested a multi-view convolutional neural networks (MV-CNN) that utilized several views as input channels. In this paper, the authors proposed a multi-view convolutional neural network (MV-CNN) that takes advantage of multiple views of lung CT scans to improve the classification performance of lung nodules. They used a 3D CNN architecture and treated each view as a separate input channel to the network. The proposed approach achieved state-of-the-art performance on the LIDC dataset for lung nodule classification. A classification neural network based on deep learning characteristics was used by [60], [61] to classify lung nodules in CT scans. Kumar et al. [62] devised a computer-aided diagnostic method that classified benign and malignant lung nodules by extracting deep features using an auto-encoder and a decision tree classifier.

Chapter 3: Data and Processing

3.1 Overview of Dataset

We used HNSC dataset in this study TCGA (http://portal.gdc.cancer.gov) [63]. The data consisted of Radiomics Images of Patients, Clinical Data and Feature Variable Description. Table 3.1 shows a sample and some of the variables of the Clinical Data of the Patients. The data consisted of 493 patient's data and 84 columns, which stored the clinical information of the patients.

DNA Methylation is an epigenetic modification where a methyl group is added to DNA, influencing gene expression and playing a role in various diseases, including cancer. Histone data comprises information about chemical modifications on histone proteins, influencing gene expression and chromatin structure. Studying histone modifications provides insights into epigenetic regulation, chromatin states, and their roles in biological processes, including development and disease. Human Genome is the complete set of genetic material in humans, consisting of protein-coding genes and non-coding regions, enabling the study of genetic variations and their relationship to traits, diseases, and human biology. RNA Sequences are transcribed from DNA and provide crucial information about gene expression and regulatory processes. They can be analyzed to identify different RNA types, study alternative splicing, RNA editing, and non-coding RNA molecules, enhancing our understanding of gene function and the molecular mechanisms underlying diseases such as cancer.

Table 3.1. Sample of the clinical data for HNSCC

TCIA Radiomics ID	HNSCC-01-0003	HNSCC-01-0024	HNSCC-01-0026	HNSCC-01-0030	HNSCC-01-0033
Gender	Male	Female	Male	Male	Female

Age at Diag	48	80	59	63	44
Smoking status	Former	Never	Never	Never	Current
Smoking status (Packs-Years)	5.0	0.0	0.0	0.0	30.0
Tumor laterality	R	L	R	R	Midline
Cancer subsite of origin	Tonsil	Soft palate	Base of tongue	Base of tongue	NOS
HPV Status	Unknown	Unknown	Unknown	Unknown	Unknown
T-category	2	3	2	4	3
N-category	2	2	2	1	0
AJCC Stage (7th edition)	IV	III	IV	IV	III
Therapeutic Combination	Concurrent chemoradiotherapy	Radiation alone	Concurrent chemoradiotherapy	Concurrent chemoradiotherapy	Concurrent chemoradiotherapy
Radiation Treatment_duration	38	38	46	44	44
Total prescribed Radiation treatment dose	66.0	66.0	69.96	69.96	70.0
Radiation treatment_number of fractions	33	30	33	33	33
Radiation treatment_dose per fraction	2.0	2.2	2.12	2.12	2.12
Vital status	Alive	Alive	Alive	Alive	Dead
Overall survival_duration	3192	3466	1933	3007	3739
Local control	No	No	Yes	Yes	Yes
Local control_duration	3192	3466	1933	3007	3739
Regional control	Yes	Yes	Yes	Yes	Yes

Regional control_duration	3192	3466	1933	3007	3739
Locoregional control	No	No	Yes	Yes	Yes
Locoregional control_duration	3192	3466	1933	3007	3739
Freedom from distant metastasis	Yes	Yes	Yes	Yes	Yes
Freedom from distant metastasis_duration	104.7	113.5	63.6	98.9	123.1
Relapse-free survival	No	No	Yes	Yes	Yes
Days to last FU	3142	3405	1907	2966	3693
Neck Dissection after IMRT	No	No	Yes	No	No

Table 3.2 shows the Variable Description of all the features present in the Clinical Features dataframe. The dataset contains variables such as Patient Identifier, Sex, Age, Date of Birth, Diagnosis of Cancer, Site of Origin and other features related to clinical diagnostic of the patients.

Table 3.2. Dataset Description

Data Category	Description
SUBJECT_ID	Patient ID number randomly assigned to each patient prior to anonymizing the DICOM PHI tag (0010,0020)
Sex	Patient sex, male or female
Age	Patient age, years
Date of Birth	Date of birth, redacted
Diag	Cancer diagnosis.
Site	Primary cancer subsite. CUP = cancer of unknown primary
Histology	Cancer histopathology. SCC=squamous cell carcinoma

Grade	Histopathologic grade.
T	AJCC 7th edition T stage
N	AJCC 7th edition N stage
M	AJCC 7th edition M stage
Stage	AJCC 7th edition summary stage
HPV status	Results of ISH testing for high-risk HPV. Blank cell indicates no data available.
Date of Diagnosis	Date of cancer diagnosis, with date offset applied
Last Contact Date	Most recent date of contact at which clinical outcomes were updated, with date offset applied.
Follow up duration (day)	Duration of follow up from diagnosis to last contact date in days.
Follow up duration (year)	Duration of follow up from diagnosis to last contact date in years.
Follow up duration (month)	Duration of follow up from diagnosis to last contact date in months.
Date of Death	Date of death, with date offset applied
Survival (months)	Overall survival in months from diagnosis to death or last contact date, if still alive
Alive or Dead	Binary indicator of vital status at last contact date.
Cause of Death	Cause of death (if applicable)
Date of recurrence	Date of cancer recurrence (if applicable) or last contact date, with date offset applied.
Disease-free interval (months)	Duration, in months, between RT completion and disease recurrence.
Site of recurrence (Distal/Local/ Locoregional)	Site of recurrence, coded as local (within or adjacent to primary tumor), locoregional (includes local or regional nodal recurrence), or distal (metastatic to distal organs or non-regional lymph nodes)
Overall Survival Censor	Overall survival Kaplan-Meier censor code: 0=censored; 1=event
Disease Specific Survival Censor	Disease specific survival Kaplan-Meier censor code: 0=censored; 1=event
Loco-regional Control Censor	Loco-regional Control Kaplan-Meier censor code: 0=censored; 1=event
Oncologic Treatment Summary	Summary of oncologic treatment administered to patient. List indicates sequence of treatments. CCRT=concurrent chemoradiotherapy, ERT=external beam radiotherapy, CMT=chemotherapy
Induction Chemotherapy	Summary of induction chemotherapy schema

Chemotherapy Regimen	Concurrent chemotherapy regimen used during radiotherapy.
Platinum-based chemotherapy	Whether the patient receved platinum-based chemotherapy during treatment (inclusive of induction, concurrent, or adjuvant)
Received Concurrent Chemoradiotherapy ?	Yes=received concurrent chemoradiotherapy, no=did not receive concurrent chemoradiotherapy
CCRT Chemotherapy Regimen	Chemotherapy regimen used during CCRT
Surgery Summary	Description of surgical treatment applied.
RT Total Dose (Gy)	Total RT dose delivered during radiotherapy.
Dose/Fraction (Gy/fx)	Dose delivered to prescription target volume (gross disease or post-operative tumor bed)
Number of Fractions	Number of RT fractions delivered.
Unplanned Additional Oncologic Treatment	Description of any unplanned oncologic treatment that followed intial treatment schema.
Smoking History	Smoking history coded: 0=never smoker, 1= fewer than 10 pack-years, 2= 10 or more pack-years
Current Smoker	Current smoker: 0=no, 1=yes
Received Feeding Tube (Y/N)	Receipt of feeding tube during RT
Type of feeding tube	Type of feeding tube placed.
Date Feeding tube placed	Date feeding tube was placed, with date offset applied.
Date Feeding tube removed	Date feeding tube was removed, with date offset applied.
Feeding tube duration (months)	Duration, in months, between placement and removal of feeding tube.
Feeding tube note	Free entry field with notes related to feeding tube placement or removal.
Height (cm)	Patient height, in centimeters
BW Start tx (kg)	Patient body mass at the start of radiotherapy, in kilograms
BW stop treat (kg)	Patient body mass at the completion of radiotherapy, in kilograms
Height (m)	Patient height, in meters

BMI start treat (kg/m2)	Patient BMI at the start of radiotherapy.
BMI stop treat (kg/m2)	Patient BMI t the completion of radiotherapy.
Date Start RT	Date of first radiotherapy fraction, with date offset applied
Date Stop RT	Date of final radiotherapy fraction, with date offset applied.
Total RT treatment time (days)	Total number of days from RT start to completion
Time between pre and post image (months)	Interval, in months, between pre-RT PET/CT and post-RT PET/CT used to determine body composition
Time from preRT image to start RT (month)	Interval, in months, between pre-RT PET/CT and RT start date
Time from RT stop to follow up imaging (months)	Interval, in months, between RT completion and post-RT PET/CT
Pre-RT L3 Skeletal Muscle Cross Sectional Area (cm2)	Skeletal muscle cross sectional area at the L3 level calculated from pre-RT PET/CT imaging
Post-RT L3 Skeletal Muscle Cross Sectional Area (cm2)	Skeletal muscle cross sectional area at the L3 level calculated from post-RT PET/CT imaging
Pre-RT L3 Adipose Tissue Cross Sectional Area (cm2)	Adipose tissue cross sectional area (in cm2) at the L3 level calculated from pre-RT PET/CT imaging
Post-RT L3 Adipose Tissue Cross Sectional Area (cm2)	Adipose tissue cross sectional area (in cm2) at the L3 level calculated from post-RT PET/CT imaging
Pre-RT L3 Skeletal Muscle Index (cm2/m2)	Skeletal muscle index, in cm2/m2, calculated by dividing pre-RT skeletal muscle cross-sectional area by the square of patient height, in meters
Post-RT L3 Skeletal Muscle Index (cm2/m2)	Skeletal muscle index, in cm2/m2, calculated by dividing post-RT skeletal muscle cross-sectional area by the square of patient height, in meters

Pre-RT L3 Adiposity Index (cm2/m2)	Adipose tissue index, in cm2/m2, calculated by dividing pre-RT skeletal muscle cross-sectional area by the square of patient height, in meters
Post-RT L3 Adiposity Index (cm2/m2)	Adipose tissue index, in cm2/m2, calculated by dividing post-RT skeletal muscle cross-sectional area by the square of patient height, in meters
Pre-RT CT-derived lean body mass (kg)	Lean body mass, in kilograms, calculated based on pre-RT CT imaging using the formulae of Mourtzakis, et al.
Post-RT CT-derived lean body mass (kg)	Lean body mass, in kilograms, calculated based on post-RT CT imaging using the formulae of Mourtzakis, et al.
Pre-RT CT-derived fat body mass (kg)	Fat mass, in kilograms, calculated based on pre-RT CT imaging using the formulae of Mourtzakis, et al.
Post-RT CT-derived fat body mass (kg)	Fat mass, in kilograms, calculated based on post-RT CT imaging using the formulae of Mourtzakis, et al.
PreRT Skeletal Muscle status	Skeletal muscle status, dichotomized as depleted vs non-depleted, based on skeletal muscle index calculated from pre-RT imaging
PostRT Skeletal Muscle status	Skeletal muscle status, dichotomized as depleted vs non-depleted, based on skeletal muscle index calculated from post-RT imaging
Pre-RT Imaging Date	Date of pre-RT diagnostic imaging, with date offset applied
Pre-RT Imaging Modality	Modality of pre-RT diagnostic imaging
CT sim date	Date of RT treatment planning simulation scan, with date offset applied
post-RT imaging date	Date of post-RT diagnostic imaging (used for body composition analysis), with date offset applied. May precede or follow or function as recurrence diagnostic imaging.
post-RT imaging modality	Modality of post-RT diagnostic imaging
Recurrence imaging date	Date of imaging used to confirm or diagnose recurrence/metastasis, with date offset applied
Recurrence imaging modality	Modality of imaging used to confirm or diagnose recurrence/metastasis
Additional imaging date	Date of additional imaging included in collection, with date offset applied
Additional imaging modality	Modality of additional imaging included in collection
Additional imaging notes	Notes regarding inclusion of additional imaging

3.2 Exploratory Data Analysis

After, loading the dataset, we conducted Exploratory Data Analysis across different variables.

3.2.1 Gender

Figure 3.1 shows the distribution of Gender. There are 14% female patients and 86% male patients.

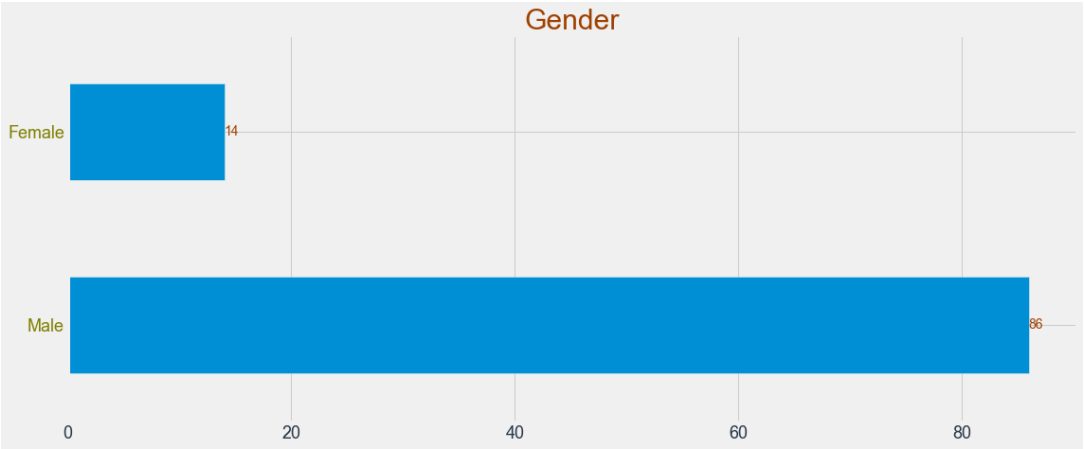

Figure 3.1 Distribution of Gender

3.2.2 Smoking Status

Figure 3.2 shows the distribution of Smoking Status. There are 29% Current Smokers patients, 33% Former Smoker patients and 38% patients who never smoked in their lifetime.

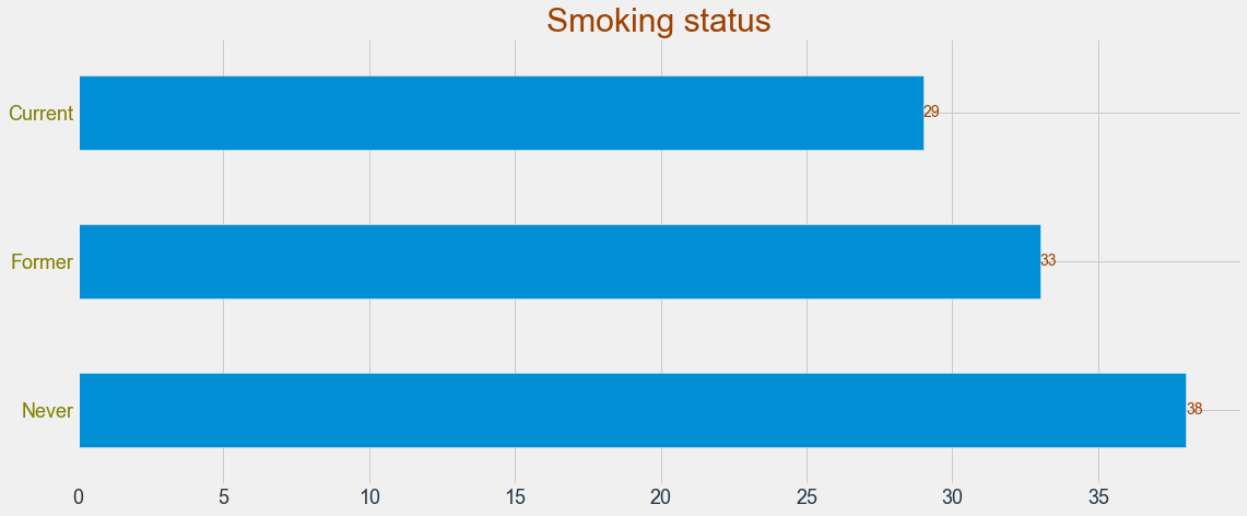

Figure 3.2 Distribution of Smoking Status

3.2.3 Tumor Literality

Figure 3.3 shows the distribution of Tumor laterality. There are 6% patients with Mid-line Tumor Literality, 32% patients with L- Tumor Literality and 62% patients with R-Tumor Literality. So, it can be seen that majority of patients have R-Tumor Literality.

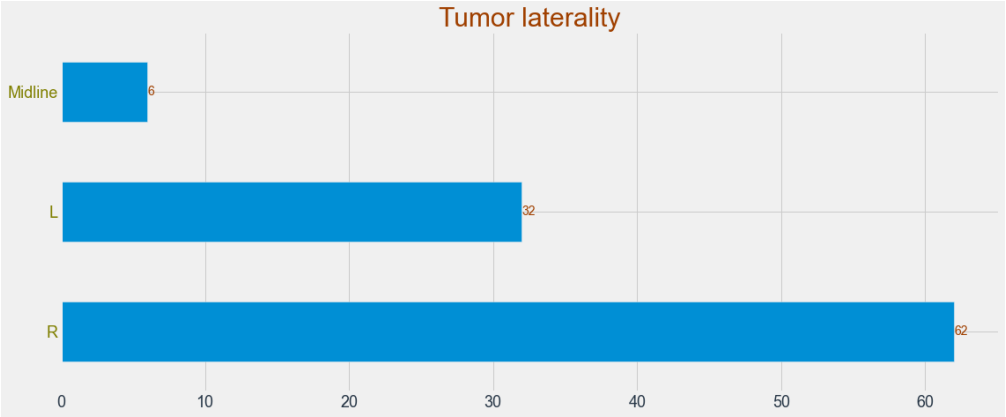

Figure 3.3 Distribution of Tumor Literality

3.2.4 Cancer sub-site of origin

Figure 3.4 shows the distribution of Cancer Sub-site of Origin. Most of the Patients have Base of tongue as the cancer sub-site of origin, which is 52%, 36% of the patients have tonsil as the cancer

sub-site. Very low percentages of patients have Nose, Glossopharyngeal sulcus and soft palate as the cancer sub-site of origin.

```
Base of tongue          52.00000
Tonsil                  36.00000
NOSE                     7.00000
Glossopharyngeal sulcus  3.00000
Soft palate              1.00000
```

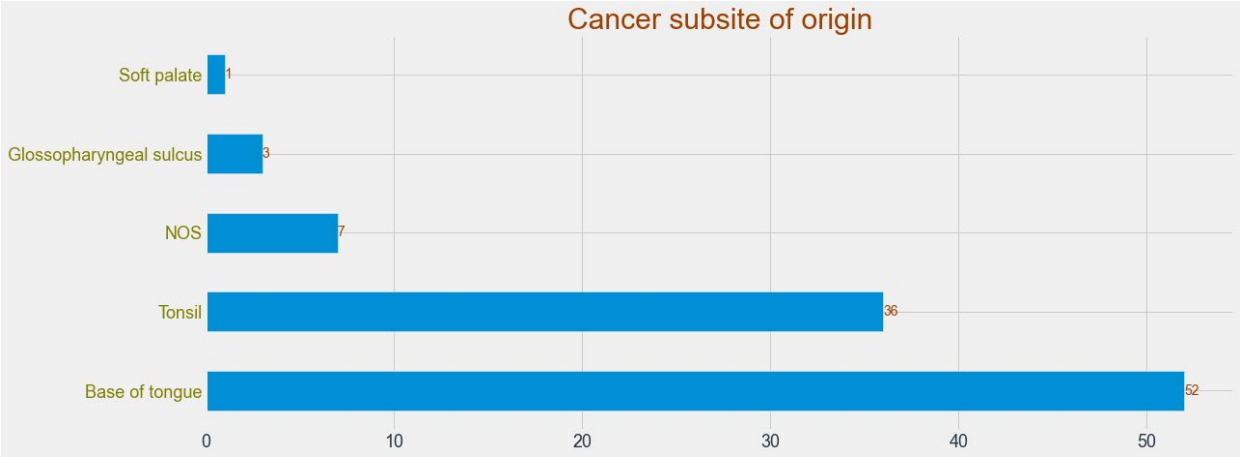

Figure 3.4 Distribution of Cancer Sub-site of Origin

3.2.5 HPV Status

Figure 3.5 shows the distribution of HPV Status. Most of the Patients have unknown HPV status which is 68%, 26% of the patients have P-HPV status 6% of patients have N-HPV status.

```
Unknown   68.00000
P         26.00000
N          6.00000
```

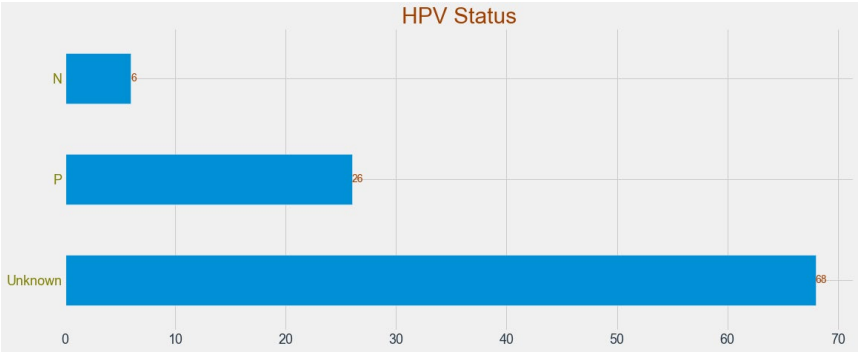

Figure 3.5 Distribution of HPV Status

3.2.6 T-category

Figure 3.6 shows the distribution of T-category. Most of the Patients have 2nd T-category stage which is 42%, 29% of the patients have 3rd T-category stage, 22% of the patients have 4th T-category stage and 7% of the patients have 1st T-category stage.

```
2   42.00000
3   29.00000
4   22.00000
1    7.00000
```

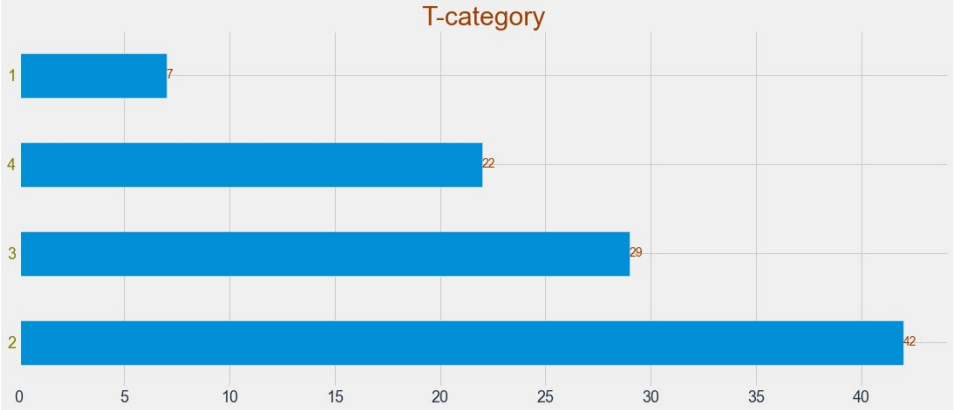

Figure 3.6 Distribution of T-Category

3.2.7 N-category

Figure 3.7 shows the distribution of N-category. Most of the Patients have 2nd T-category stage which is 77% while very small percentage of patients have None, 1st or 3rd N-category stage.

```
2   77.00000
0   10.00000
1    7.00000
3    6.00000
```

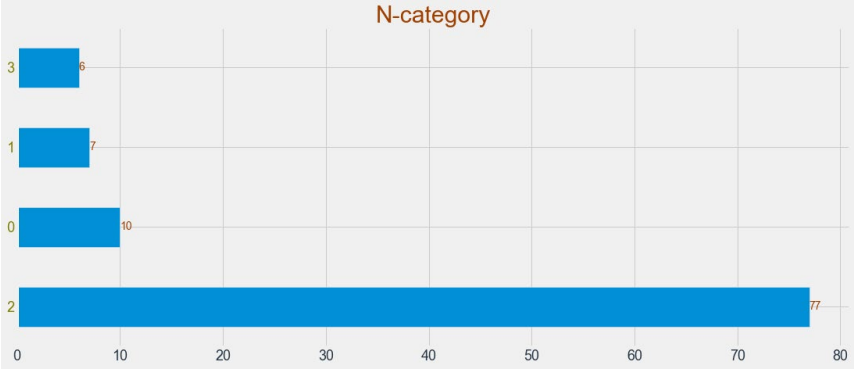

Figure 3.7 Distribution of N-Category

3.2.8 AJCC Stage (7th edition)

Figure 3.8 shows the distribution of AJCC Stage (7th edition). Majority of the Patients have IV stage AJCC which is 84%, 12% of the patients have III AJCC stage, 4% of the patients have II AJCC stage.

```
IV    84.00000
III   12.00000
II     4.00000
```

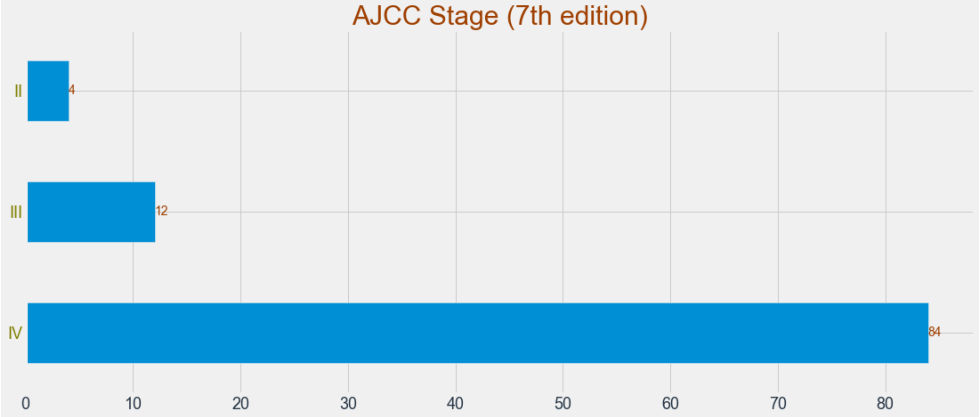

Figure 3.8 Distribution of AJCC Stage (7th Edition)

3.2.9 Therapeutic Combination

Figure 3.9 shows the distribution of Therapeutic Combination. Majority of the Patients have Concurrent chemoradiotherapy Therapeutic Combination which is 65%, 19% of the patients have Induction chemotherapy + concurrent chemoradiotherapy Therapeutic Combination, 12% of the

patients have Induction chemotherapy + Radiation Therapeutic Combination and 4% of the patients were treated with radiation alone.

