Pancreatic Cancer Survival Prediction using Deep Learning Techniques

By

Wilson Bakasa

(219098448)

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy

in the

School of Mathematics Statistics & Computer Science

College of Agriculture, Engineering and Science

University of KwaZulu-Natal,

Durban, South Africa

July, 2023

Dedication

The research described in this thesis was conducted at the University of KwaZulu-Natal under the supervision of Prof. Serestina Viriri. I hereby declare that all the materials incorporated in this thesis are my original work except where acknowledgement is made by name or in the form of a reference. The work contained herein has not been submitted in part or whole for a degree at any other university.

Signed..... .....Date.....24/08/2023.....
Wilson Bakasa

As the candidate's supervisor, I have approved this thesis for submission.

Signed.....Date.....24/08/2023.....
Prof. Serestina Viriri

Declaration 1 - Plagiarism

I, **Wilson Bakasa**, declare that;

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   (a) Their words have been re-written but the general information attributed to them has been referenced,
   (b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references sections.

Signed.....  .....Date.....24/08/2023.....
Wilson Bakasa

Declaration 2 - Thesis by Publication

THIS THESIS IS BY PUBLICATION; in that regard, some chapters are published papers. Details of contributions to publications that form part and/or include research presented in this thesis are included in chapters 3, 4 and 5.

Signed.....  .....Date.....24/08/2023.....

Wilson Bakasa

Declaration 3 - Publication

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part and/or include research presented in this thesis.

1. Bakasa, W., Viriri, S. (2021), Pancreatic Cancer Survival Prediction: A Survey of the State-of-the-art, *Computational and Mathematical Methods in Medicine journal*, vol. 2021, p. 1-17. https://doi.org/10.1155/2021/1188414.

2. Bakasa, W., Viriri, S. (2023), Light Gradient-boosting Machine Edge Detection with Cropping Layer for Semantic Segmentation of Pancreas, *Advances in Artificial Intelligence and Machine Learning,*. 2023,3(3):75, AAIML (ISSN: 2582-9793).

3. Bakasa, W., Viriri, S. (2023), Pancreas Instance Segmentation Using Deep Learning Techniques, *Pan-African Artificial Intelligence and Smart Systems,* LNICST, Springer, vol. 459, pp. 1–19, https://doi.org/10.1007/978-3-031-25271-6_13.

4. Bakasa, W., Viriri, S. (2023), Intelligent Automated Pancreas Segmentation using U-Net Model Variants, *Computational Collective Intelligence,* LNAI, Springer, vol. 46, pp. 1-13, https://doi.org/10.1007/978-3-031-41456-5_46.

5. Bakasa, W., Viriri, S. (2023), VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction, *Journal of Imaging,* vol. 9, no. 138. https://doi.org/10.3390/jimaging9070138.

6. Bakasa, W., Viriri, S. (2023), Stacked Ensemble Deep Learning for Pancreas Cancer Classification using Extreme Gradient Boosting, *Frontiers in Artificial Intelligence journal*. - Under Review

Signed.....  .....Date.....24/08/2023.....

Wilson Bakasa

Acknowledgements

This work is dedicated to Jesus ALL MIGHTY

And

to my mother, Venittah Chigwana, and my family.

Not forgetting the support from a long-time friend Mr Clopas Kwenda. I am very grateful to my supervisor, Prof. Serestina Viriri, for the help and support I received from him during my research.

Abstract

This thesis explores the significance of early cancer detection for improving survival rates, particularly focusing on pancreatic ductal adenocarcinoma cancer (PDAC). State-of-the-art techniques, including Deep Learning (DL) algorithms, are presented for predicting the overall survival of PDAC patients. The survival rates for PDAC patients at one, three, and five years are discussed, revealing the need for more accurate predictive methods.

To enhance the understanding of cancer progression and aid pathologists in patient management, various models are implemented for PDAC Computed Tomography (CT) image segmentation. These models include U-Net and Light GBM (LGBM) and Intelligent Automated Pancreas Segmentation using U-Net Model Variants. The segmentation accuracy achieved through masking and boundary classes is described, offering valuable insights for improved patient diagnosis and treatment.

Furthermore, a hybrid classification model is proposed using VGG16 as the backbone feature extractor and Extreme Gradient Boosting (XGBoost) as the classifier. This model demonstrates superior performance, achieving high accuracy and weighted F1 scores for the dataset under study.

The thesis also introduces Stacked Ensemble Deep Learning (SEDL), a pipeline for classifying pancreas CT medical images. By combining multiple models, such as Inception V3, VGG16, ResNet34, and an XGBoost-based stacking ensemble, SEDL achieves an impressive 98.4% ensemble accuracy.

The Cancer Imaging Archive (TCIA) public access dataset, consisting of 80 pancreas CT scans, is used for training and testing the models. The study's results are instrumental in diagnosing PDAC using the TNM staging system, providing valuable insights into tumour, node, and metastases classification.

This thesis advances cancer research by employing cutting-edge techniques to enhance early detection, improve survival rates, and provide accurate predictions for PDAC patients based on CT images.

Contents

List of Figures

List of Figures

List of Tables

List of Tables

List of Acronyms

PDAC	pancreatic ductal adenocarcinoma cancer
DL	Deep Learning
CT	Computed Tomography
NLM	Non-Local Means
AoU	Intersection over Union
LGBM	Light GBM
MA	Model Analyser
IAM	Intelligent Automation Model
XGBoost	Extreme Gradient Boosting
SEDL	Stacked Ensemble Deep Learning
TCIA	The Cancer Imaging Archive
TNM	tumour, node (N), and metastases (M)
TCIA	The Cancer Imaging Archive
SEDL	Stacked Ensemble Deep Learning
TCIA	The Cancer Imaging Archive
TNM	tumours (T), node (N), and metastases (M)
AI	artificial intelligence
CT	Computed Tomography
MRI	Magnetic Resonance Imaging

PET	Positron Emission Tomography
IPL	Macroscopic cystic
MCN	mucinous cystic neoplasm
PanIn	Pancreatic intraepithelial neoplasia
ASIR	age-standardized incidence rate
IA	Intelligent Automated
CNN	convolution neural networks
NLP	Natural Language Processing
RNN	recurrent neural networks
(NLP)	Natural Language Processing
RNNs	Recurrent neural networks
SE	Squeeze-and-Excitation
GVT	gross tumour volume
IPL	Macroscopic cystic
MCN	mucinous cystic neoplasm
ASIR	age-standardised incidence rate
DECT	Dual-energy computed tomography

Chapter 1

General Introduction

1. 1 Introduction

Pancreatic cancer is one of the world's