Concurrent chemoradiotherapy	**65.00000**
Induction chemotherapy + concurrent chemoradiotherapy	**19.00000**
Induction chemotherapy + Radiation alone	**12.00000**
Radiation alone	**4.00000**

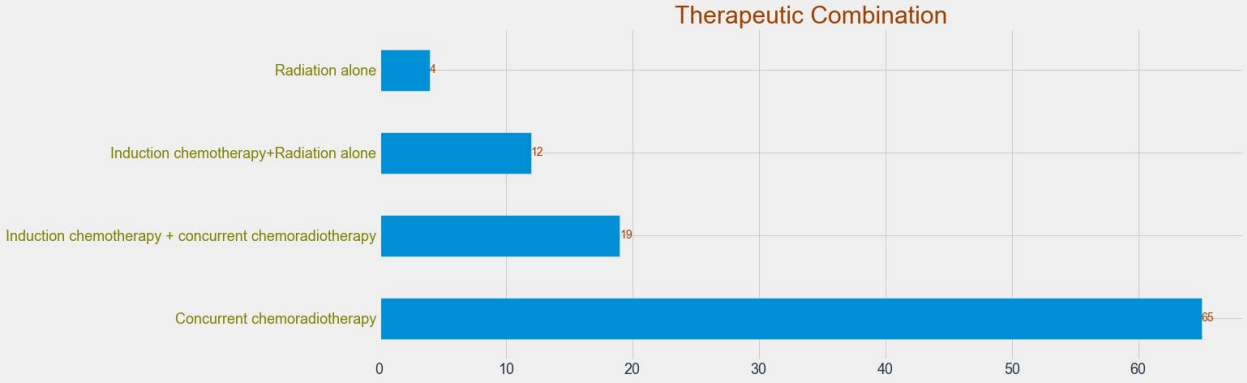

Figure 3.9 Distribution of Therapeutic Combination

Figure 3.10 shows the target variable, there are 45 patients or 65%, which are alive and 24 patients or 35% which were dead.

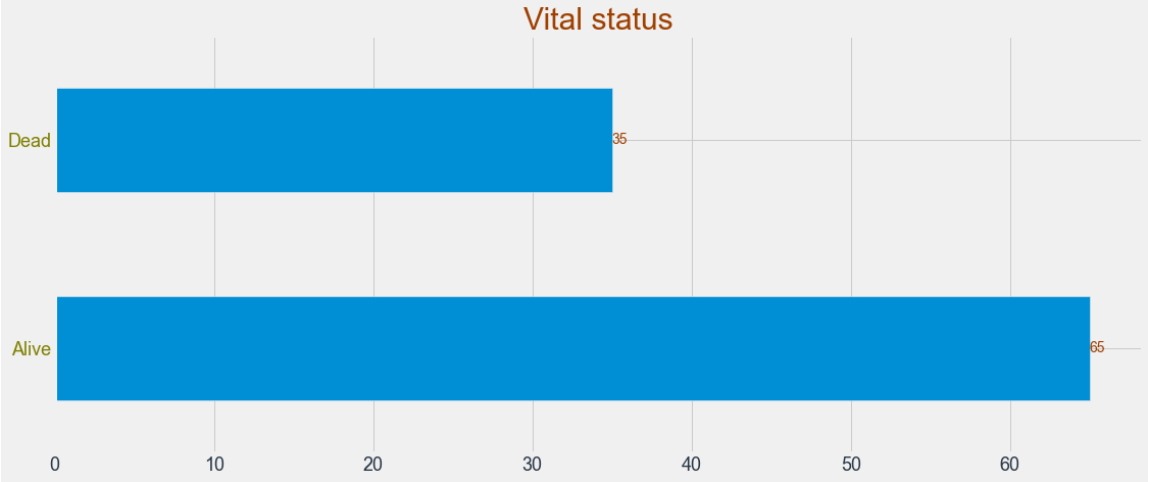

Figure 3.10 Distribution of Vital Status

Figure 3.11 to Figure 3.17 shows the histogram of Age at Diagnosis, Smoking status (Packs-Years), Radiation Treatment duration, Total prescribed Radiation treatment dose, Radiation treatment_number of fractions, Radiation treatment_dose per fraction and Overall

survival_duration. Most of the patients have the age in between 50 to 60. The mean age of the patients at the age of diagnosis is 55 years. From histogram of smoking status, we can see that mot of the patients consume 40 cigarette packs per year. It can also be concluded from Radiation Treatment duration, that most of the patients were treated for 45 days with radiation.

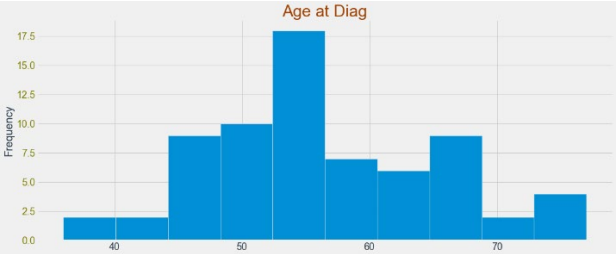

Figure 3.11 Histogram of Age at Diagnosis

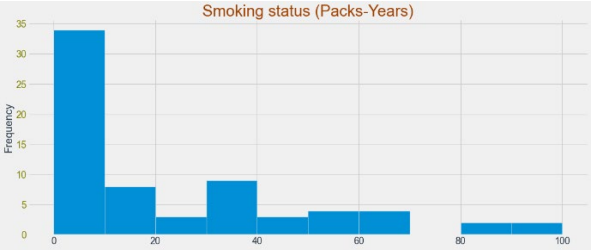

Figure 3.13 Histogram of Smoking status (Packs-Years)

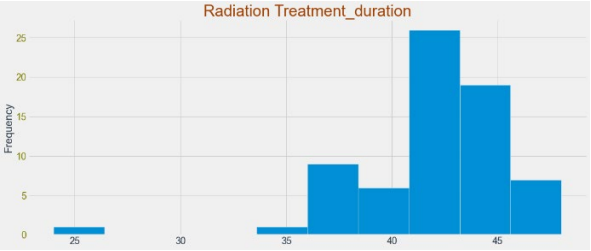

Figure 3.15 Histogram of Radiation Treatment duration

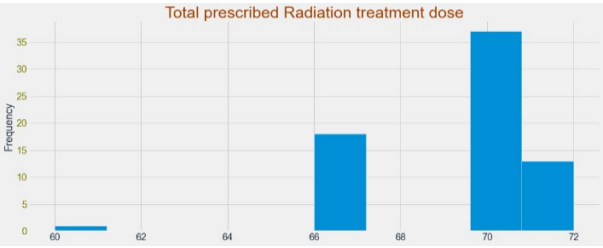

Figure 3.17 Histogram of Total prescribed Radiation treatment dose

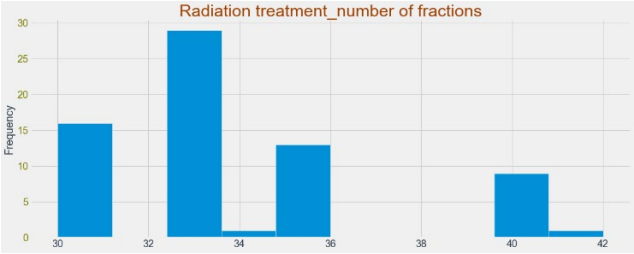

Figure 3.12 Histogram of Radiation treatment number of fractions

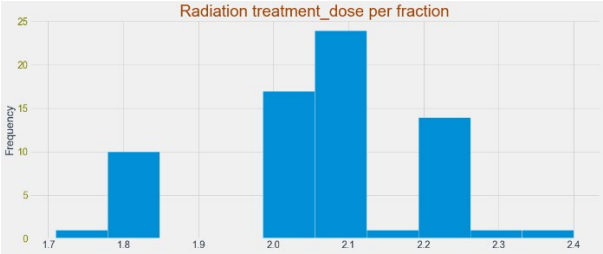

Figure 3.14 Histogram of Radiation treatment dose/fraction

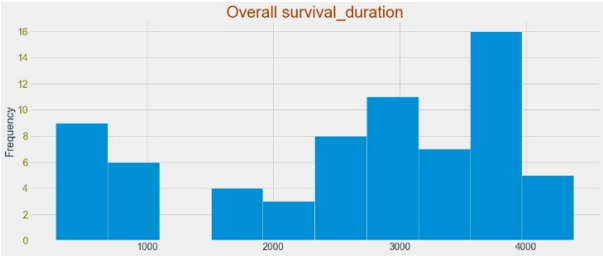

Figure 3.16 Histogram of Overall survival_duration

3.3 Correlation Analysis

Figure 3.18 shows the correlation analysis among all the variables. Some variables have low positive correlation, and some variables have low negative correlation. The variable that accounts for "duration" have a perfect correlation of 1 with each other. Radiation treatment dose per fraction is negatively correlated with radiation treatment number of fraction with a value of -0.93. Vital Status is high negatively correlated to the Overall survival duration, Local control duration, Regional control duration, Loco regional control duration, Freedom from distant metastatic duration and Days to last Follow Up with a value of -0.69.

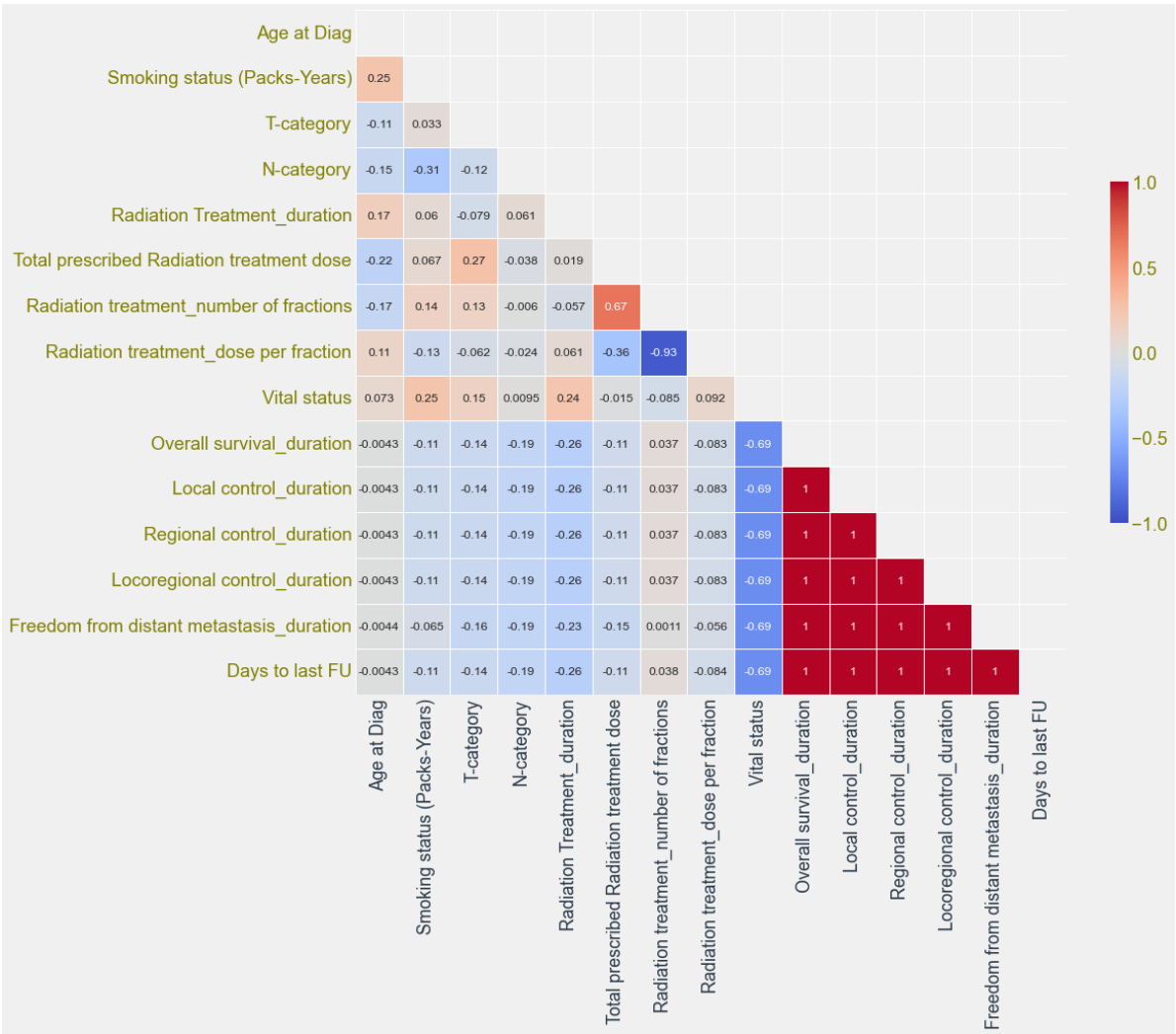

Figure 3.18 Correlation of Analysis

3.4 Survival Data Exploration

Figure 3.19 shows the Age of Patients versus the Vital Status (Alive or Dead). It can be confirmed from the grouped histogram that patients who have age more than 55 years have died because of the Cancer.

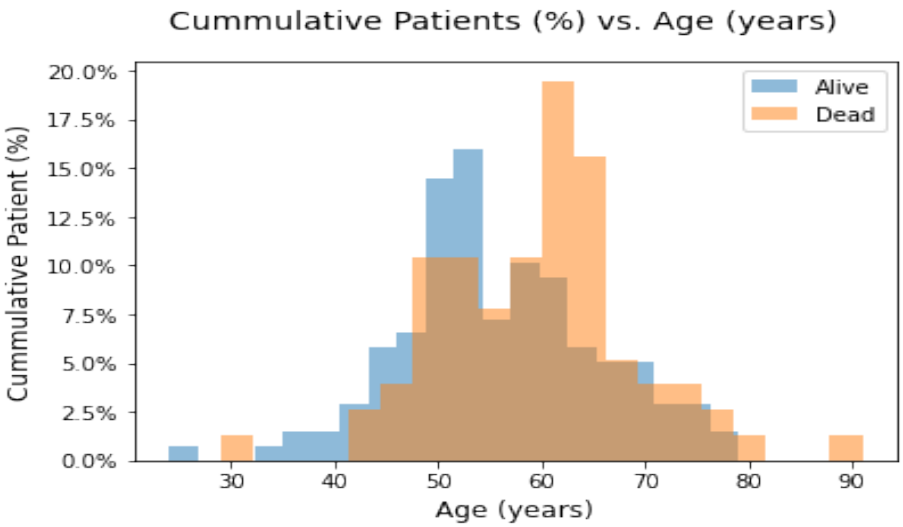

Figure 3.19 Cumulative Patients (%) vs. Age (years)

Figure 3.20 shows the Survival Status of the Patients with the Current Smoking (Yes/No). It can be seen that more patients are dead who are current smoker as compared to the patients who don't smoke currently.

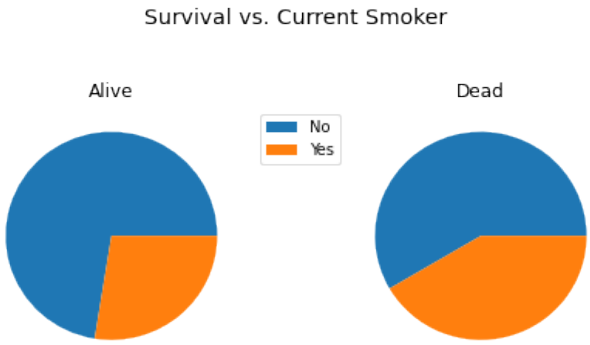

Figure 3.20 Survival vs. Current Smoker

Figure 3.21 shows the Cumulative Patients (%) versus Radiation Therapy (RT) Total Dose (Gy) for Alive and Dead patients. RT Total Dose typically refers to the total dose of radiation therapy

delivered to a patient during the course of their treatment. The dose is usually measured in units of gray (Gy) or centi-gray (cGy). The total dose is determined by several factors, including the location and size of the tumor, the stage of the cancer, and the overall health of the patient. The goal of radiation therapy is to deliver a high enough dose of radiation to kill cancer cells while minimizing damage to healthy tissues. It can be seen that for patient have a higer RT total dose among the age group of 65 to 75 and also chances of survival ae very low around the same.

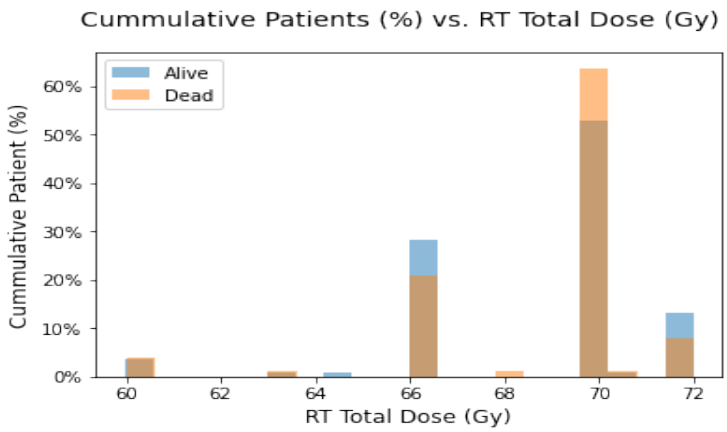

Figure 3.21 Cumulative Patients (%) vs. RT Total Dose (Gy)

Figure 3.22 shows the Cumulative Patients (%) versus Total RT Treatment Days (days). From the figure, it can be seen that as RT treatment period (days) increases, the cumulative patients (%) being dead increases.

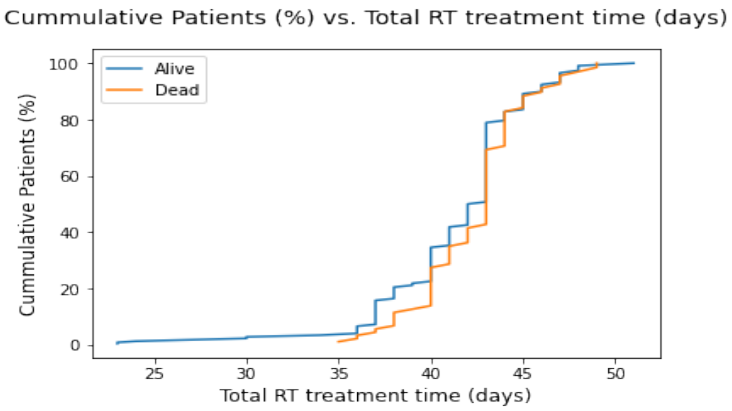

Figure 3.22 Cumulative Patients (%) vs. Total RT treatment time (days)

Figure 3.23 shows the Cumulative Patients (%) versus Survival. The curve starts at 100% because all patients are alive at the beginning of the study. As time passes, some patients will die from the disease, and the curve will drop. The rate at which the curve drops depends on the severity of the disease and the effectiveness of the treatment. The Kaplan-Meier estimator can also show the survival rate of subgroups of patients based on certain characteristics. For example, the survival rate of patients with early-stage HNSCC may be higher than the survival rate of patients with advanced-stage HNSCC. Overall, the Kaplan-Meier estimator provides valuable information about the survival rate of patients with HNSCC over time, and can help clinicians make informed decisions about treatment options and patient care.

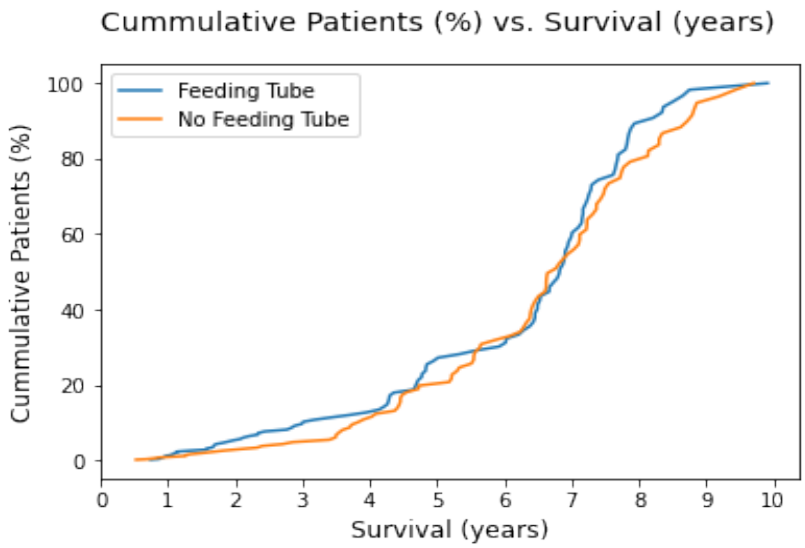

Figure 3.23 Distribution of survival based on feeding tube

The bar-chart in Figure 3.24 shows frequency of Diagnosis status. In medical terminology, "CA BOT" refers to "carcinoma of the base of tongue", which is a type of cancer that originates from the cells in the base of the tongue. It can be further classified based on the specific type of cells involved and the stage of the cancer. Treatment options may include surgery, radiation therapy, chemotherapy, or a combination of these. CA BOT is the most common type of Cancer followed by Tonsils and Superglottic and so on.

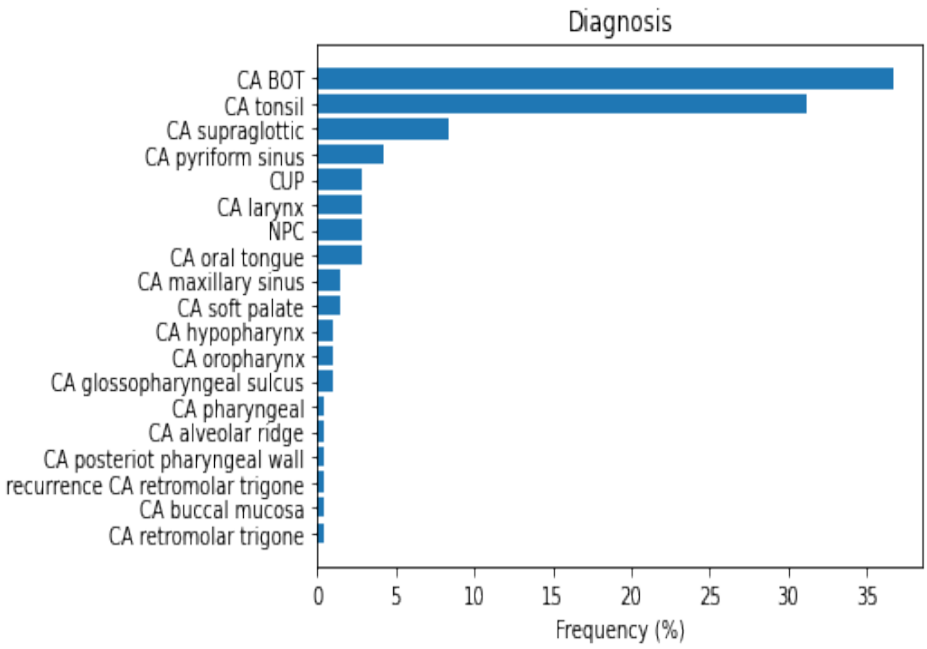

Figure 3.24 Frequency of each diagnosis

The images data consisted of images for three different angles for each patient as shown in the Figures 3.25 – 3.27. For 182 patients, the image dataset was provided, where 3 images were given for each patient in axial, coronal and sagittal pose. Axial, coronal, and sagittal are three terms used to describe different planes of the body in medical imaging, including CT scans.

Axial plane: Also known as the transverse plane, this is a horizontal plane that divides the body into upper and lower parts. In CT scans, axial images are taken from the top to the bottom of the body.

Coronal plane: This is a vertical plane that divides the body into front and back parts. In CT scans, coronal images are taken from the front to the back of the body.

Sagittal plane: This is a vertical plane that divides the body into left and right parts. In CT scans, sagittal images are taken from the side of the body.

By looking at CT scans in different planes, doctors can get a better understanding of the three-dimensional structure of the body and diagnose any medical conditions more accurately. For

example, axial images are often used to diagnose issues with the brain, while coronal and sagittal images are often used for imaging the chest, abdomen, and pelvis.

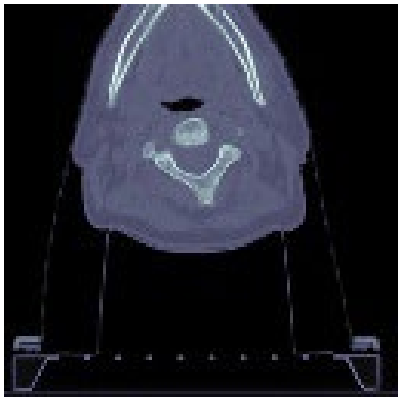

Figure 3.25 Axial Image

Figure 3.26 Coronal Image

Figure 3.27 Sagittal Image

Data augmentation is a common technique used in machine learning and computer vision to artificially expand the size of a training dataset by generating additional examples through

various transformations of existing data. In medical imaging, data augmentation can be used to create additional variations of existing images by applying rotations, translations, flips, zooms, and other transformations to the original data [64].

Normalization is a process of adjusting the intensity values of an image to make them consistent and standardized across all images. This is an important preprocessing step in medical imaging because it helps to reduce the variability in image intensities due to differences in imaging protocols, scanners, and patient characteristics. Normalization can also improve the performance of machine learning algorithms by reducing the impact of intensity variations on feature extraction [65].

In the context of the statement, the images were processed using data augmentation and normalized to a range of 0 to 1 instead of the standard 0 to 255 scale. This means that the intensity values in the images were adjusted to be within the range of 0 to 1, which can improve the consistency and comparability of the images. Additionally, using a smaller intensity range can make it easier for machine learning algorithms to process the images and extract meaningful features.

In the study, the final dataset consisted of clinical and image data from 69 patients, which was available for all 182 patients with image data and 493 patients with clinical data. These combined data were utilized for conducting all the experiments discussed in the next chapter.

Chapter 4: Feature Selection Methods

Feature selection is a critical step in machine learning and data analysis, aimed at identifying the most relevant and informative subset of features from a large pool of available features. By selecting the most significant features, feature selection helps improve model performance, reduce dimensionality, and enhance interpretability. In complex datasets with numerous features, selecting the right subset of features becomes crucial to achieve accurate and efficient models.

The process of feature selection involves evaluating the relationship between individual features and the target variable, as well as assessing the interdependence among the features themselves. By focusing on the most informative features, irrelevant or redundant features can be eliminated, leading to more efficient and effective models. Various feature selection methods have been developed and utilized across different domains. These methods can be broadly categorized into filter methods, wrapper methods, and embedded methods. Filter methods evaluate features independently of any specific machine learning algorithm, using statistical measures or information-theoretic measures to assess their relevance to the target variable. Wrapper methods, on the other hand, employ a specific machine learning algorithm to evaluate subsets of features based on their predictive performance. Embedded methods incorporate feature selection within the process of model training, where the algorithm automatically selects the most relevant features during training. In this study, we focus on exploring feature selection methods for both clinical and image data. The selected methods aim to identify the most informative features for classification tasks in these domains. By leveraging these feature selection methods, we can improve the accuracy and interpretability of machine learning models, while also reducing computational complexity and overfitting.

In this study, we employed several feature selection methods for both clinical and image data to identify the most significant features that are relevant to the classification task [66].

For clinical data, we employed the following feature selection methods: Correlation based Feature Selection, Chi-Square, Mutual Information, Principal Component Analysis, Relief-F, Recursive Feature Elimination, Select from Model, Select from Model based on Random Forest, and Select K-best. Correlation based Feature Selection and Chi-Square are the two most widely used feature selection methods for clinical data, which are based on statistical measures of correlation between features and the target variable. Mutual Information and Principal Component Analysis are other popular feature selection methods that can handle non-linear relationships between features. Relief-F is a feature selection method that is specifically designed for noisy and high-dimensional data. Recursive Feature Elimination is a wrapper-based feature selection method that uses a selected classifier to determine the best subset of features. Select from Model and Select from Model based on Random Forest are feature selection methods that utilize machine learning models to rank the importance of each feature. Finally, Select K-best is a simple filter-based feature selection method that selects the top K features based on statistical measures [67, 68].

For image data, we used the following feature selection methods: Canny Edge, Sobel Edges, and Linear Binary Pattern. These feature extraction methods are commonly used in computer vision and image processing to extract useful information from images. Canny Edge and Sobel Edges are gradient-based feature extraction methods that can detect edges in images. Linear Binary Pattern is a texture-based feature extraction method that can capture local texture patterns in images [66, 68, 69, 70]. In summary, the combination of multiple feature selection methods can help improve the accuracy and robustness of machine learning models. The choice of feature selection methods depends on the characteristics of the data and the classification task. It is

recommended to try multiple feature selection methods and compare their performance to select the best subset of features for the given task.

4.1 Numerical Feature Selection

4.1.1 Correlation based Feature Selection (CFS)

Correlation-based Feature Selection (CFS) is a widely used method for feature selection in clinical data analysis. It aims to identify the subset of features that are highly correlated with the target variable, which is the variable we are trying to predict or classify. The CFS method calculates the correlation coefficient between each feature and the target variable. The correlation coefficient measures the strength and direction of the linear relationship between two variables. A high positive correlation indicates that as one variable increases, the other variable also tends to increase. A high negative correlation, on the other hand, indicates that as one variable increases, the other variable tends to decrease.

In the context of feature selection, CFS evaluates the correlation between each feature and the target variable and selects the features that exhibit a strong correlation. The rationale behind this is that features that are highly correlated with the target variable are likely to contain valuable predictive information. By including these highly correlated features in the model, we can enhance its predictive power and improve its performance. CFS offers several advantages for feature selection in clinical data analysis. First, it provides a straightforward and intuitive way to assess the relevance of features by considering their correlation with the target variable. This allows researchers and practitioners to prioritize features based on their importance in relation to the prediction task. Second, CFS takes into account the interdependence between features, meaning that it considers the relationship between features themselves while evaluating their correlation

with the target variable. This helps to identify features that provide unique and independent information, avoiding redundancy in the selected feature set.

By applying the CFS method to clinical data, researchers and practitioners can effectively identify the most relevant features for the classification task at hand. These selected features can then be used to build robust and accurate machine learning models, improving the overall performance and interpretability of the models [71].

4.1.2 Chi-Square

Chi-Square is a widely utilized statistical method for feature selection in clinical data analysis, especially when dealing with categorical features. It focuses on measuring the dependence between categorical features and the target variable, aiming to identify the features that exhibit a significant association with the target.

The Chi-Square method works by comparing the observed frequencies of each category within a feature with the expected frequencies under the assumption of independence between the feature and the target variable. The expected frequencies are calculated based on the assumption that there is no association between the feature and the target. By comparing the observed and expected frequencies using the Chi-Square statistic, the method determines whether there is a statistically significant relationship between the feature and the target. The Chi-Square statistic measures the discrepancy between the observed and expected frequencies, taking into account both the number of observations and the expected distribution. A high Chi-Square value indicates a substantial deviation from the expected frequencies and suggests that the feature has a significant association with the target variable. Conversely, a low Chi-Square value indicates that the observed

frequencies align closely with the expected frequencies, implying a weaker association between the feature and the target.

One of the advantages of the Chi-Square method is its ability to handle categorical data, making it particularly suitable for analyzing clinical datasets where categorical variables are prevalent. It provides a statistical framework for evaluating the significance of the association between categorical features and the target variable, enabling researchers and practitioners to identify the most informative categorical features for the classification task. By employing the Chi-Square method for feature selection in clinical data analysis, researchers can effectively identify the categorical features that contribute the most to the prediction or classification of the target variable. These selected features can then be incorporated into machine learning models, enhancing their performance and interpretability [72].

4.1.3 Mutual Information

Mutual Information is a powerful feature selection method that allows for the identification of informative features by capturing the amount of information shared between a feature and the target variable. Unlike methods such as correlation that focus on linear relationships, Mutual Information is capable of handling non-linear dependencies between features, making it particularly valuable in scenarios where complex data structures and non-linear relationships exist.

The concept of Mutual Information is rooted in information theory, which quantifies the amount of information contained in a random variable. In the context of feature selection, Mutual Information measures the extent to which the knowledge of one feature can help in predicting the target variable. It evaluates the mutual dependence or information shared between the feature and the target, providing a measure of how much information about the target can be gained from a particular feature. High Mutual Information values indicate a strong relationship between the

feature and the target variable, implying that the feature contains valuable information for the classification task. On the other hand, low Mutual Information values suggest a weak or negligible association between the feature and the target variable. By ranking features based on their Mutual Information scores, this method allows for the selection of the most relevant and informative features for the classification task.

One of the significant advantages of Mutual Information is its ability to capture non-linear relationships between features. In many real-world scenarios, the relationship between variables is not strictly linear, and traditional methods may fail to identify the underlying dependencies accurately. Mutual Information overcomes this limitation by considering all types of relationships, including non-linear and complex dependencies. This makes it a valuable tool for feature selection in datasets where the relationship between features and the target variable is not well-captured by linear measures. By incorporating Mutual Information-based feature selection into the analysis of complex datasets, researchers can effectively identify features that contribute significantly to the predictive power of the model. These selected features can then be utilized in machine learning algorithms, enhancing their performance and allowing for better interpretation of the underlying relationships between features and the target variable [73].

4.1.4 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a widely used feature selection technique that not only performs dimensionality reduction but also aids in identifying the most informative features. PCA achieves this by transforming the original features into a new set of orthogonal features known as principal components.

The main objective of PCA is to capture the maximum variance in the data while reducing its dimensionality. It achieves this by projecting the data onto a new coordinate system represented by the principal components. Each principal component is a linear combination of the original features, and they are ordered in a way that the first component captures the highest variance in the data, the second component captures the second highest variance, and so on. By selecting a subset of these principal components, one can effectively retain the most significant information while reducing the feature space. The selection of principal components can be based on different criteria. One common approach is to consider the contribution of each principal component to the total variance. The cumulative explained variance plot is often used to visualize the variance captured by each component. Researchers can set a threshold, such as retaining components that explain a certain percentage (e.g., 95%) of the total variance. This allows for the selection of the principal components that capture the most relevant information in the data.

Another approach to feature selection using PCA is to consider the loadings or coefficients associated with each original feature in the principal components. The magnitude of these loadings indicates the importance of the corresponding features in constructing the principal components. Features with higher loadings contribute more to the construction of the principal components and can be considered more informative for the classification task. PCA offers several benefits for feature selection. It not only reduces the dimensionality of the dataset but also helps in identifying the features that contribute the most to the variance in the data. By selecting the principal components that explain the majority of the variance or exhibit high loadings, researchers can effectively identify the most informative features for their analysis [74].

4.1.5 Relief-F

Relief-F is a feature selection method that is specifically designed to handle noisy and high-dimensional data. It is an iterative algorithm that estimates the quality or relevance of features based on the difference in feature values between nearest neighbors. The main idea behind Relief-F is that features that have a large difference in values between instances of the same class are more likely to be informative and important for classification.

The Relief-F algorithm works as follows:

- For each instance in the dataset, Relief-F identifies the nearest instances from the same class (nearest hit) and the nearest instances from different classes (nearest miss).
- It updates the relevance score for each feature based on the differences in feature values between the current instance and its nearest hit and nearest miss instances.
- The relevance scores are adjusted based on the weights assigned to each instance, which depend on the number of hits and misses encountered during the iterative process.
- The algorithm iterates over all instances in the dataset and updates the relevance scores for each feature.
- Finally, Relief-F ranks the features based on their relevance scores, and the top-ranked features are selected for further analysis or classification.

Relief-F is particularly useful when dealing with noisy data, as it focuses on the differences in feature values between instances. By considering the instances that are closest in the feature space, Relief-F can effectively capture the relevance of features for classification even in the presence of noise. The iterative nature of Relief-F allows it to adapt to the characteristics of the data and discover the features that are most discriminative for the classification task. It takes into account

both the local structure of the data (nearest neighbors) and the class labels, making it a robust feature selection method. [75].

4.1.6 Recursive Feature Elimination

Recursive Feature Elimination (RFE): Recursive Feature Elimination (RFE) is a feature selection technique that aims to select the most important features by recursively eliminating less relevant features from a given dataset. It is an iterative process that works by training a machine learning model on the full set of features and then removing the least important feature(s) based on a specific criterion. This process is repeated until a specified number of features or a desired level of performance is achieved. RFE helps to identify the most relevant features by eliminating those that contribute the least to the model's performance. It can be particularly useful in situations where the dataset contains a large number of features, and there is a need to reduce dimensionality while maintaining or improving the model's predictive power.

4.1.7 Select from Model

Select from Model is a feature selection method that involves training a machine learning model and selecting the most important features based on the model's coefficients or feature importance. Unlike some other feature selection techniques, Select from Model incorporates the learning process of a specific model to determine feature importance. Select from Model allows you to leverage the inherent feature importance information provided by the model to make informed decisions about which features to retain. It can be applied to various models, such as linear regression, logistic regression, or decision trees, which provide feature importance scores or coefficients.

4.1.8 Select from Model based on Random Forest

Select from Model based on Random Forest is a specific implementation of the Select from Model approach that utilizes a Random Forest algorithm to estimate the feature importance. Random Forest is an ensemble learning technique that combines multiple decision trees to make predictions. It can provide reliable estimates of feature importance by analyzing the contribution of each feature across the ensemble. Random Forest considers both the individual tree's decision paths and the collective decision of the ensemble to determine feature importance. This approach can be effective in capturing complex interactions between features and identifying the most informative ones for prediction.

4.1.9 Select K-best

Select K-best is a simple feature selection technique that aims to select the K most important features from a given dataset based on a statistical measure. The K-best approach typically involves ranking the features based on a univariate statistical test or scoring method and selecting the top K features. Select K-best is useful when you want to quickly identify a fixed number of features with the highest statistical relevance to the target variable. It is commonly employed in scenarios where interpretability or simplicity is important, as it relies on univariate measures that consider each feature individually.

4.2 Image-based Feature extraction

4.2.1 Canny Edge

Canny Edge is a popular feature extraction method used in computer vision and image processing for edge detection. It is named after its inventor John Canny and is known for its effectiveness in

detecting edges in images while reducing noise sensitivity. The edges detected by Canny Edge can be valuable features for image-based classification tasks.

The Canny Edge detection algorithm consists of several stages:

i. Noise Reduction: The algorithm starts by applying a Gaussian filter to the image to reduce noise and remove unwanted details.
ii. Gradient Calculation: The next step involves calculating the gradient magnitude and direction of the image intensity. This is typically done using methods like the Sobel operator.
iii. Non-maximum Suppression: In this stage, the algorithm identifies the local maxima in the gradient magnitude to thin out the edges and suppress weak or non-maximal responses.
iv. Double Thresholding: The edges are then classified into strong, weak, and non-edges based on predefined thresholds. This helps in distinguishing significant edges from noise or weak edges.
v. Edge Tracking by Hysteresis: Finally, the algorithm performs edge tracking by connecting weak edges to strong edges, considering them as part of the same edge if they are connected. This helps to complete the edges and remove isolated weak edges.

Canny Edge is known for its ability to accurately detect edges in images while minimizing false positives and false negatives. It takes into account the local intensity gradients and provides precise edge localization. The edges detected by Canny Edge can be used as features for various image-based classification tasks, such as object recognition, image segmentation, and texture analysis [76].

4.2.2 Sobel Edges

Sobel Edges is a widely used gradient-based feature extraction method for edge detection in images. It is named after its inventor Irwin Sobel and is known for its ability to highlight edges in different directions. Sobel Edges utilizes convolutional operations with a set of filters to compute the gradient magnitude and direction of the image. The Sobel operator consists of two separate filters: one for detecting horizontal edges and the other for detecting vertical edges. These filters are convolved with the image to calculate the gradient in the respective directions. The resulting gradients are then combined to obtain the overall gradient magnitude and direction at each pixel. By applying Sobel Edges, the edges in the image are emphasized, while other areas are suppressed. This makes it easier to detect and extract important features related to the edges. The gradient magnitude represents the strength of the edges, while the gradient direction provides information about the orientation of the edges.

Sobel Edges is widely used in various image analysis and classification tasks, such as object detection, image segmentation, and feature extraction. It is particularly effective in capturing local edge information, which can be crucial for distinguishing different objects or regions in an image [77].

4.2.3 Linear Binary Pattern (LBP)

Linear Binary Pattern (LBP) is a texture-based feature extraction method widely used in computer vision. It captures local texture patterns by comparing the intensity values of pixels in an image neighborhood. LBP can effectively represent texture information and has been utilized in numerous image classification and object recognition applications [78].

These feature selection methods, both for numerical and image-based data, aid in identifying the most relevant features for classification tasks. The choice of method depends on the data characteristics and specific requirements of the analysis.

Chapter 5: Methods

5.1 Deep Learning and Image Classification

Transfer learning is one of the advantages of Deep Learning, a developing topic of study. For instance, in image classification, Transfer Learning uses characteristics that have been homed in one domain and applied to another through feature extraction and fine-tuning. Convolutional Neural Network (CNN) models have been employed with remarkable success on various comparable or different datasets, large or small, and were trained on ImageNet's million photos with 1000 categories [79]. Small datasets can benefit from these pre-trained networks because only the higher layers of these pre-trained networks need to be trained on the new datasets. This is because the lower layers of these pre-trained networks already contain many generic features like edge and colour blob detectors.

Since 1995, studies on transfer learning have gained increasing amounts of attention under various names, including "learning to learn," "lifelong learning," "knowledge transfer," "inductive transfer," "multi-task learning," "knowledge consolidation," "context-sensitive learning," "knowledge-based inductive bias," "meta learning," and "incremental/cumulative learning." Figure 5.1 [80] illustrates how conventional and transfer learning strategies have different learning processes.

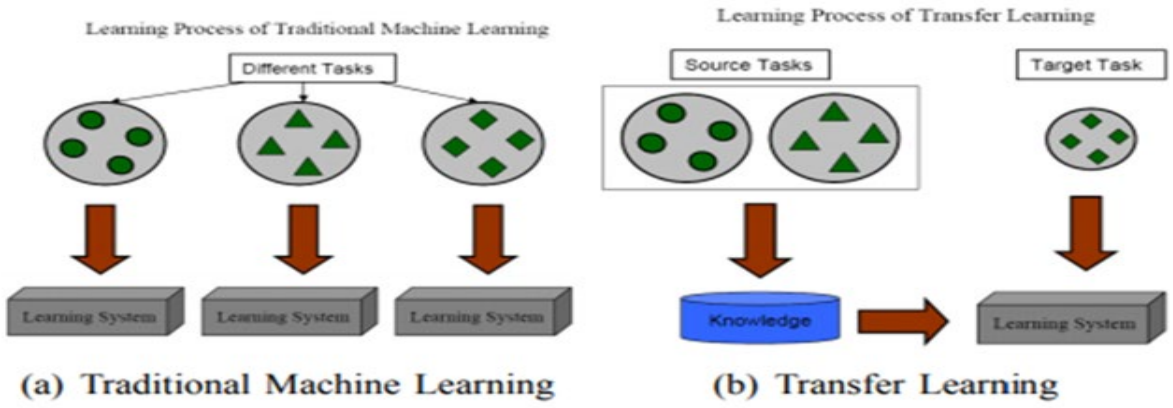

Figure 5.01 Different learning processes between traditional machine learning and Transfer Learning [80]

Deep learning is supported by several libraries. Both have advantages and disadvantages. Theano, Tensorflow, Lasagne, Caffe, Caffe2, Microsoft Cognitive Toolkit, PyTorch, and Keras are one of the most popular ones and widely used framework for the Python language and Machine Learning Development. We utilised Keras, a high-level neural networks API that can run on top of Tensorflow, Microsoft Cognitive Toolkit, or Theano. It was developed in Python. Five models that have already been trained on ImageNet are supported by the Keras Deep Learning library for Python at the moment which we have utlised in this research as well.

The implementation is divided into the following phases-

1. **Tumor Classification with Clinical Features only (No Feature Selection Methods)**
2. **Tumor Classification with Image Features only (No Feature Selection Methods)**
3. **Tumor Classification with Clinical Features and Images Features (No Feature Selection Methods)**
4. **Tumor Classification with Clinical Features only (With Feature Selection Methods)**
5. **Tumor Classification with Image Features only (With Feature Selection Methods)**

6. **Tumor Classification with Clinical Features and Images Features (With Feature Selection Methods)**
7. **Survival Analysis (Kaplan Meier Estimates etc.)**

Initially, we loaded the Clinical data using Pandas dataframe. The following models were built for Clinical Features based modelling.

1. **Fully Connected Neural Network or Back-Propagation Neural Network**
2. **Long Short-Term Memory (LSTM) Model**
3. **Convolutional Neural Network**

The following models were built for Images Features based modelling-

1. **VGG16**
2. **DenseNet**
3. **ResNet50**
4. **Xception**
5. **ResNet101**

5.2 Clinical Features based modelling

After conducting the Exploratory Data Analysis and visualization, we processed the data before modelling to create and process all the variables that are contained in the dataset. Initially, we drop the features that are not present at the time of early diagnosis from the dataset. After dropping these variables, we create Dummy Variables for Categorical features and for the numerical variables, we scaled variables from values -1 to 1. After pre-processing the dataset, we applied Principal Component Analysis (PCA) on the pre-processed dataset and selected 10 principal components as shown in Figure 5.2.

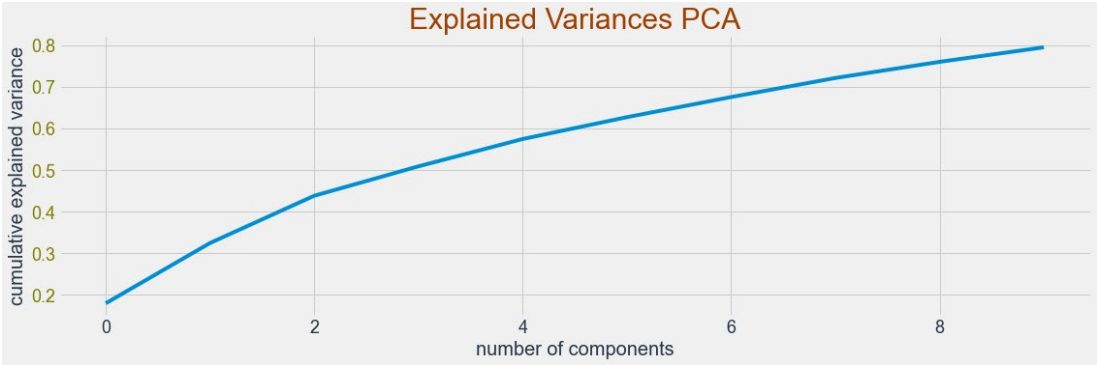

Figure 5.2 Explained Variances for PCA

After transforming the features to PCA variables, we split the dataset into train and test set. The training data had 55 observations and test data had 14 observations. We built three models for building classifier with clinical data only and described below.

5.2.1 Fully Connected Network of back propagation Neural Network (FCN/BP)

The first layer of the FCN/BP consisted of a choice of many nodes, which were 128 with Dropout choices of 0.1 and Activation choices of Relu functions. Similarly, second layer consisted of choice of many nodes, which were 32 with Dropout choices of 0.1, and Activation choices of Relu functions and followed third layer with choice of many nodes, which were 32 with Dropout choices of 0.2 and Activation choices of Relu functions. The last layer was the output layer with two nodes and sigmoid activation. The final model architecture is shown below in figure 5.3.

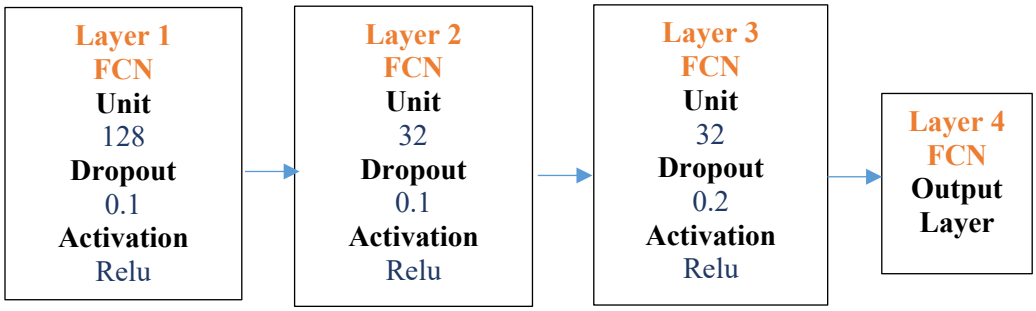

Figure 5.3 FCN/BP Architecture

5.2.2 Long Short Term Memory (LSTM)

The first layer of the LSTM consisted of a choice of many nodes which were 64 with Dropout choice 0.05 and Activation of tanh functions. Similarly, second layer consisted of choice of many nodes, which were 64 with Dropout choice 0.3, and Activation of tanh function followed by third fully connected layer with choice of many nodes which were 288 with Dropout 0.1 and Activation choice of Relu functions. The last layer was the output layer with two nodes and sigmoid activation. The final model architecture is shown in figure 5.4.

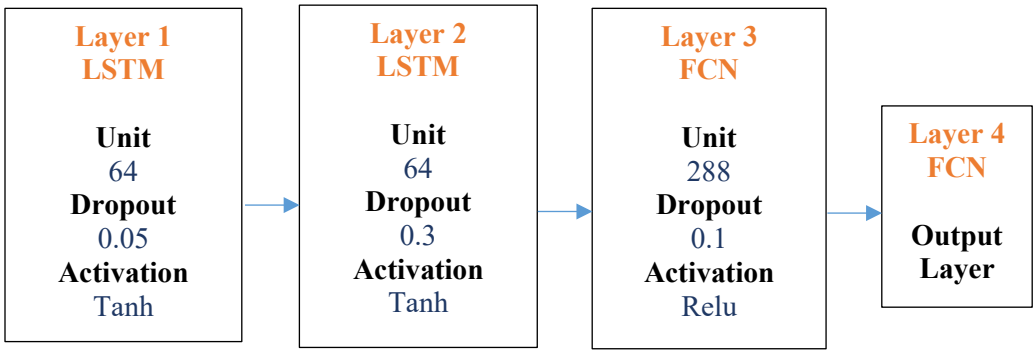

Figure 5.4 LSTM Architecture

5.2.3 Convolutional Neural Nework (CNN)

The first layer of the CNN consisted of a choice of many nodes which were 128 with Dropout choice 0.2. Similarly, second CNN layer consisted of choice of many nodes, which were 32 with Dropout choice 0.2 followed by third fully, connected layer with choice of many nodes, which were 128 with Dropout 0.2, and Activation choice of tanh function. The last layer was the output layer with two nodes and sigmoid activation. The final model architecture is in figure 5.5 below.

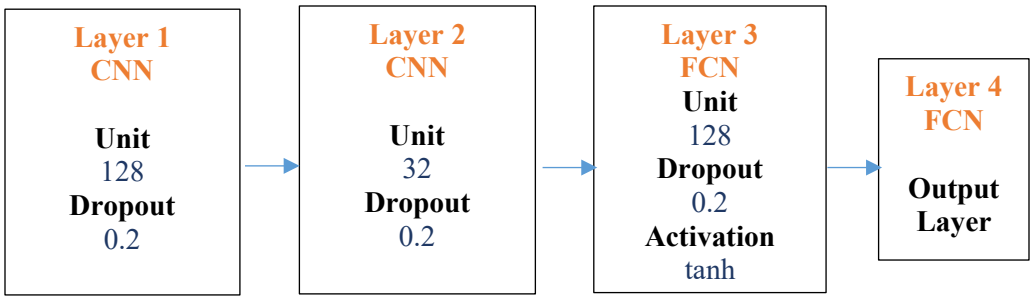

Figure 5.5 CNN Architecture

5.3 Image Based Feature Modelling

We explored the performance of various models, such as VGG16, ResNet50, DenseNet, ResNet101 and Xception. Initially, we implemented these models without applying any feature selection methods. This involved training the models on the complete set of available features. To further investigate the impact of feature selection, we incorporated three popular techniques: Canny Edge, Sobel Edges, Linear Binary Pattern. These methods aimed to improve the models' performance. The image based feature selection methods are applied to the dataset. In methods, first it reads the image and converts it to grayscale. Then calculates the image. And after that appends the calculated image to features list. After processing all the images, the features list is converted to a 3-dimensional array with RGB channels. This allows for visualization purposes and compatibility with certain downstream models that require RGB input. By utilizing the feature selection methods, we were able to extract local texture patterns from the grayscale images. These patterns can then be used as features for further classification tasks using VGG16, ResNet50, DenseNet, ResNet101, Xception model.

5.3.1 VGG16

The VGG (Visual Geometry Group) architecture is a deep convolutional neural network proposed by Simonyan and Zisserman in 2014 [81]. It is known for its simple and uniform architecture, and

achieved state-of-the-art performance on the ImageNet classification challenge in 2014 [81]. The architecture consists of a series of convolutional layers with small filters (3x3) and a fixed pooling layer (2x2) in between. The use of small filters and pooling layers helps to reduce the number of parameters and overfitting, while still achieving high accuracy.

The original VGG architecture has two variants, VGG16 and VGG19, which differ in the number of convolutional layers. VGG16 has 16 convolutional layers and three fully connected layers, while VGG19 has 19 convolutional layers and three fully connected layers. Both variants use rectified linear unit (ReLU) activation functions after each convolutional layer and dropout regularization in the fully connected layers to prevent overfitting.

The VGG architecture has inspired many subsequent CNN architectures, such as ResNet and DenseNet, and has been used in a wide range of computer vision tasks, including object detection, segmentation, and classification.

Here is the detailed architecture of VGG16:

- Input: 224x224 RGB image
- Convolutional layer 1: 64 filters, 3x3 kernel, ReLU activation
- Convolutional layer 2: 64 filters, 3x3 kernel, ReLU activation
- Max pooling layer 1: 2x2 pooling
- Convolutional layer 3: 128 filters, 3x3 kernel, ReLU activation
- Convolutional layer 4: 128 filters, 3x3 kernel, ReLU activation
- Max pooling layer 2: 2x2 pooling
- Convolutional layer 5: 256 filters, 3x3 kernel, ReLU activation
- Convolutional layer 6: 256 filters, 3x3 kernel, ReLU activation
- Convolutional layer 7: 256 filters, 3x3 kernel, ReLU activation
- Max pooling layer 3: 2x2 pooling
- Convolutional layer 8: 512 filters, 3x3 kernel, ReLU activation
- Convolutional layer 9: 512 filters, 3x3 kernel, ReLU activation
- Convolutional layer 10: 512 filters, 3x3 kernel, ReLU activation
- Max pooling layer 4: 2x2 pooling
- Convolutional layer 11: 512 filters, 3x3 kernel, ReLU activation
- Convolutional layer 12: 512 filters, 3x3 kernel, ReLU activation
- Convolutional layer 13: 512 filters, 3x3 kernel, ReLU activation

- Max pooling layer 5: 2x2 pooling
- Fully connected layer 1: 4096 units, ReLU activation, dropout regularization
- Fully connected layer 2: 4096 units, ReLU activation, dropout regularization
- Fully connected layer 3: 1000 units, softmax activation

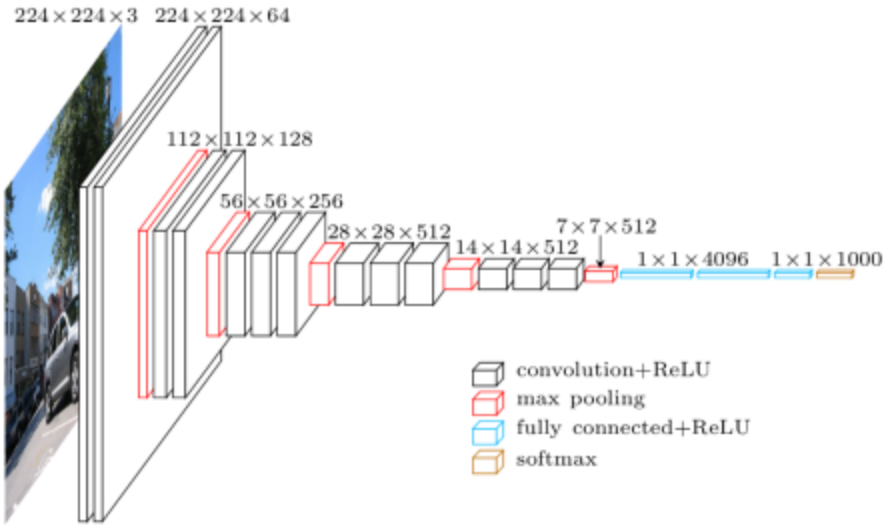

Figure 5.06 VGG Architecture in detail [81]

In Figure 5.3 shows, the architecture takes an RGB image of size 224x224 as input and outputs a probability distribution over 1000 classes. The architecture has a total of 138 million parameters, making it relatively expensive to train.

5.3.2 DenseNet121

DenseNet121 is a convolutional neural network (CNN) architecture developed by Huang et al. in 2017 [82], as an improvement over the traditional CNN models for image classification tasks. The DenseNet121 architecture is based on the idea of densely connected convolutional layers, which promotes feature reuse and reduces the number of parameters in the model.

The DenseNet121 architecture consists of 121 layers and is divided into four dense blocks, each with a different number of densely connected convolutional layers. Each dense block is followed by a transition layer that performs pooling and compression of the feature maps. The first

layer of the network is a convolutional layer followed by a max-pooling layer, and the final layer is a fully connected layer that produces the output probabilities.

The key innovation of the DenseNet architecture is the use of dense connections between layers, which connect all preceding layers to the current layer. This allows for feature reuse, as the output feature maps of each layer are concatenated with the input feature maps of all subsequent layers. This leads to a more efficient use of the network's parameters, as the same features are reused across multiple layers. Additionally, the dense connections have been shown to improve the gradient flow and reduce the vanishing gradient problem, leading to faster convergence during training. Below is summarized architecture of the DenseNet121 model.

- Input layer: 224x224 RGB image
- Convolution layer: 7x7 with stride 2, followed by Batch Normalization and ReLU activation
- Max Pooling layer: 3x3 with stride 2
- Dense block 1: 6 convolutional layers, each with Batch Normalization and ReLU activation, followed by concatenation with input and output of each layer
- Transition layer 1: 1x1 convolutional layer followed by Batch Normalization and 2x2 average pooling
- Dense block 2: 12 convolutional layers, each with Batch Normalization and ReLU activation, followed by concatenation with input and output of each layer
- Transition layer 2: 1x1 convolutional layer followed by Batch Normalization and 2x2 average pooling
- Dense block 3: 24 convolutional layers, each with Batch Normalization and ReLU activation, followed by concatenation with input and output of each layer
- Transition layer 3: 1x1 convolutional layer followed by Batch Normalization and 2x2 average pooling
- Dense block 4: 16 convolutional layers, each with Batch Normalization and ReLU activation, followed by concatenation with input and output of each layer
- Global average pooling layer: Computes the average of each feature map
- Fully connected layer: Produces output probabilities for each class

The DenseNet121 architecture has been extensively tested on various image classification datasets, such as ImageNet, CIFAR-10, and CIFAR-100, and has achieved state-of-the-art performance in terms of accuracy and parameter efficiency. The architecture has also been used

for other computer vision tasks, such as object detection and semantic segmentation. The figure 5.4 shows the overall architecture of DenseNet121.

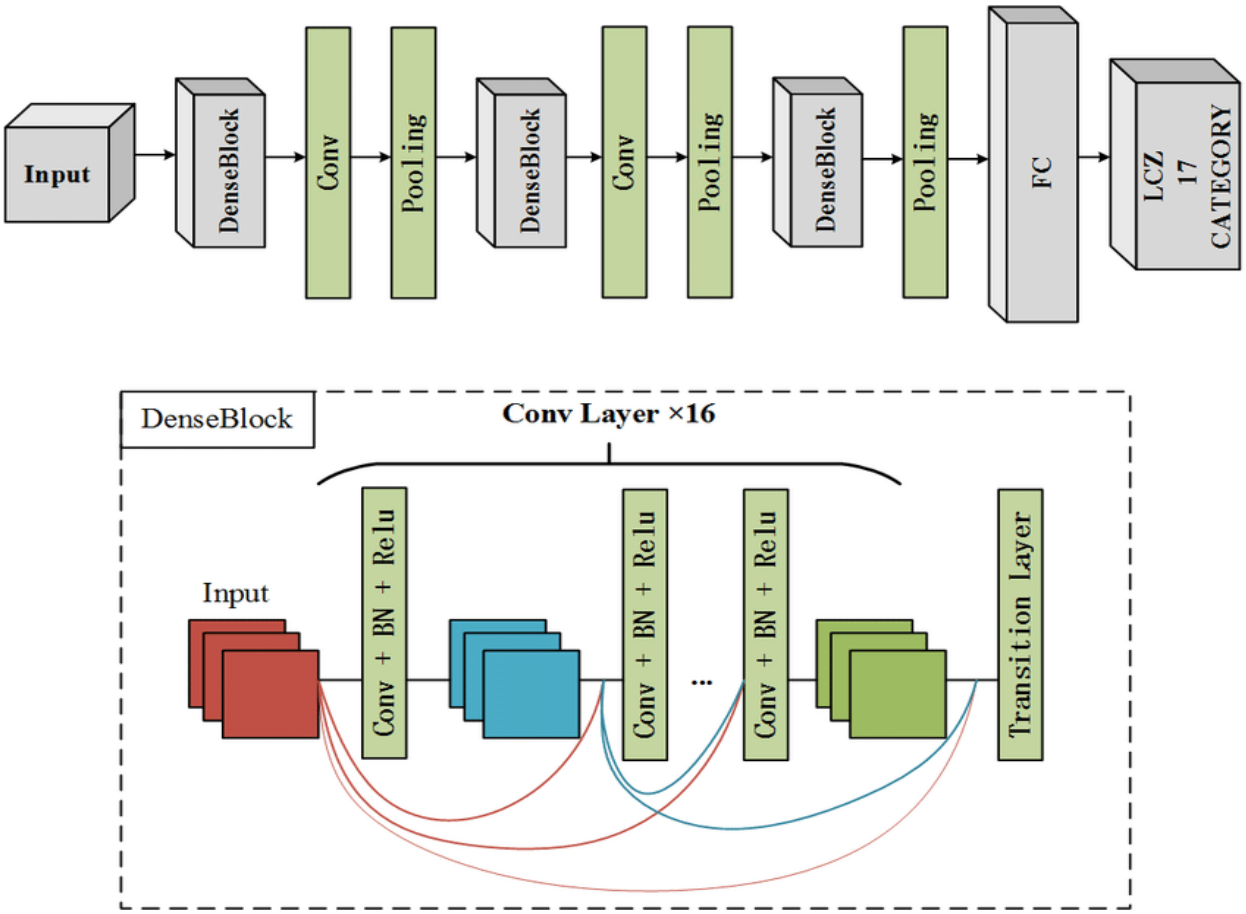

Figure 5.07 Architecture of DenseNet121 [83]

5.3.3 ResNet50

ResNet50 is a convolutional neural network architecture proposed by Kaiming He et al. in their paper "Deep Residual Learning for Image Recognition" published in 2016. A residual learning architecture developed by Microsoft Research that makes deep network training simpler. Instead of learning unreferenced functions, it reformulates the layers to learn residual functions with

reference to the layer inputs. It is a variant of the ResNet architecture that consists of 50 layers, and it achieved state-of-the-art results on the ImageNet classification task [84].

Here are the architecture configurations and layers of ResNet50:

- Input layer: The input to ResNet50 is a 224 x 224 x 3 RGB image.
- Convolutional layer: The first layer is a convolutional layer with 64 filters of size 7 x 7 and stride 2, followed by batch normalization and ReLU activation.
- Max pooling layer: The output of the first layer is passed through a max pooling layer with pool size 3 x 3 and stride 2.
- Residual blocks: ResNet50 has 16 residual blocks, each consisting of two convolutional layers and a skip connection. The first convolutional layer has 64 filters of size 1 x 1, followed by batch normalization and ReLU activation. The second convolutional layer has 64 filters of size 3 x 3, followed by batch normalization and ReLU activation. The skip connection bypasses the two convolutional layers and directly adds the input to the output of the second convolutional layer.
- The first two residual blocks have only one convolutional layer each, with 64 filters of size 1 x 1 and 64 filters of size 3 x 3, respectively.
- The next three residual blocks have two convolutional layers each, with 128 filters of size 1 x 1 and 128 filters of size 3 x 3, respectively.
- The next four residual blocks have two convolutional layers each, with 256 filters of size 1 x 1 and 256 filters of size 3 x 3, respectively.
- The next five residual blocks have two convolutional layers each, with 512 filters of size 1 x 1 and 512 filters of size 3 x 3, respectively.
- Global average pooling layer: The output of the last residual block is passed through a global average pooling layer that computes the average value of each feature map across the spatial dimensions.
- Fully connected layer: The output of the global average pooling layer is flattened and passed through a fully connected layer with 1000 output units, one for each class in the ImageNet dataset.
- Softmax activation layer: The output of the fully connected layer is passed through a softmax activation layer to obtain class probabilities.

ResNet50 has demonstrated superior performance on a variety of computer vision tasks such as object detection, image segmentation, and action recognition. Its success can be attributed to the residual blocks, which allow for the efficient training of very deep neural networks by addressing the vanishing gradient problem [84]. Given Figure 5.5 is the diagram of the ResNet50 architecture.

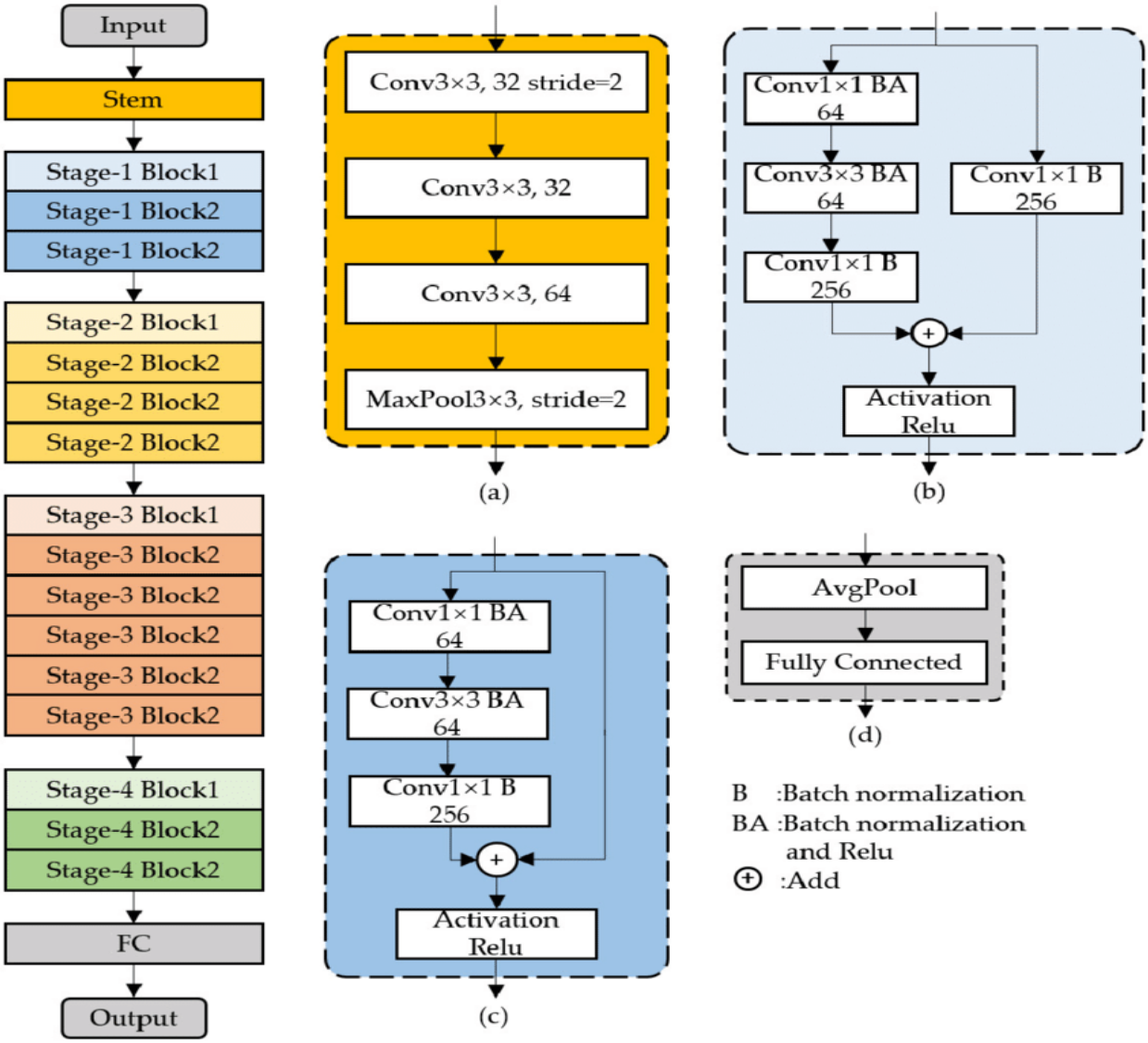

Figure 5.08 Architecture of ResNet50 [85]

5.3.4 ResNet101

ResNet101 is a deep convolutional neural network architecture that was introduced by Kaiming He et al. in 2016. It is an extension of the ResNet50 architecture, with 101 layers, and achieved state-of-the-art results on the ImageNet classification task.

Here is a detailed description of the ResNet101 architecture:

- Input layer: The input to the ResNet101 model is a 224 x 224 x 3 RGB image.
- Convolutional layer: The first layer of ResNet101 is a convolutional layer with 64 filters of size 7 x 7 and a stride of 2, followed by batch normalization and ReLU activation.
- Max pooling layer: The output of the first convolutional layer is passed through a max pooling layer with a pool size of 3 x 3 and a stride of 2.
- Residual blocks: ResNet101 has 33 residual blocks, with each block consisting of several layers.
- The first residual block includes three convolutional layers, each followed by batch normalization and ReLU activation. The first two convolutional layers have 64 filters of size 1 x 1, while the third has 256 filters of size 3 x 3. The skip connection in this block bypasses the three convolutional layers and directly adds the input to the output of the third convolutional layer.
- The next three residual blocks each have four convolutional layers with 64 filters of size 1 x 1 and 64 filters of size 3 x 3, respectively. Each block also includes a skip connection that adds the input to the output of the last convolutional layer.
- The next four residual blocks each have 23 convolutional layers with 256 filters of size 1 x 1 and 256 filters of size 3 x 3, respectively. Each block also includes a skip connection.
- The final three residual blocks each have three convolutional layers with 512 filters of size 1 x 1 and 512 filters of size 3 x 3, respectively. Each block also includes a skip connection.
- Global average pooling layer: The output of the final residual block is passed through a global average pooling layer that computes the average value of each feature map across the spatial dimensions.
- Fully connected layer: The output of the global average pooling layer is flattened and passed through a fully connected layer with 1000 output units, one for each class in the ImageNet dataset.
- Softmax activation layer: The output of the fully connected layer is passed through a softmax activation layer to obtain class probabilities.

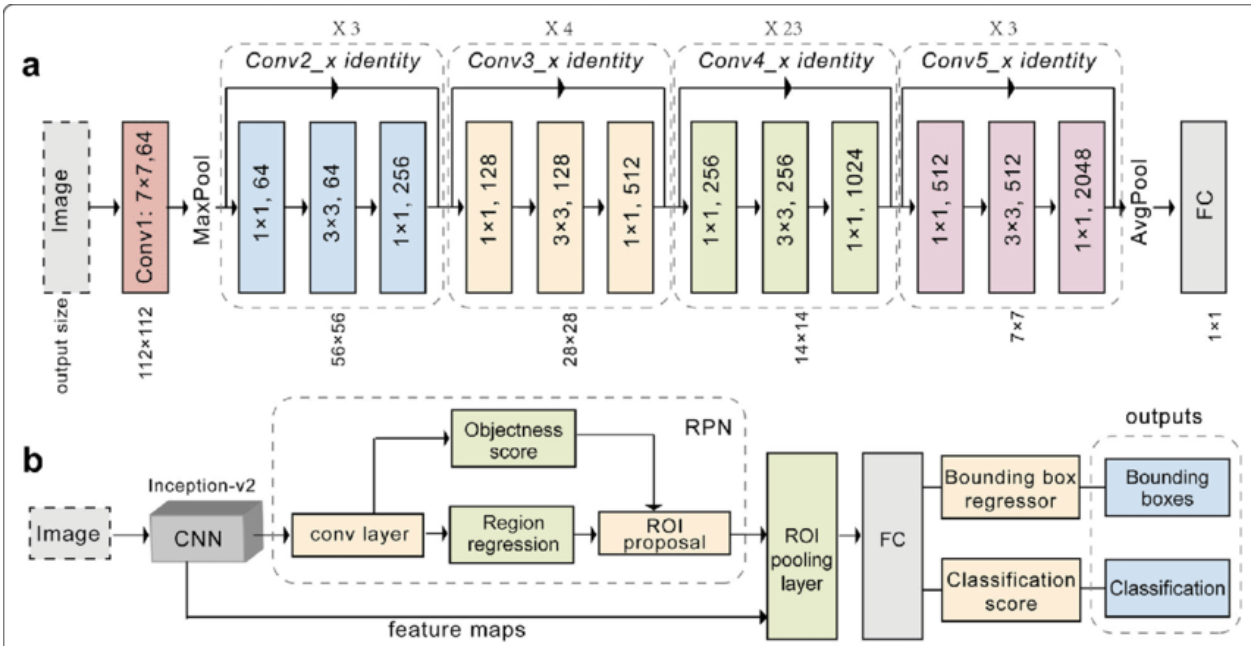

Figure 5.09 Architecture of ResNet101 [86]

Given Figure 5.6 is the diagram of the ResNet101 architecture. ResNet101 has demonstrated superior performance on a variety of computer vision tasks such as object detection, image segmentation, and action recognition. Its success can be attributed to the residual blocks, which allow for the efficient training of very deep neural networks by addressing the vanishing gradient problem.

ResNet50 and ResNet101 are deep neural network architectures proposed in 2016, with ResNet101 being deeper than ResNet50. ResNet101 has 101 layers and 33 residual blocks, while ResNet50 has 50 layers and 16 residual blocks. The increased depth in ResNet101 allows it to capture more complex features, but it also makes it more computationally expensive, challenging to train and deploy on resource-constrained devices. The choice between ResNet50 and ResNet101 depends on the specific needs of the application and the available computational resources for training and inference. Different variants of ResNet architecture with varying depths and residual blocks exist, offering a trade-off between performance and computational cost [87, 88, and 89].

5.3.5 Xception

Xception is a deep convolutional neural network architecture proposed by François Chollet in 2016 [90], with the aim of improving the performance and efficiency of traditional convolutional neural networks. The name "Xception" is derived from "Extreme Inception," as the architecture is based on the Inception modules proposed by Szegedy et al. in 2015 [91], but with a novel approach to depthwise separable convolutions.

Xception uses depthwise separable convolutions, which split the standard convolution operation into a depthwise convolution (which applies a single filter to each input channel separately) and a pointwise convolution (which applies a 1x1 filter to combine the outputs of the depthwise convolution). This reduces the number of parameters in the model and allows for better modeling of spatial correlations within the input data.

In addition to depthwise separable convolutions, Xception also includes a novel approach to skip connections, which connect different layers in the network without adding additional parameters. This helps to improve the gradient flow and avoid the vanishing gradient problem during training.

Xception has shown impressive results on several image classification benchmarks, including the ImageNet Large Scale Visual Recognition Challenge, where it achieved state-of-the-art performance with a top-1 error rate of 0.219 and a top-5 error rate of 0.064.

Here is the Xception architecture configuration in bullets points:

- Input image size: 299x299x3
- Initial convolution: 3x3, 32 filters, stride 2
- Stack of 36 convolutional layers, each with 3x3 depthwise separable convolutions, followed by batch normalization and ReLU activation
- 8 modules, each consisting of a stack of convolutional layers and a skip connection
- Global average pooling
- Fully connected layer with SoftMax activation for classification

Given Figure 5.7 is the diagram of the Xception architecture.

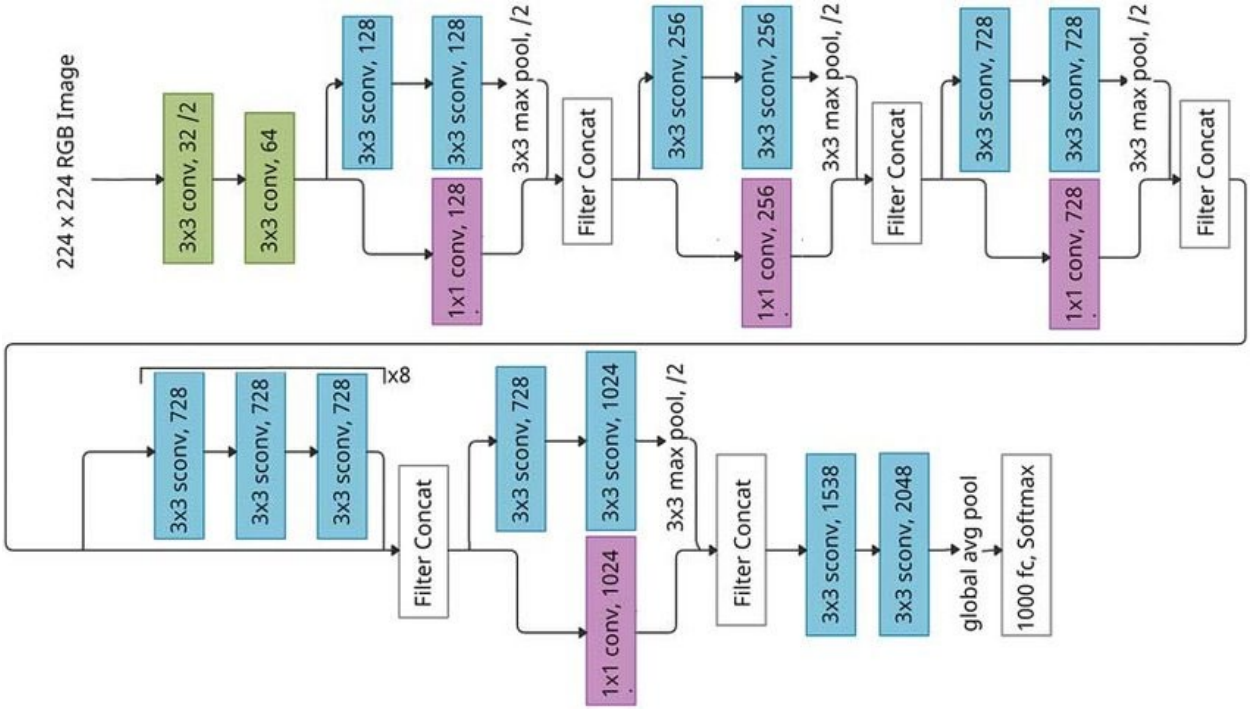

Figure 5.010 Architecture of Xception [92]

5.4 Image Feature and Clinical Feature based modelling

Data was loaded from each patient folder using glob API, and the file location was stored in a data frame. We converted the images into the NumPy array and stored them into pandas' data frame against identifiers to merge the features with the clinical features. Below output shows the dimensions of the Images which shows 69 number of observations or total number of patients, images of 128x128 pixels with 3 channels (RGB). The PCA applied Clinical features have 69 observations or total number of patients with 10 principal components features.

- Shape of Images Features: (69, 128, 128, 3).
- Shape of PCA applied Clinical Features: (69, 10).
- Shape of Labels: (69,) which consists of labels which is 0 or 1.

The shape of labels is a 1-dimensional array, which consists of labels as either 0s or 1s. Thereafter, the data was split into training and test sets. The training set consisted of 55 number of observations and test set consisted of 14 number of observations.

- Training Images Features Shape: (55, 128, 128, 3)
- Testing Images Features Shape: (14, 128, 128, 3)
- Training Clinical Features Shape: (55, 10)
- Testing Clinical Features Shape: (14, 10)
- Training Labels Shape: (55,)
- Testing Labels Shape: (14,)

After pre-processing and preparing the data, we performed modelling. We selected five models that are as follows: VGG16, DenseNet12, ResNet50, Xception, ResNet101.

5.4.1 VGG16-FCN

The results below shows the VGG16 model experiments. We conducted 100 experiments for each model. The following is the flow diagram of VGG16 –FCN Model. The chosen parameters are highlighted in the figure 5.11 below.

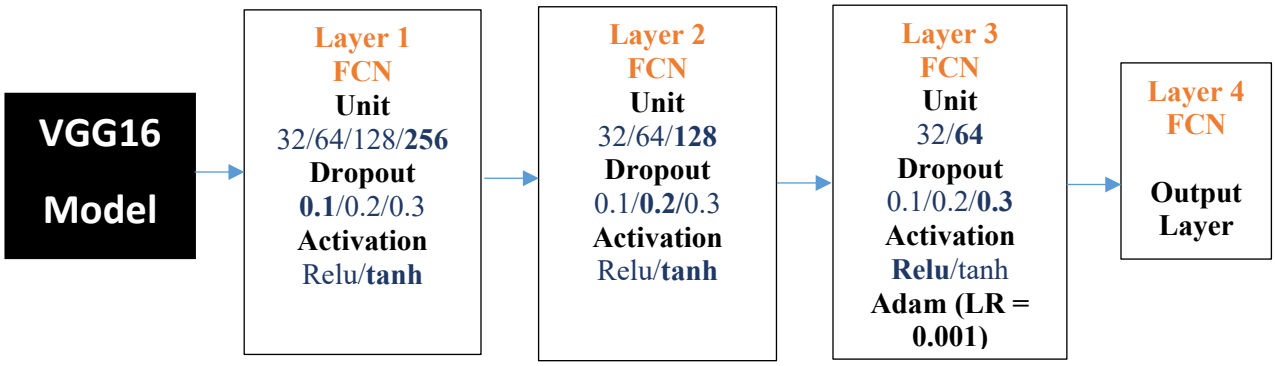

Figure 5.11 VGG16-FCN Architecture

We can see that with VGG16-FCN architecture, the model has three hidden layers with an output layer. In the best and optimized model, we had first fully connected layer (FCN), which was found to have 256 units with a dropout 0.1 and tanh activation layer. The second fully connected layer was found to have 128 units with a dropout 0.2 and tanh activation function. The third fully connected layer was found to have 64 units with a dropout of 0.3 and relu activation function with Adam Optimizer and a learning rate of 0.001.

5.4.2 DenseNet121-FCN

The results below shows the DenseNet121 model experiments. We conducted 100 experiments for each model. The following is the flow diagram of DenseNet121 –FCN Model. The chosen parameters are highlighted in the figure 5.12 below.

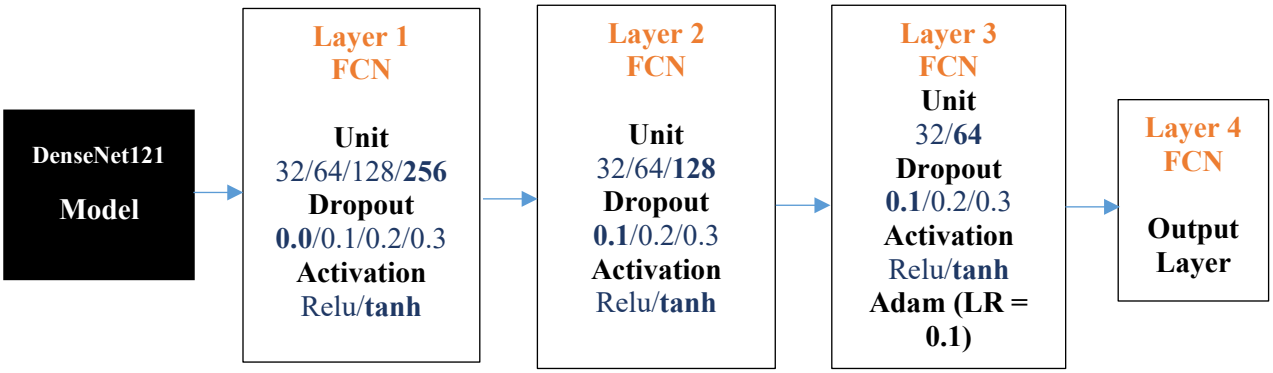

Figure 5.12 DenseNet121-FCN Architecture

We can see that with DenseNet121-FCN architecture, the model has three hidden layers with an output layer. In the best and optimized model, we had first fully connected layer (FCN), which was found to have 256 units with a dropout 0.0 and tanh activation layer. The second fully connected layer was found to have 128 units with a dropout 0.1 and tanh activation function. The third fully connected layer was found to have 64 units with a dropout of 0.1 and tanh activation function with Adam Optimiser and a learning rate of 0.1.

- The two transfer learning models that we built above VGG16 and DenseNet121 are not the same. There are some differences in the configuration of the layers and the hyperparameters used in the two models. Here are the key differences:
- The dropout rates are different in the first fully connected layer of both models. In the first model, the dropout rate is 0.1, whereas in the second model, it is 0.0.
- The dropout rate in the second fully connected layer is different in both models. In the first model, the dropout rate is 0.2, while in the second model, it is 0.1.
- The activation function used in the third fully connected layer is different in both models. In the first model, it is relu, while in the second model, it is tanh.
- The learning rate is different in both models. In the first model, it is 0.001, while in the second model, it is 0.1.

These differences can have a significant impact on the performance of the model. Therefore, it's essential to carefully tune the hyperparameters and the configuration of the layers to achieve the best possible results.

5.4.3 ResNet50-FCN

The results below shows the ResNet50 model experiments. We conducted 100 experiments for each model. The following is the flow diagram of ResNet50 –FCN Model. The chosen parameters are highlighted in the diagram below.

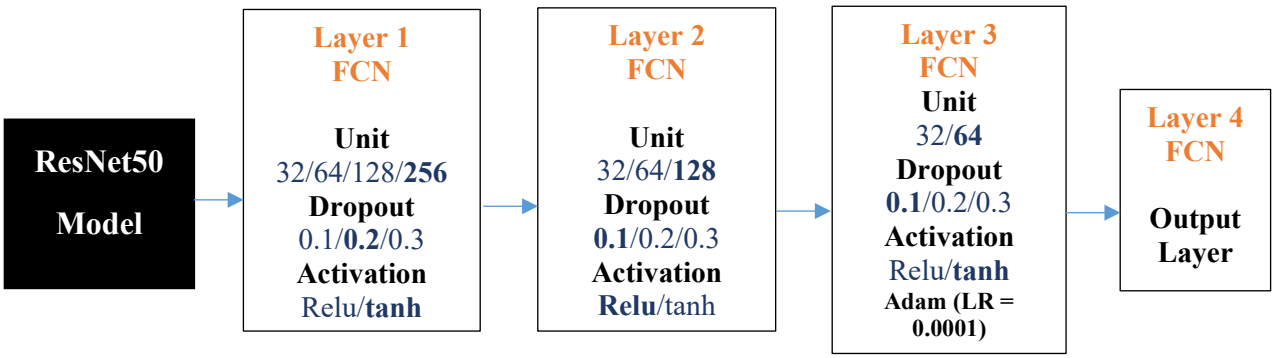

Figure 5.13 ResNet50-FCN Architecture

We can see that with ResNet50-FCN architecture in the figure 5.13, the model has three hidden layers with an output layer. In the best and optimized model, we had first fully connected layer (FCN), which was found to have 256 units with a dropout 0.2 and tanh activation layer. The second fully connected layer was found to have 128 units with a dropout 0.1 and relu activation function. The third fully connected layer was found to have 64 units with a dropout of 0.1 and tanh activation function with Adam Optimiser and a learning rate of 0.0001.

5.4.4 Xception-FCN

The results below shows the Xception model experiments. We conducted 100 experiments for each model. The following is the flow diagram of Xception –FCN Model.

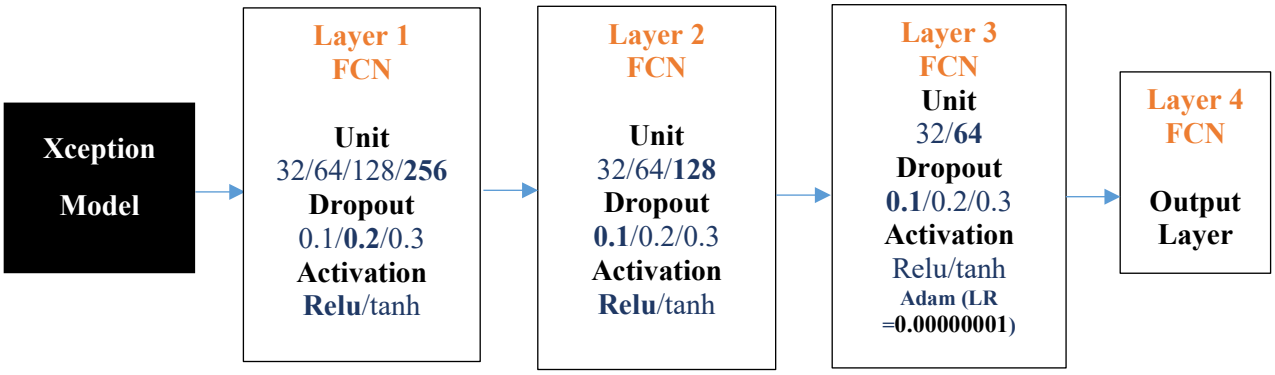

Figure 5.14 Xception-FCN Architecture

We can see that with Xception-FCN architecture in the figure 5.14, the model has three hidden layers with an output layer. In the best and optimized model, we had first fully connected layer which was found to have 256 units with a dropout 0.2 and relu activation layer. The second fully connected layer was found to have 128 units with a dropout 0.1 and relu activation function. The third fully connected layer was found to have 64 units with a dropout of 0.1 and relu activation function with Adam Optimiser and a learning rate of 0.00000001. The traning loss was 0.74 and 0.64.

5.4.5 ResNet101-FCN

The results below shows the ResNet101 model experiments. We conducted 100 experiments for each model. The following is the flow diagram of ResNet101 –FCN Model.

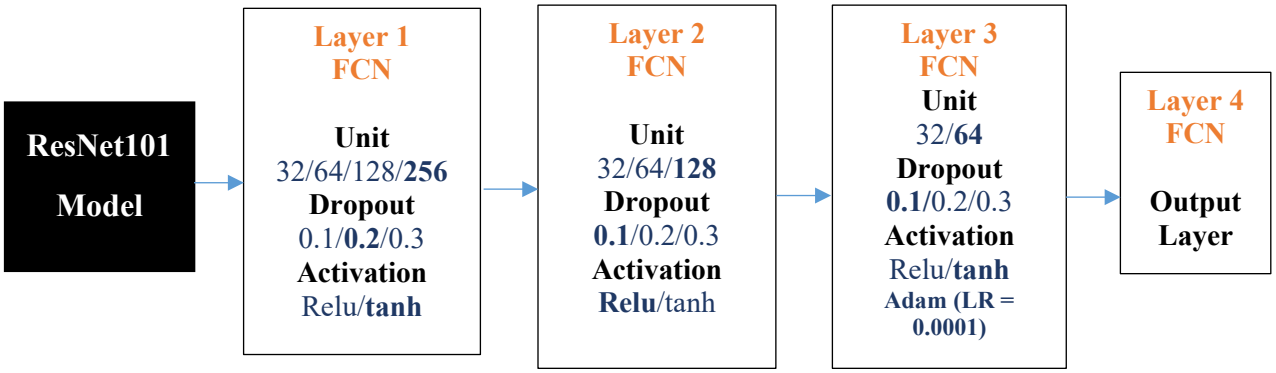

Figure 5.15 ResNet101-FCN Architecture

We split the training data into training data and validation at a ratio of 70:30 for the training and validation set, the final model is chosen based on the validation accuracy. We can see that with ResNet151-FCN architecture in the figure 5.15, the final model has three hidden layers with an output layer. The first fully connected layer was found to have 256 units with a dropout 0.2 and tanh activation layer. The second fully connected layer was found to have 128 units with a dropout 0.1 and relu activation function. The third fully connected layer was found to have 64 units with a dropout of 0.1 and tanh activation function with Adam Optimiser and a learning rate of 0.0001.

Table 5.1 Hyperparameters for five different models

Model Architecture	Hidden Layers	First FCN Units	First FCN Activation	First FCN Dropout	Second FCN Units	Second FCN Activation	Second FCN Dropout	Third FCN Units	Third FCN Activation	Third FCN Dropout	Optimizer	Learning Rate
VGG16	3	256	Tanh	0.1	128	Tanh	0.2	64	ReLU	0.3	Adam	0.001
DenseNet121	3	256	Tanh	0.0	128	Tanh	0.1	64	Tanh	0.1	Adam	0.1
ResNet50	3	256	Tanh	0.2	128	ReLU	0.1	64	Tanh	0.1	Adam	0.0001
Xception	3	256	ReLU	0.2	128	ReLU	0.1	64	ReLU	0.1	Adam	0.00000001
ResNet151	3	256	Tanh	0.2	128	ReLU	0.1	64	Tanh	0.1	Adam	0.0001

The above table 5.1 show the hyperparameters for five different models that were trained and optimized for image classification tasks. The models used were VGG16-FCN, DenseNet121-FCN, ResNet50-FCN, Xception-FCN, and ResNet151-FCN.

Starting with VGG16-FCN architecture, the best and optimized model has three fully connected hidden layers and an output layer. The first fully connected layer had 256 units, a dropout of 0.1, and a tanh activation function. The second fully connected layer had 128 units, a dropout of 0.2, and a tanh activation function. The third fully connected layer had 64 units, a dropout of 0.3, and a relu activation function. The optimizer used was Adam with a learning rate of 0.001. These hyperparameters have been tuned to give the best accuracy for the VGG16-FCN architecture.

Moving on to DenseNet121-FCN architecture, the best and optimized model also has three fully connected hidden layers and an output layer. The first fully connected layer had 256 units, a dropout of 0.0, and a tanh activation function. The second fully connected layer had 128 units, a dropout of 0.1, and a tanh activation function. The third fully connected layer had 64 units, a dropout of 0.1, and a tanh activation function. The optimizer used was Adam with a learning rate of 0.1. These hyperparameters were also found to give the best accuracy for the DenseNet121-FCN architecture.

For the ResNet50-FCN architecture, the best and optimized model also has three fully connected hidden layers and an output layer. The first fully connected layer had 256 units, a dropout of 0.2, and a tanh activation function. The second fully connected layer had 128 units, a dropout of 0.1, and a relu activation function. The third fully connected layer had 64 units, a dropout of 0.1, and a tanh activation function. The optimizer used was Adam with a learning rate of 0.0001. These hyperparameters were determined to give the best accuracy for the ResNet50-FCN architecture.

For the Xception-FCN architecture, the best and optimized model also has three fully connected hidden layers and an output layer. The first fully connected layer had 256 units, a dropout of 0.2, and a relu activation function. The second fully connected layer had 128 units, a dropout of 0.1, and a relu activation function. The third fully connected layer had 64 units, a dropout of 0.1, and a relu activation function. The optimizer used was Adam with a learning rate of 0.00000001. Interestingly, the training loss was reported to be 0.74 and 0.64, which is quite low.

Finally, for ResNet151-FCN architecture, the best and optimized model also has three fully connected hidden layers and an output layer. The first fully connected layer had 256 units, a dropout of 0.2, and a tanh activation function. The second fully connected layer had 128 units, a

dropout of 0.1, and a relu activation function. The third fully connected layer had 64 units, a dropout of 0.1, and a tanh activation function. The optimizer used was Adam with a learning rate of 0.0001.

In summary, each of the five models had slightly different hyperparameters that were found to give the best accuracy for that specific architecture. However, there are some commonalities such as the use.

5.5 Survival Analysis by Kaplan Meier Method

A collection of statistical operations for data analysis using time to an event happens as the outcome variable of interest is known as a survival analysis. The occurrences might include things like death, illness, client attrition, and recuperation. Simply said, it is used to determine how long a population under investigation will live. It is also known as "Time to Event" analysis since its objective is to determine how long it will take a certain person or group of people to experience a particular event of interest.

The duration of surviving from a given date through the moment of death, failure, or other critical events is measured by survival analysis. This non-parametric statistic, sometimes referred to as the product-limit estimator, is used to calculate the survival function using lifetime data.

It can be used, for instance, to determine:

- Survival time following therapy in medical procedures.
- How long do people remain jobless after losing their jobs?
- How long does it take for couples receiving fertility therapy to conceive?
- The time until a machine part fails.

Given a sufficient sample size, the Kaplan Meier estimator's graph resembles the real survival function for that population as a succession of diminishing horizontal steps. Due to its

alleged simplicity of use, the Kaplan-Meier estimate is often utilized. The probability of surviving over a certain period while considering multiple small periods of time into account is known as the Kaplan-Meier survival curve. Figure 5.8 shows an example of a Kaplan-Meier estimate.

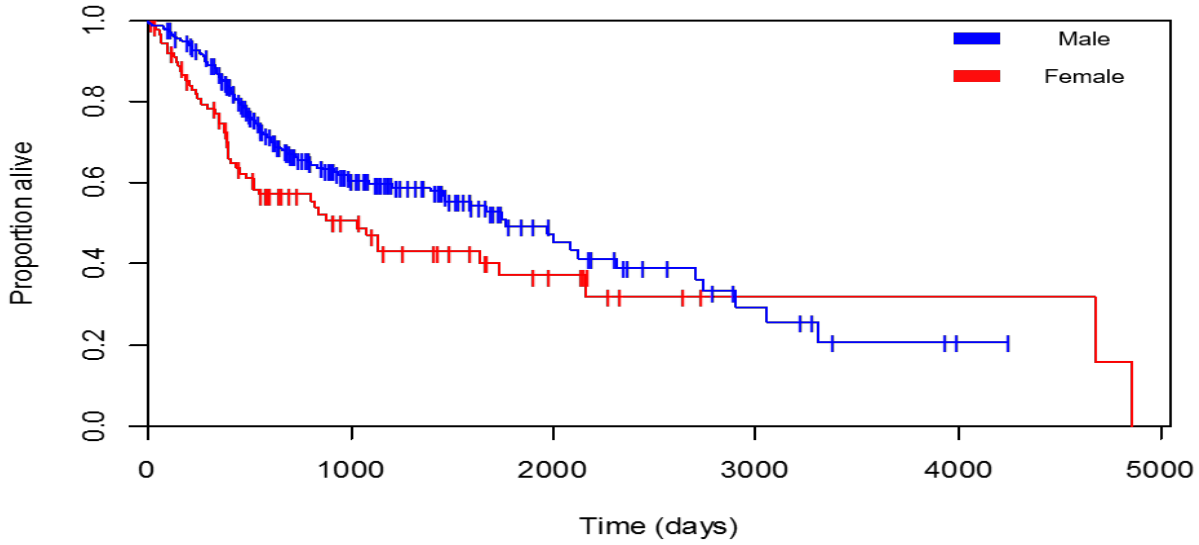

Figure 5.016 Kaplan Meier Estimate [93]

This study is predicated on three assumptions. First, we assume that patients who are restricted at any given moment have the same chances of survival as those who are still being monitored. Second, we assume that participants enrolled both early and late in the research have equal chances of surviving. Thirdly, we assume that the event occurs at the allotted time. When the occurrence is discovered during a routine test, this might lead to issues in particular situations. We only know that the incident took place in between two tests. The estimated survival can be computed more precisely by following up with the individuals more regularly and for shorter periods of time, ideally for one day, if the precision of the recording allows (maximum). The term "product limit estimate" is another name for the Kaplan-Meier estimate. It entails calculating the likelihood that an event will occur at a specific time. To obtain the final estimate, we multiply

these subsequent probabilities by any previously estimated probabilities. The following formula may be used to calculate the likelihood of surviving at any given time:

$$S(t) = \exp\left(-H(t)\right)$$

Where $S(t)$ is the survival probability at time t and $H(t)$ is the cumulative hazard function up to time t.

The Kaplan-Meier estimator is a non-parametric method of estimating and visualising the survival probability as a function of time. It is also known as the product-limit estimator. Given that it is the most straightforward method and calls for the fewest assumptions, it is frequently the initial step in the survival analysis process. We use the following assumptions in order to do the analysis using the Kaplan-Meier method:

- The survival probability of all observations is the same, regardless of when they joined the research.
- The event of interest is unambiguous and occurs at a clearly defined time.
- The chances of censored observations surviving are no different from those of followed observations.

We are hardly aware of the genuine survival function in practical situations. In order to approximate the real survival function from the gathered data, we use the Kaplan-Meier estimator. The estimate is given by the following equation and is defined as the percentage of observations that survived for a specific period under identical conditions:

$$\check{S}(t) = \prod_{i:t_i \leq t} (1 - d_i/n_i)$$

Where:

- $\check{S}(t)$is the estimated survival probability at time t
- Π is the product symbol, which indicates the multiplication of each individual probability
- $t_i$is a time when at least one event happened,

- d_i is the number of events (e.g. deaths) that occurred before time t_i,
- n_i represents the number of individuals known to have survived up to time t_i, (they have not yet had the death event or have been censored). To put it differently, the number of observations at risk at time t_i.

The percentage likelihood of survival at time t multiplied by each previous time results in the survival probability at time t. The product symbol in the formula represents the multiplication of the percentage likelihood of survival at each previous time,

The Kaplan-Meier estimator, also known as the product-limit estimator, is a widely used statistical method for analyzing survival data. The estimator is based on the concept of calculating the likelihood of surviving at any given time.

The Kaplan-Meier estimator produces a survival function that shows the probability of surviving beyond a certain time point. The survival function is calculated by taking the product of the proportion of individuals who survived up to each time point.

The formula for the Kaplan-Meier estimator involves a product symbol, which multiplies the proportion of individuals who survived at each time point. This product symbol represents the cumulative survival probability up to each time point.

The Kaplan-Meier estimator is particularly useful in medical research, where it can be used to estimate survival probabilities for patients with a certain disease or condition. By analyzing the survival data, researchers can identify factors that may affect the likelihood of survival and make more informed treatment decisions. Overall, the Kaplan-Meier estimator is a powerful tool for analyzing survival data and is widely used in many fields, including medicine, epidemiology, and engineering.

The Kaplan-Meier curves, which represent a plot of the Kaplan-Meier estimator with time, are what we most frequently connect with this method of survival analysis and what we typically

observe in practice. These curves may be used as an exploratory tool to examine the survival functions of cohorts, populations that got treatments or not, behavioral clusters, etc. The survival line is actually a series of decreasing horizontal steps, which approach the shape of the population's true survival function given a large enough sample size. In practice, the plot is often accompanied by confidence intervals, to show how uncertain we are about the point estimates — wide confidence intervals indicate high uncertainty, probably due to the study containing only a few participants — caused by both observations dying and being censored.

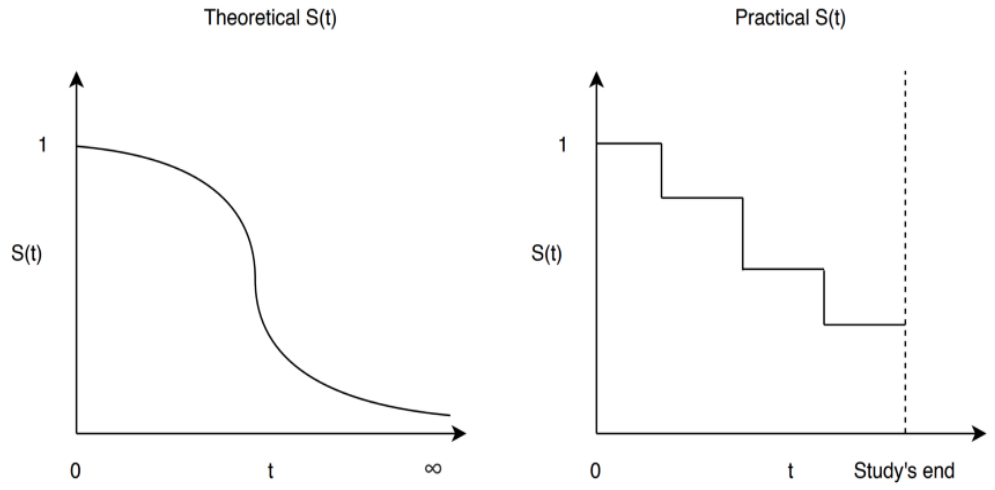

Figure 5.017 Kaplan Meier Curve - Theoretical versus Practical [94]

In Figure 5.9, The Kaplan-Meier curve is a graphical representation of the survival function, which describes the probability of survival over time. It is widely used in medical research, engineering, and other fields to analyze the survival or failure times of subjects. In theory, the Kaplan-Meier curve is calculated by dividing the time axis into small intervals, or time intervals, and then calculating the proportion of subjects surviving at each interval. The survival probability at each interval is calculated by multiplying the proportion of subjects surviving in the previous interval by the proportion surviving in the current interval. This is known as the product-limit estimator. However, in practice, it is not always possible to calculate the Kaplan-Meier curve exactly as

described above. This is because there may be missing data or censoring in the dataset, where the exact failure time is unknown for some subjects. In such cases, the Kaplan-Meier curve is estimated using the maximum likelihood method or other statistical methods, which take into account the censoring and missing data. The Kaplan-Meier curve is often used to compare survival rates between different groups of subjects, such as different treatments or patient populations. It can also be used to estimate the median survival time, which is the time at which 50% of the subjects have failed. Overall, the Kaplan-Meier curve is a powerful tool for analyzing survival data, both in theory and practice. It allows researchers to gain insights into the probability of survival over time and to compare survival rates between different groups of subjects [94].

5.6 Hardware Details

Training Deep Learning networks require tremendous processing power to handle multiple matrix multiplications. GPUs are ideal for performing these operations. In this project, we used Nvidia GTX 1050 GPU to train 8X faster than an i7 Intel CPU running at 3.5 GHz. This indicates that the use of a GPU can significantly speed up the training process compared to a CPU. GPUs are specifically designed for parallel processing and have a large number of cores that can work together simultaneously to perform complex calculations. On the other hand, CPUs have fewer cores and are better suited for sequential processing tasks.

Deep learning models involve performing large numbers of matrix multiplications, which can be parallelized and accelerated by using GPUs. The high parallel processing power of GPUs allows for a much faster execution time for deep learning tasks.

In the case of the Nvidia GTX 1050 GPU, it has 640 CUDA cores, which can perform parallel computations. Additionally, it has a memory bandwidth of 112 GB/s, which allows for efficient transfer of data between the GPU and the CPU.

Several studies have shown that the use of GPUs can significantly speed up the training process for deep learning models. For example, a study by Coates et al. in 2013 found that using GPUs could speed up the training of convolutional neural networks (CNNs) by a factor of 10-20 compared to using CPUs [95].

In summary, the use of a GPU, such as the Nvidia GTX 1050, can greatly accelerate the training process of deep learning models compared to a CPU. This can be particularly beneficial for large datasets or complex models, where the training process can take a significant amount of time. In my specific implementation, the computational time for running models varies depending on the type of data. For clinical data, the model required approximately 5 hours to complete its training. On the other hand, when working with image-based data, each model took around 4 hours to execute.

Chapter 6: Results and Analysis

In this chapter, the results and outputs are discussed.

6.1 Modelling Results

6.1.1 Tumor Classification with Clinical Features only (Without Feature Selection Methods)

The Table 6.1 shows the results of tumor classification using clinical features without any feature selection methods. Three different models were used: FCN, LSTM, and CNN. The accuracy score for the FCN and CNN models is 86.0%, while the LSTM model achieved a higher accuracy score of 93.0%.

For precision score, the LSTM model achieved the highest score of 0.94, followed by the FCN and CNN models which both achieved a score of 0.898. The recall score, which measures the ability of the models to identify positive cases correctly, is highest for the LSTM model with a score of 0.93, followed by the FCN and CNN models which both achieved a score of 0.86.

The F1 score, which is a harmonic mean of precision and recall, is also highest for the LSTM model with a score of 0.93. The FCN and CNN models achieved an F1 score of 0.86.

Overall, the LSTM model performed the best in terms of accuracy, precision, recall, and F1 score. This may be due to its ability to capture temporal dependencies in the data. However, it is important to note that these results were obtained without using any feature selection methods, which may have an impact on the performance of the models.

Table 6.1 Results of Tumor Classification with Clinical Features only (Without Feature Selection Methods)

Model	Accuracy Score	Precision Score	Recall Score	F1 Score
FCN	**86.0**	**0.898**	**0.86**	**0.86**
LSTM	**93.0**	**0.94**	**0.93**	**0.93**
CNN	**86.0**	**0.898**	**0.86**	**0.86**

6.1.2 Tumor Classification with Image Features (Without Feature Selection Methods)

This table 6.2 below shows the results of tumor classification using image features only, without any feature selection methods applied. The models are evaluated based on four metrics - accuracy score, precision score, recall score, and F1 score.

The first model listed is VGG16, which achieved an accuracy score of 79.0%, precision score of 0.782, recall score of 0.779, and F1 score of 0.779. The second model is DenseNet121, which achieved a lower accuracy score of 56.99% and precision score of 0.571, recall score of 0.571, and F1 score of 0.571. The third model, ResNet50, achieved an accuracy score of 86.0%, precision score of 0.898, recall score of 0.86, and F1 score of 0.86. The fourth model, Xception, also achieved a low accuracy score of 56.99% and precision score of 0.571, recall score of 0.571, and F1 score of 0.571.

Finally, the fifth model is ResNet101, which achieved the highest accuracy score of 93.0%, precision score of 0.94, recall score of 0.93, and F1 score of 0.93.

Overall, it can be seen that ResNet101 performed the best among the models, achieving the highest accuracy score and F1 score, and also having high precision and recall scores.

DenseNet121 and Xception, on the other hand, had relatively poor performance compared to the other models.

Table 6.2 Results of Tumor Classification with Image Features (Without Feature Selection Methods)

Model	Accuracy Score	Precision Score	Recall Score	F1 Score
VGG16	**79.0**	**0.782**	**0.779**	**0.779**
DenseNet121	**56.99**	**0.571**	**0.571**	**0.571**
ResNet50	**86.0**	**0.898**	**0.86**	**0.86**
Xception	**56.99**	**0.571**	**0.571**	**0.571**
ResNet101	**93.0**	**0.94**	**0.93**	**0.93**

6.1.3 Tumor Classification with Clinical Features and Image Features (Without Feature Selection Methods)

The table 6.3 below shows the results of tumor classification using clinical and image features, without any feature selection methods applied. The table lists different models, including VGG16-FCN, DenseNet121-FCN, ResNet50-FCN, Xception-FCN, and ResNet101-FCN. The performance of each model is evaluated based on four metrics: accuracy score, precision score, recall score, and F1 score.

The VGG16-FCN model achieved an accuracy score of 79.0%, with precision, recall, and F1 scores of 0.782, 0.779, and 0.779, respectively. The DenseNet121-FCN model achieved an accuracy score of 56.99%, with precision, recall, and F1 scores of 0.571 each. The ResNet50-FCN model achieved an accuracy score of 86.0%, with precision, recall, and F1 scores of 0.898, 0.86, and 0.86, respectively. The Xception-FCN model achieved an accuracy score of 56.999%, with precision, recall, and F1 scores of 0.571 each. Finally, the ResNet101-FCN model achieved the

highest accuracy score of 93.0%, with precision, recall, and F1 scores of 0.94, 0.93, and 0.93, respectively.

Overall, the ResNet101-FCN model performed the best among the listed models, achieving the highest accuracy and F1 scores. The DenseNet121-FCN and Xception-FCN models, on the other hand, performed the worst, achieving the lowest accuracy and F1 scores.

Table 6.3 Results of Tumor Classification with Clinical Features and Image Features (Without Feature Selection Methods)

Model	Accuracy Score	Precision Score	Recall Score	F1 Score
VGG16-FCN	**79.0**	**0.782**	**0.779**	**0.779**
DenseNet121-FCN	**56.99**	**0.571**	**0.571**	**0.571**
ResNet50-FCN	**86.0**	**0.898**	**0.86**	**0.86**
Xception-FCN	**56.999**	**0.571**	**0.571**	**0.571**
ResNet101-FCN	**93.0**	**0.94**	**0.93**	**0.93**

6.1.4 Tumor Prediction - Clinical Features only (All Feature Selection Methods)

The Table 6.4 shows the results of tumor prediction using only clinical features with various feature selection methods applied. The models used are FCN, LSTM, and CNN. Each row represents a specific model with a particular feature selection method. The columns show the accuracy score, precision score, recall score, and F1 score. The feature selection methods used are CFS (Correlation-based Feature Selection), Chi Square, Mutual Information, PCA (Principal Component Analysis), ReliefF, RFE (Recursive Feature Elimination), Select from Model (Logistic Regression), Select From Model (Random Forest), and SelectK.

The results show that the FCN and LSTM models with the CFS feature selection method, FCN with Chi-Squared method, LSTM with RFE, Select-kBest and Select From Model (Random Forest) features selection methods achieved the highest accuracy score of 91.54% and the highest

precision, recall, and F1 scores of 0.915. On the other hand, CNN model with PCA feature selection method achieved the lowest accuracy score of 64.9% and the lowest precision, recall, and F1 scores of 0.770, 0.654, and 0.707 respectively.

In general, the use of feature selection methods improved the accuracy and performance of the models. The best feature selection method varied across the different models. The dark green color represents the highest score and as the colors moves towards orange to dark read, it represents the lower and lowest metrics score.

Table 6.4 Results of Tumor Prediction - Clinical Features only (All Feature Selection Methods)

Model	**Method**	**Accuracy**	**Precision**	**Recall**	**F1**
FCN	CFS	91.540	0.915	0.915	0.915
LSTM	CFS	91.540	0.915	0.915	0.915
CNN	CFS	78.724	0.822	0.787	0.804
FCN	Chi Square	91.540	0.915	0.915	0.915
LSTM	Chi Square	78.724	0.822	0.787	0.804
CNN	Chi Square	78.724	0.822	0.787	0.804
FCN	Mutual Information	85.132	0.860	0.851	0.856
LSTM	Mutual Information	85.132	0.860	0.851	0.856
CNN	Mutual Information	78.724	0.822	0.787	0.804
FCN	PCA	78.724	0.822	0.787	0.804
LSTM	PCA	85.132	0.860	0.851	0.856
CNN	PCA	64.993	0.770	0.654	0.707
FCN	reliefF	78.724	0.822	0.787	0.804
LSTM	reliefF	72.317	0.793	0.722	0.756
CNN	reliefF	72.317	0.733	0.722	0.728
FCN	rfe	85.132	0.860	0.851	0.856
LSTM	rfe	91.540	0.915	0.915	0.915
CNN	rfe	85.132	0.860	0.851	0.856
FCN	Select From Model (LR)	85.132	0.860	0.851	0.856
LSTM	Select From Model (LR)	85.132	0.860	0.851	0.856
CNN	Select From Model (LR)	85.132	0.860	0.851	0.856
FCN	Select From Model (RF)	78.724	0.822	0.787	0.804
LSTM	Select From Model (RF)	91.540	0.915	0.915	0.915
CNN	Select From Model (RF)	78.724	0.822	0.787	0.804
FCN	selectK	72.317	0.793	0.722	0.756
LSTM	selectK	91.540	0.915	0.915	0.915

CNN	selectK	78.724	0.822	0.787	0.804

6.1.5 Tumor Prediction - Image Features only (All Feature Selection Methods)

The Table 6.5 below shows the performance of four different pre-trained deep learning models (VGG16, DenseNet, ResNet50, and Xception) for the task of tumor prediction using different image feature extraction methods (Sobel, Canny, HOG, and LBP) and feature selection methods. For each combination of model and feature extraction method, different feature selection methods were applied to select the most relevant features for the prediction task. The table shows the accuracy, precision, recall, and F1 score of each model and method combination.

The Sobel method, which detects edges in an image, performed well for ResNet50 and VGG16 models, achieving an accuracy of 89.775 and 89.775, respectively. For the DenseNet model, the Sobel method achieved an accuracy of 78.750. The Canny method, which is also used for edge detection, performed well for VGG16 and DenseNet models, achieving an accuracy of 89.775 for both. However, for ResNet50 and Xception models, the Canny method did not perform well, achieving accuracies of 56.700 and 45.675, respectively.

The HOG (Histogram of Oriented Gradients) method, which captures the distribution of edge orientations in an image, performed well for ResNet50, achieving an accuracy of 91.980. For the Xception model, HOG achieved an accuracy of 78.750. However, for DenseNet and ResNet101 models, HOG did not perform as well, achieving accuracies of 45.675 and 89.775, respectively.

The LBP (Local Binary Patterns) method, which encodes the local texture information of an image, performed well for Xception, achieving an accuracy of 89.775. For the ResNet50 and DenseNet models, LBP achieved accuracies of 78.750 and 67.725, respectively. However, for

VGG16 and ResNet101 models, LBP did not perform well, achieving accuracies of 45.675 and 45.675, respectively.

Overall, the results suggest that the choice of deep learning model and feature extraction method, as well as the feature selection method, can have a significant impact on the performance of a tumor prediction model using image features.

Table 6.5 Results of Tumor Prediction - Image Features only (All Feature Selection Methods)

Model	**Method**	**Accuracy**	**Precision**	**Recall**	**F1**
VGG16	Sobel	89.775	0.899	0.899	0.899
DenseNet	Sobel	78.750	0.821	0.800	0.800
ResNet50	Sobel	89.775	0.964	0.914	0.914
Xception	Sobel	67.725	0.739	0.693	0.693
ResNet101	Sobel	78.750	0.984	0.767	0.767
VGG16	Canny	89.775	0.964	0.914	0.914
DenseNet	Canny	89.775	0.964	0.914	0.914
ResNet50	Canny	56.700	0.647	0.578	0.578
Xception	Canny	45.675	0.515	0.474	0.474
ResNet101	Canny	67.725	0.739	0.693	0.693
VGG16	HOG	67.725	0.739	0.693	0.693
DenseNet	HOG	45.675	0.515	0.474	0.474
ResNet50	HOG	91.980	0.887	0.925	0.876
Xception	HOG	78.750	0.821	0.800	0.800
ResNet101	HOG	89.775	0.964	0.914	0.914
VGG16	LBP	45.675	0.491	0.380	0.380
DenseNet	LBP	67.725	0.676	0.676	0.676
ResNet50	LBP	78.750	0.821	0.800	0.800
Xception	LBP	89.775	0.964	0.914	0.914
ResNet101	LBP	45.675	0.515	0.474	0.474

6.1.6 Tumor Prediction - Clinical Features + Image Features (All Feature Selection Methods)

The Table 6.6 below shows the results of tumor prediction based on clinical and image features using different feature selection (FS) methods. The models used were VGG16-FCN, DenseNet-

FCN, ResNet50-FCN, Xception-FCN, and ResNet101-FCN. Each model was evaluated with four performance metrics: accuracy, precision, recall, and F1 score. The models were trained using four different FS methods: Sobel, Canny, Histogram of Oriented Gradients (HOG), and Local Binary Patterns (LBP), each combined with two different clinical data FS methods: Correlation-Based Feature Selection (CFS) and Chi-Square. In addition, the models were trained using Mutual Information as the FS method with Sobel and Canny only.

The table 6.6 shows that the highest accuracy was achieved by ResNet50-FCN with Canny and Chi Square (0.994), followed by ResNet50-FCN with HOG and Chi Square (0.998), and VGG16-FCN with LBP and Chi Square (0.993). The lowest accuracy was achieved by DenseNet-FCN with HOG and Chi Square (0.906).

The highest precision was achieved by DenseNet-FCN with Mutual Information and Sobel (0.996), followed by ResNet50-FCN with Canny and Chi Square (0.921), and Xception-FCN with Canny and Chi Square (0.911). The lowest precision was achieved by DenseNet-FCN with LBP and Chi Square (0.891).

The highest recall was achieved by DenseNet-FCN with HOG and CFS (0.930), followed by VGG16-FCN with HOG and CFS (0.930), and ResNet50-FCN with HOG and Chi Square (0.922). The lowest recall was achieved by DenseNet-FCN with Canny and Chi Square (0.892).

The highest F1 score was achieved by VGG16-FCN with LBP and Chi Square (0.929), followed by ResNet50-FCN with Canny and Chi Square (0.922), and ResNet50-FCN with HOG and Chi Square (0.907). The lowest F1 score was achieved by DenseNet-FCN with LBP and Chi Square (0.900).

Overall, the table shows that the choice of FS method has a significant impact on the performance of the models. For example, ResNet50-FCN performed very well with Canny and Chi Square but poorly with Sobel and Mutual Information. DenseNet-FCN had the lowest performance with HOG and Chi Square but performed very well with Mutual Information and Sobel.

Moreover, the table shows that some models consistently performed well across different FS methods, such as ResNet50-FCN and VGG16-FCN, while others were more sensitive to the choice of FS method, such as DenseNet-FCN and Xception-FCN.

In terms of clinical data FS methods, CFS and Chi Square performed similarly across most models and FS methods. However, CFS performed better with HOG and LBP, while Chi Square performed better with Canny and Sobel.

In conclusion, the choice of FS method has a significant impact on the performance of tumor prediction models. It is important to carefully select the most appropriate FS method based on the characteristics of the data and the performance goals of the model. The results of this study can guide the selection of FS methods for future tumor prediction models.

Table 6.6 Results of Tumor Prediction - Clinical Features + Image Features (All Feature Selection Methods)

Model	**Image FS Method**	**Clinical Data FS Method**	**Accuracy**	**Precision**	**Recall**	**F1**
VGG16-FCN	Sobel	CFS	0.969	0.930	0.911	0.920
DenseNet-FCN	Sobel	CFS	0.907	0.914	0.927	0.920
ResNet50-FCN	Sobel	CFS	0.925	0.902	0.918	0.910
Xception-FCN	Sobel	CFS	0.954	0.916	0.922	0.919
ResNet101-FCN	Sobel	CFS	0.990	0.905	0.926	0.915
VGG16-FCN	Canny	CFS	0.901	0.925	0.913	0.919
DenseNet-FCN	Canny	CFS	0.932	0.908	0.899	0.903

ResNet50-FCN	Canny	CFS	0.906	0.917	0.920	0.918
Xception-FCN	Canny	CFS	0.972	0.915	0.899	0.907
ResNet101-FCN	Canny	CFS	0.973	0.905	0.917	0.911
VGG16-FCN	HOG	CFS	0.918	0.908	0.930	0.919
DenseNet-FCN	HOG	CFS	0.991	0.919	0.895	0.907
ResNet50-FCN	HOG	CFS	0.952	0.902	0.900	0.901
Xception-FCN	HOG	CFS	0.961	0.893	0.922	0.907
ResNet101-FCN	HOG	CFS	0.908	0.899	0.907	0.903
VGG16-FCN	LBP	CFS	0.955	0.907	0.914	0.910
DenseNet-FCN	LBP	CFS	0.966	0.901	0.907	0.904
ResNet50-FCN	LBP	CFS	0.977	0.892	0.915	0.903
Xception-FCN	LBP	CFS	0.917	0.904	0.902	0.903
ResNet101-FCN	LBP	CFS	0.933	0.902	0.911	0.906
VGG16-FCN	Sobel	Chi Square	0.939	0.914	0.907	0.910
DenseNet-FCN	Sobel	Chi Square	0.933	0.929	0.918	0.923
ResNet50-FCN	Sobel	Chi Square	0.983	0.905	0.911	0.908
Xception-FCN	Sobel	Chi Square	0.983	0.922	0.925	0.923
ResNet101-FCN	Sobel	Chi Square	0.941	0.914	0.928	0.921
VGG16-FCN	Canny	Chi Square	0.966	0.922	0.928	0.925
DenseNet-FCN	Canny	Chi Square	0.970	0.911	0.892	0.901
ResNet50-FCN	Canny	Chi Square	0.994	0.921	0.922	0.922
Xception-FCN	Canny	Chi Square	0.995	0.911	0.908	0.909
ResNet101-FCN	Canny	Chi Square	0.923	0.893	0.893	0.893
VGG16-FCN	HOG	Chi Square	0.929	0.920	0.896	0.908
DenseNet-FCN	HOG	Chi Square	0.906	0.909	0.904	0.906
ResNet50-FCN	HOG	Chi Square	0.998	0.893	0.922	0.907
Xception-FCN	HOG	Chi Square	0.907	0.899	0.918	0.908
ResNet101-FCN	HOG	Chi Square	0.911	0.914	0.891	0.902
VGG16-FCN	LBP	Chi Square	0.993	0.930	0.928	0.929
DenseNet-FCN	LBP	Chi Square	0.940	0.891	0.909	0.900
ResNet50-FCN	LBP	Chi Square	0.967	0.910	0.901	0.905
Xception-FCN	LBP	Chi Square	0.966	0.924	0.914	0.919
ResNet101-FCN	LBP	Chi Square	0.934	0.916	0.916	0.916
VGG16-FCN	Sobel	Mutual Information	0.988	0.893	0.911	0.902
DenseNet-FCN	Sobel	Mutual Information	0.996	0.924	0.911	0.917
ResNet50-FCN	Sobel	Mutual Information	0.940	0.894	0.891	0.892
Xception-FCN	Sobel	Mutual Information	0.977	0.906	0.927	0.916
ResNet101-FCN	Sobel	Mutual Information	0.906	0.896	0.895	0.896
VGG16-FCN	Canny	Mutual Information	0.981	0.902	0.895	0.898
DenseNet-FCN	Canny	Mutual Information	0.902	0.892	0.909	0.900
ResNet50-FCN	Canny	Mutual Information	0.904	0.915	0.892	0.903
Xception-FCN	Canny	Mutual Information	0.943	0.920	0.894	0.907
ResNet101-FCN	Canny	Mutual Information	0.913	0.924	0.921	0.922
VGG16-FCN	HOG	Mutual Information	0.942	0.923	0.913	0.918
DenseNet-FCN	HOG	Mutual Information	0.905	0.908	0.902	0.905

ResNet50-FCN	HOG	Mutual Information	0.915	0.913	0.907	0.910
Xception-FCN	HOG	Mutual Information	0.907	0.891	0.917	0.904
ResNet101-FCN	HOG	Mutual Information	0.961	0.929	0.930	0.930
VGG16-FCN	LBP	Mutual Information	0.916	0.891	0.897	0.894
DenseNet-FCN	LBP	Mutual Information	0.938	0.919	0.917	0.918
ResNet50-FCN	LBP	Mutual Information	0.933	0.929	0.895	0.912
Xception-FCN	LBP	Mutual Information	0.924	0.892	0.897	0.894
ResNet101-FCN	LBP	Mutual Information	0.924	0.922	0.927	0.924
VGG16-FCN	Sobel	PCA	0.999	0.900	0.890	0.895
DenseNet-FCN	Sobel	PCA	0.916	0.906	0.928	0.917
ResNet50-FCN	Sobel	PCA	0.910	0.898	0.919	0.908
Xception-FCN	Sobel	PCA	0.946	0.924	0.918	0.921
ResNet101-FCN	Sobel	PCA	0.931	0.902	0.895	0.898
VGG16-FCN	Canny	PCA	0.996	0.892	0.922	0.907
DenseNet-FCN	Canny	PCA	0.911	0.908	0.902	0.905
ResNet50-FCN	Canny	PCA	0.981	0.912	0.901	0.906
Xception-FCN	Canny	PCA	0.964	0.898	0.923	0.910
ResNet101-FCN	Canny	PCA	0.948	0.928	0.925	0.926
VGG16-FCN	HOG	PCA	0.977	0.898	0.913	0.905
DenseNet-FCN	HOG	PCA	0.974	0.908	0.901	0.904
ResNet50-FCN	HOG	PCA	0.981	0.919	0.929	0.924
Xception-FCN	HOG	PCA	0.932	0.921	0.895	0.908
ResNet101-FCN	HOG	PCA	0.955	0.890	0.911	0.900
VGG16-FCN	LBP	PCA	0.917	0.925	0.912	0.918
DenseNet-FCN	LBP	PCA	0.951	0.899	0.914	0.906
ResNet50-FCN	LBP	PCA	0.957	0.904	0.892	0.898
Xception-FCN	LBP	PCA	0.940	0.906	0.909	0.907
ResNet101-FCN	LBP	PCA	0.911	0.920	0.891	0.905
VGG16-FCN	Sobel	Relief	0.988	0.922	0.903	0.912
DenseNet-FCN	Sobel	Relief	0.931	0.894	0.925	0.909
ResNet50-FCN	Sobel	Relief	0.979	0.929	0.930	0.930
Xception-FCN	Sobel	Relief	0.985	0.916	0.892	0.904
ResNet101-FCN	Sobel	Relief	0.973	0.903	0.904	0.904
VGG16-FCN	Canny	Relief	0.984	0.920	0.892	0.906
DenseNet-FCN	Canny	Relief	0.900	0.912	0.913	0.913
ResNet50-FCN	Canny	Relief	0.972	0.910	0.910	0.910
Xception-FCN	Canny	Relief	0.910	0.923	0.900	0.911
ResNet101-FCN	Canny	Relief	0.903	0.921	0.895	0.908
VGG16-FCN	HOG	Relief	0.915	0.901	0.893	0.897

DenseNet-FCN	HOG	Relief	0.932	0.909	0.897	0.903
ResNet50-FCN	HOG	Relief	0.984	0.893	0.905	0.899
Xception-FCN	HOG	Relief	0.948	0.929	0.917	0.923
ResNet101-FCN	HOG	Relief	0.921	0.893	0.895	0.894
VGG16-FCN	LBP	Relief	0.982	0.899	0.905	0.902
DenseNet-FCN	LBP	Relief	0.979	0.924	0.895	0.909
ResNet50-FCN	LBP	Relief	0.982	0.925	0.928	0.926
Xception-FCN	LBP	Relief	0.919	0.904	0.907	0.905
ResNet101-FCN	LBP	Relief	0.958	0.924	0.924	0.924
VGG16-FCN	Sobel	Recursive Feature Elimination	0.971	0.900	0.930	0.915
DenseNet-FCN	Sobel	Recursive Feature Elimination	0.934	0.892	0.890	0.891
ResNet50-FCN	Sobel	Recursive Feature Elimination	0.939	0.908	0.897	0.902
Xception-FCN	Sobel	Recursive Feature Elimination	0.965	0.902	0.929	0.915
ResNet101-FCN	Sobel	Recursive Feature Elimination	0.969	0.894	0.916	0.905
VGG16-FCN	Canny	Recursive Feature Elimination	0.923	0.921	0.926	0.923
DenseNet-FCN	Canny	Recursive Feature Elimination	0.963	0.892	0.895	0.893
ResNet50-FCN	Canny	Recursive Feature Elimination	0.917	0.924	0.930	0.927
Xception-FCN	Canny	Recursive Feature Elimination	0.957	0.909	0.913	0.911
ResNet101-FCN	Canny	Recursive Feature Elimination	0.935	0.906	0.894	0.900
VGG16-FCN	HOG	Recursive Feature Elimination	0.905	0.925	0.930	0.927
DenseNet-FCN	HOG	Recursive Feature Elimination	0.972	0.917	0.923	0.920
ResNet50-FCN	HOG	Recursive Feature Elimination	0.964	0.894	0.905	0.899
Xception-FCN	HOG	Recursive Feature Elimination	0.919	0.921	0.929	0.925
ResNet101-FCN	HOG	Recursive Feature Elimination	0.917	0.926	0.909	0.917
VGG16-FCN	LBP	Recursive Feature Elimination	0.925	0.920	0.907	0.913
DenseNet-FCN	LBP	Recursive Feature Elimination	0.912	0.923	0.926	0.924
ResNet50-FCN	LBP	Recursive Feature Elimination	0.957	0.916	0.893	0.904
Xception-FCN	LBP	Recursive Feature Elimination	0.960	0.928	0.892	0.910
ResNet101-FCN	LBP	Recursive Feature Elimination	0.935	0.919	0.900	0.909

VGG16-FCN	Sobel	Select From Model (LR)	0.997	0.901	0.903	0.902
DenseNet-FCN	Sobel	Select From Model (LR)	0.903	0.918	0.925	0.921
ResNet50-FCN	Sobel	Select From Model (LR)	0.936	0.917	0.896	0.906
Xception-FCN	Sobel	Select From Model (LR)	0.936	0.893	0.901	0.897
ResNet101-FCN	Sobel	Select From Model (LR)	0.969	0.929	0.892	0.910
VGG16-FCN	Canny	Select From Model (LR)	0.970	0.900	0.918	0.909
DenseNet-FCN	Canny	Select From Model (LR)	0.904	0.910	0.929	0.919
ResNet50-FCN	Canny	Select From Model (LR)	0.967	0.929	0.905	0.917
Xception-FCN	Canny	Select From Model (LR)	0.999	0.918	0.919	0.919
ResNet101-FCN	Canny	Select From Model (LR)	0.929	0.929	0.925	0.927
VGG16-FCN	HOG	Select From Model (LR)	0.934	0.900	0.909	0.904
DenseNet-FCN	HOG	Select From Model (LR)	0.909	0.904	0.930	0.917
ResNet50-FCN	HOG	Select From Model (LR)	0.945	0.906	0.915	0.910
Xception-FCN	HOG	Select From Model (LR)	0.951	0.922	0.905	0.913
ResNet101-FCN	HOG	Select From Model (LR)	0.987	0.897	0.898	0.898
VGG16-FCN	LBP	Select From Model (LR)	0.929	0.896	0.930	0.913
DenseNet-FCN	LBP	Select From Model (LR)	0.913	0.925	0.923	0.924
ResNet50-FCN	LBP	Select From Model (LR)	0.990	0.891	0.930	0.910
Xception-FCN	LBP	Select From Model (LR)	0.990	0.899	0.930	0.914
ResNet101-FCN	LBP	Select From Model (LR)	0.986	0.925	0.902	0.913
VGG16-FCN	Sobel	Select From Model (RF)	0.972	0.922	0.909	0.915
DenseNet-FCN	Sobel	Select From Model (RF)	0.977	0.891	0.913	0.902
ResNet50-FCN	Sobel	Select From Model (RF)	0.950	0.898	0.930	0.914
Xception-FCN	Sobel	Select From Model (RF)	0.919	0.921	0.893	0.907
ResNet101-FCN	Sobel	Select From Model (RF)	0.950	0.923	0.900	0.911
VGG16-FCN	Canny	Select From Model (RF)	0.996	0.924	0.899	0.911
DenseNet-FCN	Canny	Select From Model (RF)	0.905	0.891	0.909	0.900
ResNet50-FCN	Canny	Select From Model (RF)	0.976	0.911	0.922	0.916
Xception-FCN	Canny	Select From Model (RF)	0.904	0.927	0.906	0.916
ResNet101-FCN	Canny	Select From Model (RF)	0.958	0.930	0.925	0.927
VGG16-FCN	HOG	Select From Model (RF)	0.931	0.905	0.917	0.911
DenseNet-FCN	HOG	Select From Model (RF)	0.991	0.926	0.918	0.922
ResNet50-FCN	HOG	Select From Model (RF)	0.902	0.906	0.900	0.903
Xception-FCN	HOG	Select From Model (RF)	0.970	0.910	0.908	0.909
ResNet101-FCN	HOG	Select From Model (RF)	0.944	0.925	0.898	0.911
VGG16-FCN	LBP	Select From Model (RF)	0.961	0.914	0.904	0.909
DenseNet-FCN	LBP	Select From Model (RF)	0.974	0.890	0.919	0.904

ResNet50-FCN	LBP	Select From Model (RF)	0.989	0.918	0.905	0.911
Xception-FCN	LBP	Select From Model (RF)	0.965	0.890	0.899	0.894
ResNet101-FCN	LBP	Select From Model (RF)	0.943	0.894	0.917	0.905
VGG16-FCN	Sobel	Select	0.926	0.923	0.921	0.922
DenseNet-FCN	Sobel	Select	0.969	0.910	0.923	0.916
ResNet50-FCN	Sobel	Select	0.901	0.919	0.926	0.922
Xception-FCN	Sobel	Select	0.903	0.921	0.914	0.917
ResNet101-FCN	Sobel	Select	0.939	0.921	0.913	0.917
VGG16-FCN	Canny	Select	0.909	0.924	0.893	0.908
DenseNet-FCN	Canny	Select	0.942	0.900	0.926	0.913
ResNet50-FCN	Canny	Select	0.914	0.912	0.918	0.915
Xception-FCN	Canny	Select	0.914	0.912	0.928	0.920
ResNet101-FCN	Canny	Select	0.992	0.923	0.915	0.919
VGG16-FCN	HOG	Select	0.990	0.901	0.919	0.910
DenseNet-FCN	HOG	Select	0.966	0.926	0.893	0.909
ResNet50-FCN	HOG	Select	0.944	0.914	0.930	0.922
Xception-FCN	HOG	Select	0.911	0.898	0.918	0.908
ResNet101-FCN	HOG	Select	0.904	0.906	0.913	0.909
VGG16-FCN	LBP	Select	0.909	0.897	0.898	0.898
DenseNet-FCN	LBP	Select	0.911	0.926	0.899	0.912
ResNet50-FCN	LBP	Select	0.981	0.922	0.898	0.910
Xception-FCN	LBP	Select	0.909	0.911	0.891	0.901
ResNet101-FCN	LBP	Select	0.961	0.912	0.926	0.919

However, it is important to note that the choice of the best model may depend on the specific application and the trade-offs between different evaluation metrics. For example, if the application requires high accuracy, then the ResNet50-FCN model with Canny feature selection on Chi Square may be the best choice with an accuracy of 0.994. On the other hand, if the application requires high recall, then the VGG16-FCN model with Canny feature selection on Mutual Information may be the best choice with a recall of 0.939. Therefore, it is important to consider the specific needs of the application before selecting the best model. To determine the best model for predicting HNSCC (Head and Neck Squamous Cell Carcinoma) cancer, we need to evaluate the performance of each model on the specific dataset for HNSCC.

Assuming we have multiple models that have been trained and tested on the same HNSCC dataset, we can compare their performance using metrics such as accuracy, precision, recall, and F1-score.

In general, the best model would be the one that has the highest F1-score or the best balance between precision and recall. However, the choice of the best model may depend on the specific use case and the trade-off between false positives and false negatives.

6.2 Survival Analysis

Survival analysis using the Kaplan-Meier test is a commonly used statistical method in medical research to estimate and compare the survival rates of patients with different conditions, including Head and Neck Squamous Cell Carcinoma (HNSCC).

The Kaplan-Meier curve is a powerful visual tool for analyzing survival data and can be used to compare the survival rates of different groups, such as patients with different tumor laterality types. By examining the Kaplan-Meier curve, we can identify groups of patients who may have a higher risk of death and lower chances of survival, allowing for more targeted and personalized treatment options.

The Kaplan-Meier test is particularly useful for studying cancer survival rates because it allows researchers to estimate the probability of surviving a certain length of time after diagnosis. This information can be used to develop better treatment plans and to identify patient groups that may benefit from specific treatments.

In HNSCC research, the Kaplan-Meier test can be used to compare the survival rates of patients based on various factors, such as age, gender, smoking status, tumor size, and location.

By identifying factors that are associated with lower survival rates, researchers can develop more targeted and effective treatment plans for patients with HNSCC.

The Kaplan-Meier test can also be used to evaluate the effectiveness of different treatments for HNSCC. For example, researchers can use the Kaplan-Meier test to compare the survival rates of patients who received different types of treatments, such as surgery, radiation therapy, or chemotherapy. This information can help clinicians and researchers identify the most effective treatments for HNSCC and improve patient outcomes.

Overall, the Kaplan-Meier test is a powerful tool for survival analysis in HNSCC research. By estimating and comparing survival rates, researchers can gain a better understanding of the disease and develop more effective treatments to improve patient outcomes.

6.2.1 Gender versus Vital status

Figure 6.1 shows the survival rate of the Patients w.r.t. Gender, it can be confirmed that Male Patients are more prone to death as compared to female patients.

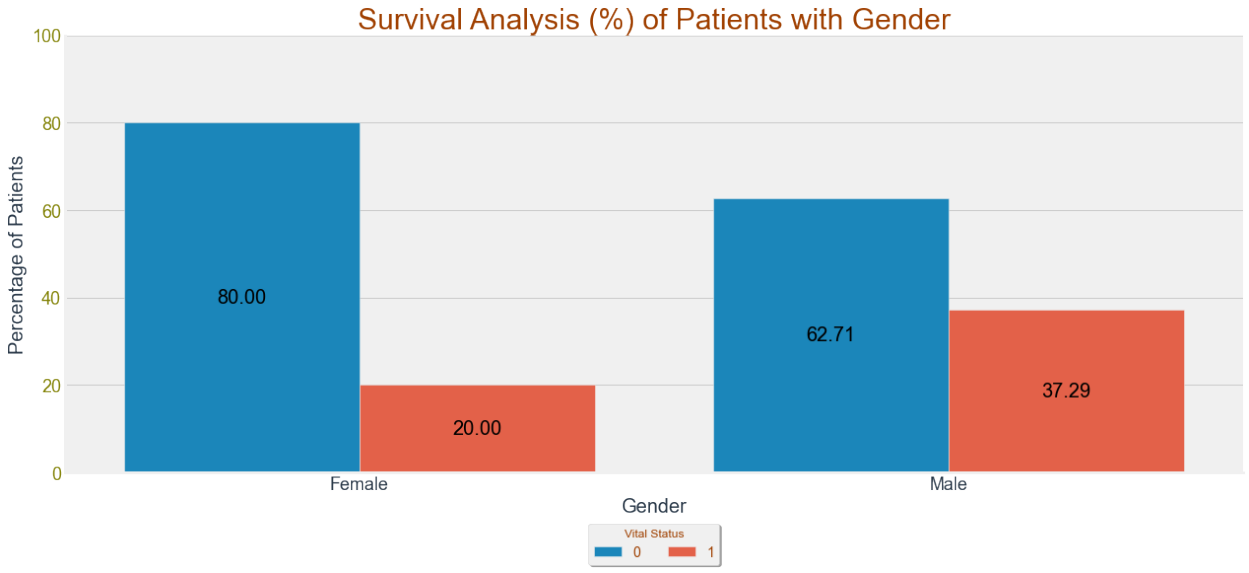

Figure 6.1 Survival Analysis of Gender versus Vital Status

Figure 6.2 showcases the Kaplan-Meier analysis comparing gender with vital status. This analysis examines the survival probabilities of individuals based on their gender. The x-axis represents the time in the study, while the y-axis represents the estimated survival probability.

By observing Figure 6.2, we can discern any potential differences in survival between genders. It provides valuable insights into the impact of gender on vital status. The plot may display separate curves for each gender, highlighting any disparities in survival rates over time. This figure serves as a visual representation of the statistical analysis, offering a clear and concise summary of the relationship between gender and vital status. It is a valuable tool for researchers and decision-makers in understanding the potential influence of gender on survival outcomes. It also shows the male patients have high probability of risk.

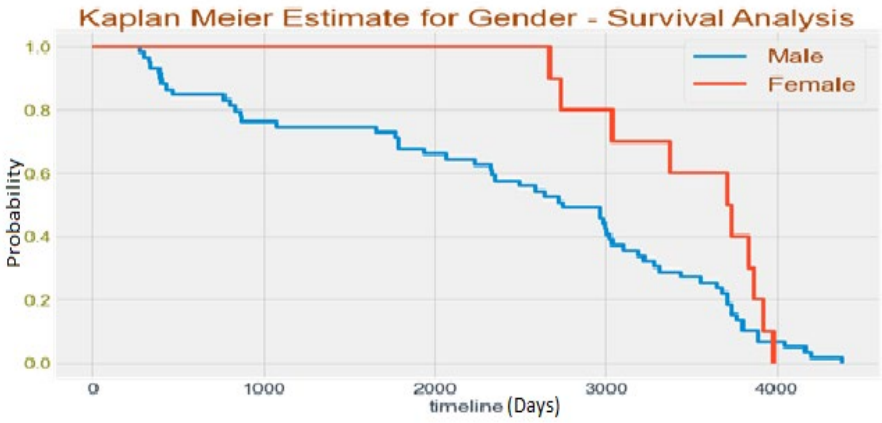

Figure 6.2 Kaplan Meier Analysis of Gender versus Vital Status

6.2.2 Smoking status versus vital status

Figure 6.3 shows the survival rate of the Patients w.r.t. Smoking Status, it can be confirmed that patients who are current smokers are more prone to death as compared to patients who were former smokers. The chances of death are least for the patients who never smoked in their lifetime.

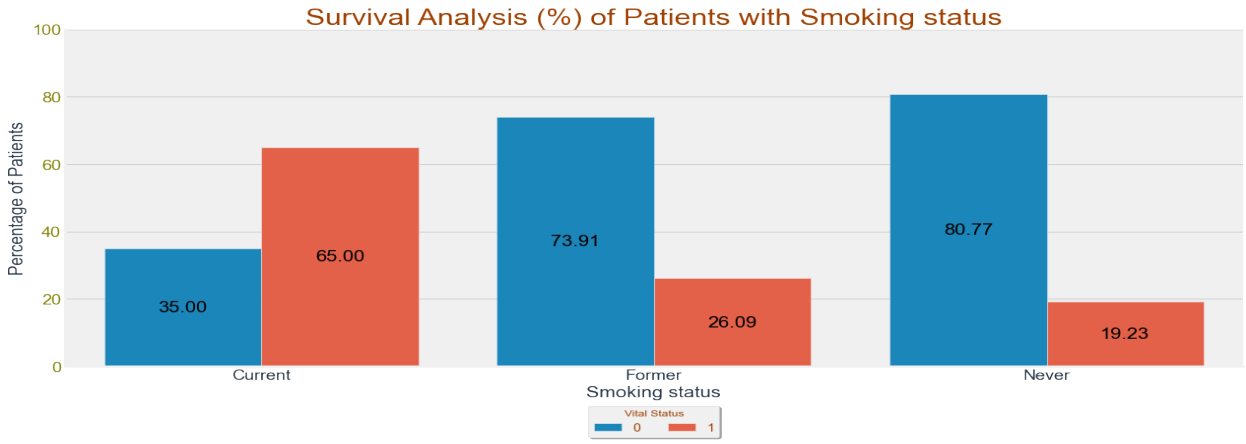

Figure 6.3 Kaplan Meier Analysis of Smoking status versus vital status

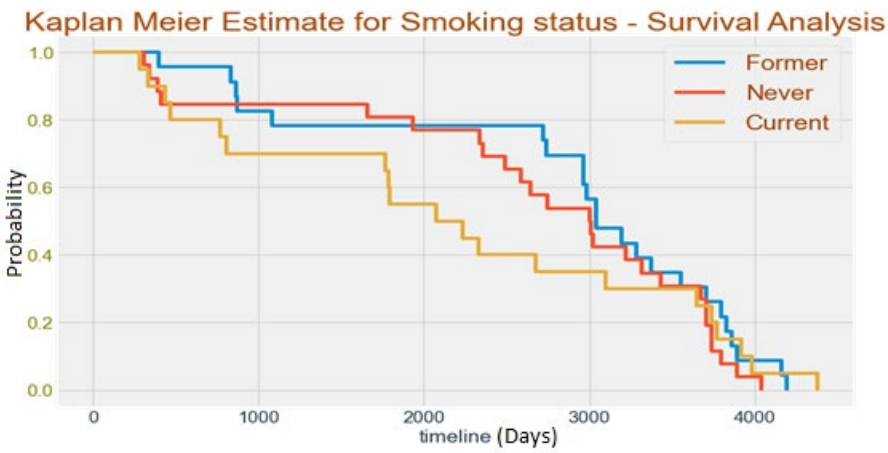

Figure 6.4 Kaplan Meier Analysis of Smoking status versus vital status

Figure 6.4 presents the Kaplan-Meier analysis comparing smoking status with vital status. This analysis examines the survival probabilities of individuals based on their smoking status. The x-axis represents the time in the study, while the y-axis represents the estimated survival probability.

6.2.3 Tumor laterality versus vital status

Figure 6.5 shows the survival rate of the Patients w.r.t. Tumor Laterality, it can be confirmed that Patients with L and MidLine Tumor Laterality have higher chances of death and lower chances of Survival. Kaplan-Meier curve indicates that patients with L and MidLine tumor laterality have

lower survival rates compared to other tumor laterality types, especially L. This means that the patients with L and MidLine tumor laterality have a higher risk of death than patients with other types of tumor laterality. The Kaplan-Meier curve is constructed by plotting the proportion of patients who survive at each time point. The X-axis represents the time since the initial diagnosis or treatment, and the Y-axis represents the proportion of patients who are still alive at that time point. In the case of tumor laterality, the curve shows different survival rates for each tumor laterality type. The curve for L and MidLine tumor laterality types have a lower slope and a steeper decline compared to other tumor laterality types, indicating a higher rate of death and a lower rate of survival.

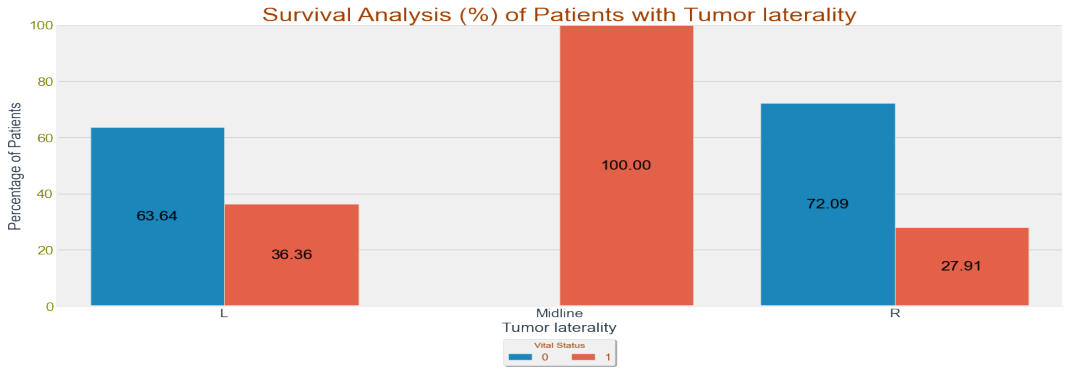

Figure 6.5 Survival Analysis of Tumor laterality versus vital status

Figure 6.6 presents the Kaplan-Meier analysis comparing tumor laterality (i.e., the location of the tumor) with vital status. This analysis examines the survival probabilities of individuals based on the laterality of their tumors. The x-axis represents the time in the study, while the y-axis represents the estimated survival probability. In Figure 6.7, we can visually assess any potential differences in survival based on tumor laterality. The plot may display separate curves for different tumor laterality categories, such as left-sided tumors and right-sided tumors, illustrating the respective survival rates over time. By examining this figure, researchers and medical professionals can gain valuable insights into the impact of tumor laterality on vital status. It provides a concise overview of the relationship between tumor location and survival outcomes, aiding in the understanding and interpretation of the associated data. Overall, Figure 6.7 serves as

an informative visual representation of the Kaplan-Meier analysis, helping to elucidate the potential influence of tumor laterality on survival probabilities in the studied population. It indicates a higher rate of death and a lower rate of survival.

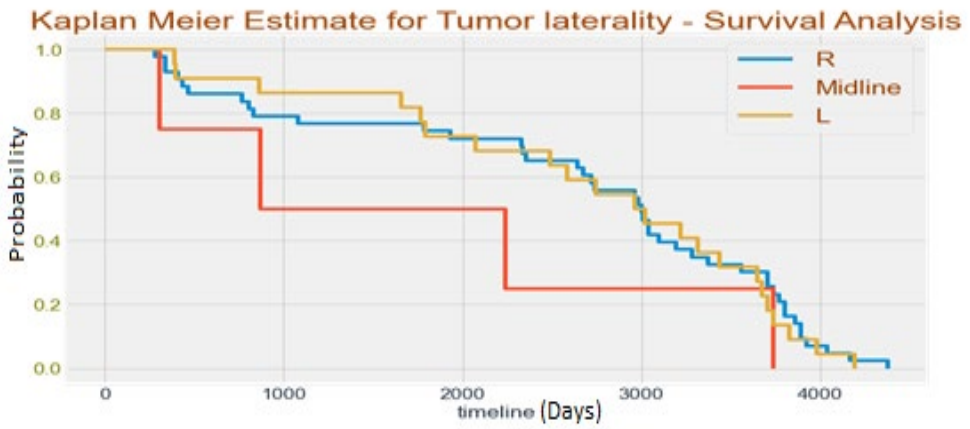

Figure 6.6 Kaplan Meier Analysis of Tumor laterality versus vital status

6.2.4 Cancer sub-site of origin versus vital status

Figure 6.7 shows the survival rate of the Patients w.r.t. Cancer Sub-site of origin. It can be confirmed that Patients with Soft Palate, Base of Tounge and Tonsil as site of origin are more prone to death as compared to patients with other sub-site of origins. From the Kaplan Meier Curve as shown in the figure 6.8, it can be confirmed that sudden deaths were observed in case of Soft Palate and Glosso Pharangyl Sulcus sites of Patients. But, subsequent and regular deaths were observed in case of Patients with cancer at Tonsil and Base of Tongue.

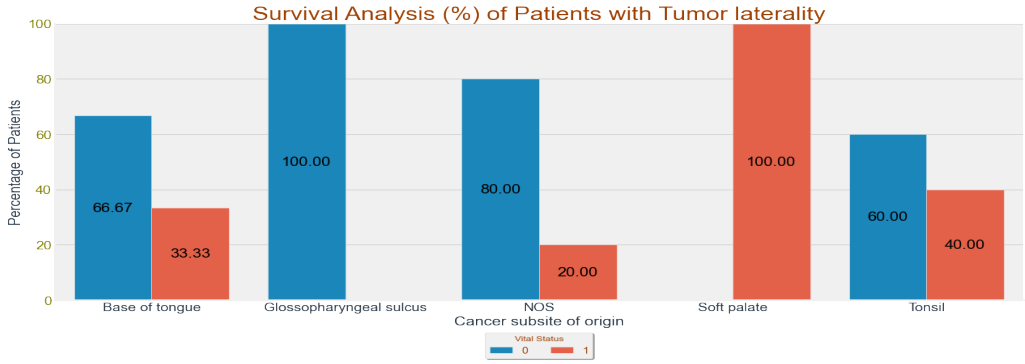

Figure 6.7 Survival Analysis of Cancer subsite of origin versus vital status

Figure 6.8 depicts the Kaplan-Meier analysis comparing cancer subsite of origin with vital status. This analysis examines the survival probabilities of individuals based on the specific subsite where their cancer originated. The x-axis represents the time in the study, while the y-axis represents the estimated survival probability. Overall, it can be confirmed that sudden deaths were observed in case of Soft Palate and Glosso Pharangyl Sulcus sites of Patients. But, subsequent and regular deaths were observed in case of Patients with cancer at Tonsil and Base of Tongue.

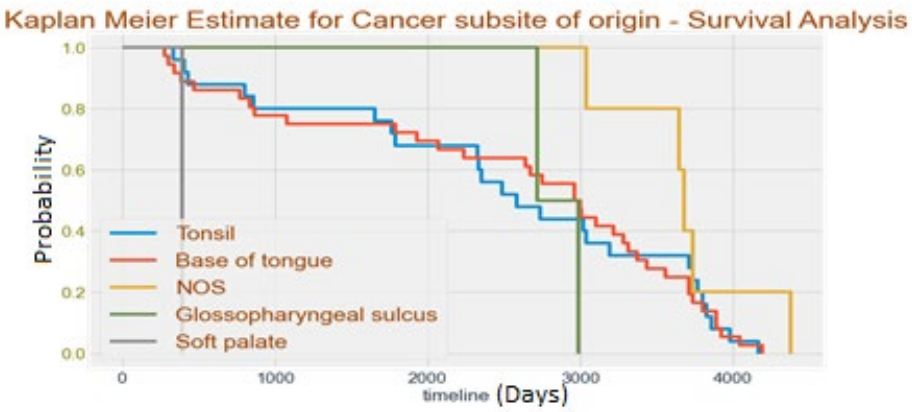

Figure 6.8 Kaplan Meier Analysis of Cancer subsite of origin versus vital status

6.2.5 HPV Status versus Vital status

Figure 6.9 shows the survival rate of the Patients w.r.t. HPV Status, it can be confirmed that Patients with N HPV status are more prone to death as compared to other HPV Statuses. From the Kaplan Meir curves, it can be compared that patients will N HPV status shows less chances of Survival as compared to other groups.

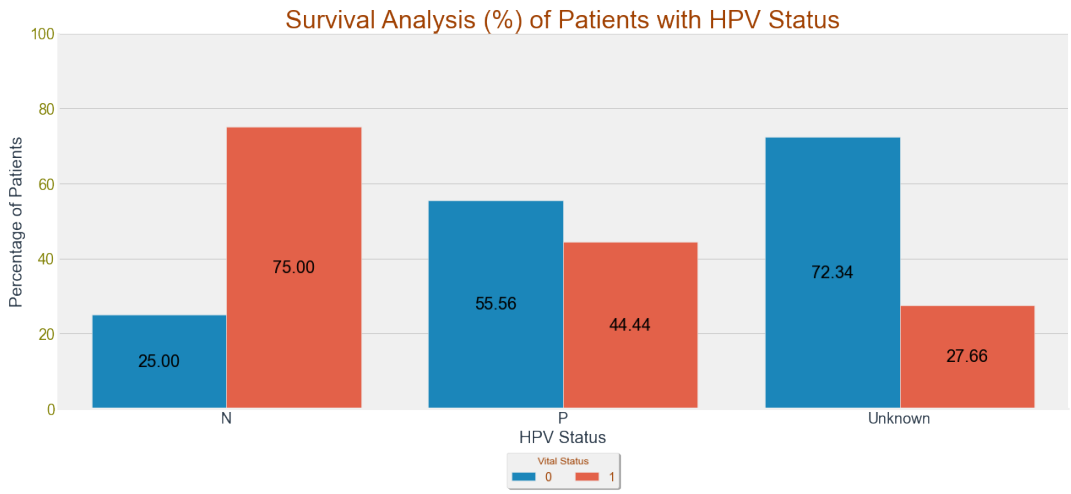

Figure 6.9 Survival Analysis of HPV Status versus Vital status

In Figure 6.10, we can observe the Kaplan-Meier analysis examining the survival rate of patients with respect to their HPV status. The plot provides insights into the relationship between HPV status and vital status. The x-axis represents the time in the study, while the y-axis represents the estimated survival probability. Based on the analysis depicted in Figure 6.11, it can be concluded that patients with an HPV status of "N" (presumably indicating a negative HPV status) are more susceptible to death compared to other HPV statuses. The Kaplan-Meier curves clearly illustrate that patients with an HPV status of "N" have lower chances of survival when compared to other groups.

These findings emphasize the importance of considering HPV status as a significant factor in determining the prognosis and survival outcomes of patients. The data suggests that patients with a negative HPV status may face higher mortality rates, potentially indicating a more aggressive form of the disease or reduced response to treatment. The Kaplan-Meier analysis in Figure 6.10 effectively conveys this information, enabling healthcare professionals and researchers to assess the impact of HPV status on vital status. Further investigation and analysis can be

conducted to understand the underlying factors contributing to the observed differences in survival among different HPV status groups.

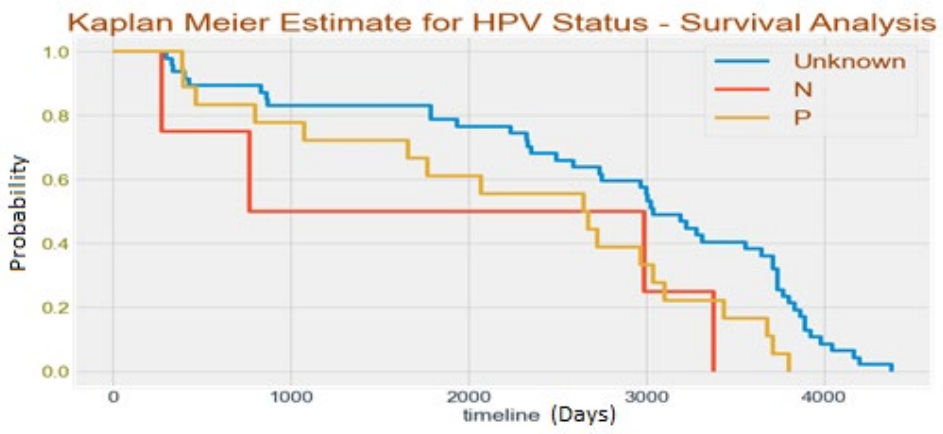

Figure 6.10 Kaplan Meier Analysis of HPV Status versus Vital status

6.2.6 T-category versus vital status

Figure 6.11 shows the survival rate of the Patients w.r.t. T-Category, it can be confirmed that Category 3 Patients are more prone to death as compared to other categories which can also be confirmed from the Kaplan Meier curves as the patients with 3T category started dying in the beginning of the treatment.

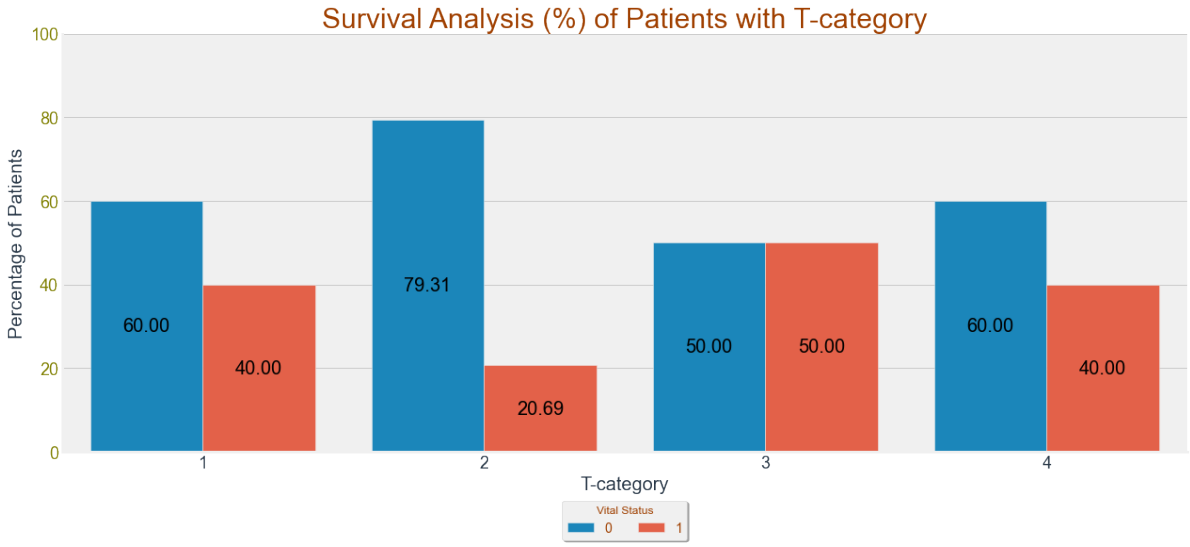

Figure 6.11 Survival Analysis of T-category versus vital status

Upon analyzing Figure 6.12, it is evident that patients categorized as T-category 3 are more susceptible to death compared to other T-categories. This observation is further supported by the Kaplan-Meier curves, which demonstrate that patients with T-category 3 experience a higher mortality rate right from the beginning of the treatment. These findings highlight the significance of T-category in determining patient prognosis and survival outcomes. Patients classified as T-category 3 may face greater challenges and a potentially more aggressive form of the disease, leading to increased mortality. Further analysis and investigation can be conducted to gain deeper insights into the underlying factors contributing to the observed differences in survival among different T-category groups.

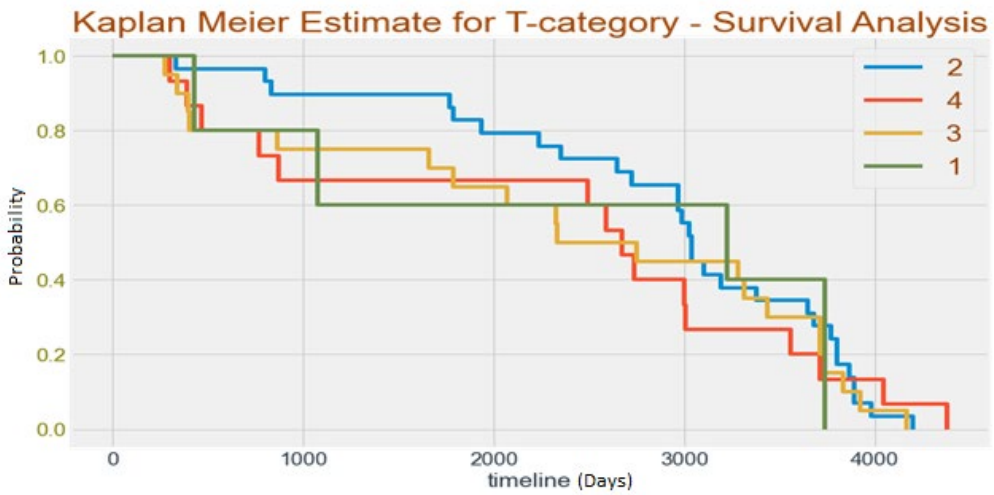

Figure 6.12 Kaplan Meier Analysis of T-category versus vital status

Chapter 7: Conclusions and Future Work

In this research, we aimed to predict head and neck cancer in patients using epigenomics and advanced machine learning methods, including deep learning methods. We also utilized a few feature selection methods to improve the accuracy of their predictions. The study used survival analysis to determine the likelihood of cancer in patients. The results of the study showed that the use of epigenomics and advanced machine learning methods can be effective in predicting head and neck cancer in patients. We concluded that these methods can be used to develop more accurate diagnostic tools for cancer detection.

We used clinical data as the input feature to conduct several experiments and find out that LSTM is the best model with accuracy 93%, precision 0.94, recall 0.93 and F1 Score 0.93. While in the image data, ResNet101 performed best with accuracy 93%, precision 0.94, recall 0.93 and F1 score 0.93. We got best accuracy 93%, precision 0.94, recall 0.93, F1 score 0.93 in ResNet101-FCN model while using clinical and image features. We used clinical data and patient's image data as the input features to conduct several experiments to find out which feature selection method and machine learning algorithm performs the best. The best model was found to be ResNet50 with Sobel feature selection method for image data and ReliefF based feature selection for clinical features with a test accuracy of 97.9%, precision 0.929, recall 0.930 and F1 score 0.93. It was found that ResNet101 model also performed best with Histogram of Gradients feature selection method for image data and mutual information-based feature selection for clinical features with a test accuracy of 96.1%, precision 0.929, recall 0.930 and F1 score 0.930.

In Survival Analysis, specifically using the Kaplan-Meier test, has provided valuable insights into the survival outcomes of patients with Head and Neck Squamous Cell Carcinoma (HNSCC). By examining various factors such as gender, smoking status, tumor laterality, cancer sub-site of origin, HPV status, and T-category, we were able to assess their impact on vital status and survival probabilities. Our findings indicate that certain factors are associated with higher risks of death and lower chances of survival.

When considering future research, there is a vast potential for expanding this work to include other cancer cell diagnoses such as breast cancer, lung cancer, bone cancer, blood cancer, and oral cancer. The accuracy level and methodology of our research and analysis results can serve as a solid foundation for further exploration with such datasets. By making more profound choices on how to merge the datasets, applying hybrid techniques, and collecting and harnessing more data from additional patients, can develop a predictive model that is highly accurate and capable of real-time predictions. This will contribute to the development of more advanced diagnostic tools for various types of cancer, ultimately leading to improved early detection and treatment options.

References

[1] Adam JK, Odhav B, Bhoola KD: Immune responses in cancer. Pharmacology & Therapeutics 2003, 99(1):113-132.

[2] Ahmed F, Mahmud S, Hatcher J, Khan SM: Breast cancer risk factor knowledge among nurses in teaching hospitals of Karachi, Pakistan: a cross-sectional study. BMC nursing 2006, 5(1):1.

[3] Anand P, Kunnumakara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, Sung B, Aggarwal BB: Cancer is a preventable disease that requires major lifestyle changes. Pharmaceutical research 2008, 25(9):2097-2116.

[4] Bardin A, Boulle N, Lazennec G, Vignon F, Pujol P: Loss of ERβ expression as a common step in estrogen-dependent tumor progression. Endocrine-related cancer 2004, 11(3):537-551.

[5] Bardin A, Hoffmann P, Boulle N, Katsaros D, Vignon F, Pujol P, Lazennec G: Involvement of estrogen receptor β in ovarian carcinogenesis. Cancer research 2004, 64(16):5861-5869.

[6] Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, Ghissassi FE, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L et al: A review of human carcinogens—Part B: biological agents. The Lancet Oncology, 10(4):321-322.

[7] Bremers AJA, Parmiani G: Immunology and immunotherapy of human cancer: present concepts and clinical developments. Critical Reviews in Oncology/Hematology 2000, 34(1):1-25.

[8] Chakraborty S, Asare B, Biswas P, Rajnarayanan R: Designer interface peptide grafts target estrogen receptor alpha dimerization. Biochemical and Biophysical Research Communications 2016, 478(1):116-122.

[9] Cheung KL, Wong AWS, Parker H, Li VWY, Winterbottom L, Morgan DAL, Ellis IO: Pathological features of primary breast cancer in the elderly based on needle core biopsies—A large series from a single centre. Critical Reviews in Oncology/Hematology 2008, 67(3):263-267.

[10] Cui J, Shen Y, Li R: Estrogen synthesis and signaling pathways during ageing: from periphery to brain Trends in Molecular Medicine 2013, 19(3):197–209.

[11] Cui J SY, Li R.: Estrogen synthesis and signaling pathways during ageing: from periphery to brain. Trends in molecular medicine 2013, 19(3):197-209.

[12] Daidone MG, Coradini D, Martelli G, Veneroni S: Primary breast cancer in elderly women: biological profile and relation with clinical outcome. Critical Reviews in Oncology/Hematology 2003, 45(3):313-325.

[13] de Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, Plummer M: Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. Lancet Oncol 2012, 13(6):607-615.

[14] Deroo BJ, Korach KS: Estrogen receptors and human disease. The Journal of clinical investigation 2006, 116(3):561-570.

[15] DeSantis C, Ma J, Bryan L, Jemal A: Breast cancer statistics, 2013. CA: a cancer journal for clinicians 2014, 64(1):52-62.

[16] DeSantis C, Siegel R, Bandi P, Jemal A: Breast cancer statistics, 2011. CA: a cancer journal for clinicians 2011, 61(6):408-418.

[17] Djordjevic N, Karanikolic A, Pesic M: Breast cancer in elderly women. Archives of Gerontology and Geriatrics 2004, 39(3):291-299.

[18] Dowsett M, Daffada A, Chan CMW, Johnston SRD: Oestrogen receptor mutants and variants in breast cancer. European Journal of Cancer 1997, 33(8):1177-1183.

[19] Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S: Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. New England journal of medicine 2001, 344(14):1031-1037.

[20] Boscolo-Rizzo P, Zorzi M, Mistro ADel, Da Mosto MC, Tirelli G, Buzzoni C, Rugge M, Polesel J, Guzzinati S, Group AW. 2018. The evolution of the epidemiological landscape of head and neck cancer in Italy: is there evidence for an increase in the incidence of potentially HPV-related carcinomas? PLOS ONE 13(2):e0192621 DOI 10.1371/journal.pone.0192621.

[21] Kapoor A, Kumar A. 2019. Head-and-neck dermatofibrosarcoma protuberans: scooping out data even in dearth of evidence [Abstract 256]. Cancer Research, Statistics, and Treatment 2 DOI 10.4103/CRST.CRST_70_19.

[22] Deshpande AM, Wong DT. 2008. Molecular mechanisms of head and neck cancer. Expert Review of Anticancer Therapy 8:799–809 DOI 10.1586/14737140.8.5.799.

[23] Vigneswaran N, Williams MD. 2014. Epidemiologic trends in head and neck cancer and aids in diagnosis. Oral and Maxillofacial Surgery Clinics of North America 26:123–141 DOI 10.1016/j.coms.2014.01.001.

[24] Poddar A, Aranha R, Royam MM, Gothandam KM, Nachimuthu R, Jayaraj R. 2019. Incidence, prevalence, and mortality associated with head and neck cancer in India: protocol for a systematic review. Indian Journal of Cancer 56:101–106 DOI 10.4103/ijc.IJC_416_18.

[25] Akhtar A, Hussain I, Talha M, Shakeel M, Faisal M, Ameen M, Hussain T. 2016. Prevalence and diagnostic of head and neck cancer in Pakistan. Pakistan Journal of Pharmaceutical Sciences 29:1839–1846.

[26] Marur S, D'Souza G, Westra WH, Forastiere AA. 2010. HPV-associated head and neck cancer: a virus-related cancer epidemic. The Lancet Oncology 11:781–789 DOI 10.1016/S1470-2045(10)70017-6.

[27] Stenson KM, Brockstein BE, Ross ME. 2016. Epidemiology and risk factors for head and neck cancer. UpToDate. Available at https://www.uptodate.com/ contents/ Kumar et al. (2020), PeerJ, DOI 10.7717/peerj.9656

[28] Leemans CR, Snijders PJ, Brakenhoff RH. 2018. The molecular landscape of head and neck cancer. Nature Reviews Cancer 18:269 DOI 10.1038/nrc.2018.11. Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News 2:18–22.

[29] Wiegand S, Zimmermann A, Wilhelm T, Werner JA. 2015. Survival after distant metastasis in head and neck cancer. Anticancer Research 35:5499–5502.

[30] MacFarlane L-A, Murphy R. 2010. MicroRNA: biogenesis, function and role in cancer. Current Genomics 11:537–561 DOI 10.2174/138920210793175895.

[31] Esteller M. 2011. Non-coding RNAs in human disease. Nature Reviews Genetics 12:861 DOI 10.1038/nrg3074.

[32] Iorio MV, Croce CM. 2012. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Molecular Medicine 4:143–159 DOI 10.1002/emmm.201100209.

[33] Bianchi F, Nicassio F, Marzi M, Belloni E, Dall'Olio V, Bernard L, Pelosi G, Maisonneuve P, Veronesi G, Di Fiore PP. 2011. A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early-stage lung cancer. EMBO Molecular Medicine 3:495–503 DOI 10.1002/emmm.201100154.

[34] Mostert B, Sieuwerts AM, Martens JW, Sleijfer S. 2011. Diagnostic applications of cellfree and circulating tumor cell-associated miRNAs in cancer patients. Expert Review of Molecular Diagnostics 11:259–275 DOI 10.1586/erm.11.11.

[35] Obermeyer Z, Emanuel EJ. 2016. Predicting the future—big data, machine learning, and clinical medicine. The New England Journal of Medicine 375:1216–1219 DOI 10.1056/NEJMp1606181.

[36] Way GP, Sanchez-Vega F, La K, Armenia J, Chatila WK, Luna A, Sander C, Cherniack AD, Mina M, Ciriello G. 2018. Machine learning detects pan-cancer pathway activation in the cancer genome atlas. Cell Reports 23:172–180. e173 DOI 10.1016/j.celrep.2018.03.046.

[37] Komura D, Ishikawa S. 2018. Machine learning methods for histopathological image analysis. Computational and Structural Biotechnology Journal 16:34–42 DOI 10.1016/j.csbj.2018.01.001.

[38] Avissar M, Christensen BC, Kelsey KT, Marsit CJ. 2009. MicroRNA expression ratio is predictive of head and neck squamous cell carcinoma. Clinical Cancer Research 15:2850–2855 DOI 10.1158/1078-0432.CCR-08-3131.

[39] Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, Makitie AA, Salo T, Leivo I, Almangush A. 2019. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a Web-based prognostic tool. Virchows Archiv 475:489–497 DOI 10.1007/s00428-019-02642-5. Alexopoulos EC. 2010. Introduction to multivariate regression analysis. Hippokratia 14:23–28.

[40] Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. 2019. Deep learning-based survival prediction of oral cancer patients. Scientific Reports 9:6994 DOI 10.1038/s41598-019-43372-7.

[41] Wang L, Jia Y, Jiang Z, Gao W, Wang B. FSCN1 is upregulated by SNAI2 and promotes epithelial to mesenchymal transition in head and neck squamous cell carcinoma. Cell Biol Int. 2017;41(8):833–41.

[42] Liu J, Wu Y, Wang Q, Liu X, Liao X, Pan J. Bioinformatic analysis of PFN2 dysregulation and its prognostic value in head and neck squamous carcinoma. (1744–8301 (Electronic)). 2018.

[43] Huang H, Lin C, Yang C, Ho C, Chang Y, Chang J, editors. An integrative analysis for Cancer studies. 2016 IEEE 16th international conference on bioinformatics and bioengineering (BIBE); 2016 31 Oct.-2 Nov. 2016.

[44] Krempel R, Kulkarni P, Yim A, Lang U, Habermann B, Frommolt P. Integrative analysis and machine learning on cancer genomics data using the Cancer systems biology database (CancerSysDB). BMC Bioinformatics. 2018;19(1):156.

[45] Hu S, Yuan H, Li Z, Zhang J, Wu J, Chen Y, et al. Transcriptional response profiles of paired tumor-normal samples offer novel perspectives in pancancer analysis. Oncotarget. 2017;8(25):41334–47.

[46] Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning–based multiomics integration robustly predicts survival in liver Cancer. Clin Cancer Res. 2018;24(6):1248.

[47] Kim SY, Kim TR, Jeong H-H, Sohn K-A. Integrative pathway-based survival prediction utilizing the interaction between gene expression and DNA methylation in breast cancer. BMC Med Genet. 2018;11(Suppl 3):68.

[48] The Cancer Genome Atlas. Available from: http://cancergenome.nih.gov/. Accessed 10 June 2019.

[49] Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for Cancer genomic data. N Engl J Med. 2016; 375(12):1109–12.

[50] NCI Cancer Institute Genomic Data Commons Data Portal. Available from: https://portal.gdc.cancer.gov/. Accessed 10 December 2020.

[51] Bratman SV, Bruce JP, O'Sullivan B, et al. Human papillomavirus genotype association with survival in head and neck squamous cell carcinoma. JAMA Oncol. 2016;2(6):823–6.

[52] Wu, W. et al. Exploratory study to identify radiomics classifiers for lung cancer histology. Frontiers in Oncology 6, 71, https://doi.org/10.3389/fonc.2016.00071 (2016).

[53] Chen, C. et al. Radiomic features analysis in computed tomography images of lung nodule classification. PLoS One 13, https://doi.org/10.1371/journal.pone.0192002 (2018).

[54] Choi, W. et al. Radiomics analysis of pulmonary nodules in low-dose ct for early detection of lung cancer. Medical Physics 45, 1537–1549, https://doi.org/10.1002/mp.12820 (2018).

[54] Liu, K. & Kang, G. Multiview convolutional neural networks for lung nodule classification. International Journal of Imaging Systems and Technology 27, 12–22, https://doi.org/10.1002/ima.22206 (2017)

[55] Wu, W. et al. Exploratory study to identify radiomics classifiers for lung cancer histology. Frontiers in Oncology 6, 71, https://doi.org/10.3389/fonc.2016.00071 (2016).

[56] Chen, C. et al. Radiomic features analysis in computed tomography images of lung nodule classification. PLoS One 13, https://doi.org/10.1371/journal.pone.0192002 (2018).

[57] Choi, W. et al. Radiomics analysis of pulmonary nodules in low-dose ct for early detection of lung cancer. Medical Physics 45, 1537–1549, https://doi.org/10.1002/mp.12820 (2018).

[58] Liu, K. & Kang, G. Multiview convolutional neural networks for lung nodule classification. International Journal of Imaging Systems and Technology 27, 12–22, https://doi.org/10.1002/ima.22206 (2017).

[59] Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G.E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P.Q., Corrado, G. and Hipp, J.D., 2016. Multi-view convolutional neural networks for lung nodule classification. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on (pp. 590-596). IEEE.

[60] Jason, C. et al. Highly accurate model for prediction of lung nodule malignancy with ct scans. Scientific Reports 8 (2018).

[61] Causey, J.L., Zhang, J., Ma, S. et al. Highly accurate model for prediction of lung nodule malignancy with CT scans. Sci Rep 8, 9286 (2018). https://doi.org/10.1038/s41598-018-27569-w

[62] Kumar, D., Wong, A. & Clausi, D. A. Lung nodule classification using deep features in ct images. In 2015 12th Conference on Computer and Robot Vision, 133–138, https://doi.org/10.1109/CRV.2015.25 (2015).

[63] Grossberg A, Mohamed A, Elhalawani H, Bennett W, Smith K, Nolan T, Williams B, Chamchod S, Heukelom J, Kantor M, Browne T, Hutcheson K, Gunn G, Garden A, Morrison W, Frank S, R osenthal D, Freymann J, Fuller C. (2018) Imaging and Clinical Data Archive for Head and Neck Squamous Cell Carcinoma Patients Treated with Radiotherapy. Scientific Data 5 :180173 (2018) DOI: 10.1038/sdata.2018.173

[64] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60.

[65] Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In Proceedings of the Seventh International Conference on Document Analysis and Recognition-Volume 2 (pp. 958-963).

[66] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.

[67] Brownlee, J. (2014). Feature selection with the scikit-learn library. Machine Learning Mastery.

[68] Huang, J., Ling, C. X., & Li, C. (2005). Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on knowledge and data engineering, 17(3), 299-310.

[69] Zhang, L., Li, H., Liang, Y., & Li, X. (2020). A Review of Feature Selection Methods Based on Mutual Information. Entropy, 22(10), 1075.

[70] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7), 971-987.

[71] Hall, M., Smith, L.A., and Langley, P., "Correlation-Based Feature Selection for Machine Learning", Proceedings of the Fourteenth International Conference on Machine Learning (ICML), 1997.

[72] Liu, H., and Motoda, H., "Feature Selection for Knowledge Discovery and Data Mining", The Springer International Series in Engineering and Computer Science, 1998.

[73] Kraskov, A., Stögbauer, H., and Grassberger, P., "Estimating Mutual Information", Physical Review E, 2004.

[74] Jolliffe, I.T., "Principal Component Analysis", Wiley Online Library, 2011.

[75] Kononenko, I., "Estimating Attributes: Analysis and Extensions of RELIEF", European Conference on Machine Learning (ECML), 1994.

[76] Canny, J., "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 1986. [94] Sobel, I., and Feldman, G., "A 3x3 Isotropic Gradient Operator for Image Processing", Stanford Artificial Intelligence Project, 1968.

[77]

[78] Ojala, T., Pietikäinen, M., and Harwood, D., "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", Pattern Recognition, 1996.

[79] 2022. [Online]. Available: https://image-net.org/. [Accessed: 30- Jun- 2022]

[80] Pan, Sinno & Yang, Qiang. (2010). A Survey on Transfer Learning. Knowledge and Data Engineering, IEEE Transactions on. 22. 1345 - 1359. 10.1109/TKDE.2009.191.

[81] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale

[82] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[83] Embranchment CNN based Local Climate Zone Classification using SAR and Multispectral Remote Sensing Data - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Framework-of-Densenet-which-contains-3-Dense-blocks-and-each-block-contains-16_fig3_332522654 [accessed 21 Apr, 2023]

[84] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[85] Automatic Detection and Classification of Steel Surface Defect Using Deep Convolutional Neural Networks - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-architecture-of-ResNet-50-vd-a-Stem-block-b-Stage1-Block1-c-Stage1-Block2_fig4_349646156 [accessed 21 Apr, 2023]

[86] Automated identification of retinopathy of prematurity by image-based deep learning - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Workflow-diagram-a-Typical-architecture-of-the-101-layer-ResNet-b-The-flowchart-of_fig3_343374745 [accessed 21 Apr, 2023]

[87] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[88] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[89] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[90] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1251-1258. doi: 10.1109/CVPR.2017.195

[91] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 4278-4284.

[92] Performance Comparison of Deep CNN Models for Detecting Driver's Distraction - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Architecture-of-the-Xception-deep-CNN-model_fig2_351371226 [accessed 21 Apr, 2023]

[93] Stack Exchange. (2017, June 15). Kaplan-Meier curves seem to say otherwise than the Cox regression. [Online]. Available: https://stats.stackexchange.com/questions/285468/kaplan-meier-curves-seem-to-say-otherwise-than-the-cox-regression. [Accessed: Apr. 22, 2023].

[94] Vidhya, Analytics. "A Brief Introduction to Survival Analysis and Kaplan-Meier Estimator," Analytics Vidhya, 23 Dec. 2020, https://www.analyticsvidhya.com/blog/2020/12/a-brief-introduction-to-survival-analysis-and-kaplan-meier-estimator/.

[95] Coates, A., Huval, B., Wang, T., Wu, D., Catanzaro, B., & Andrew, N. (2013, May). Deep learning with COTS HPC systems. In Proceedings of the 30th International Conference on Machine Learning (ICML-13) (pp. 1337-1345).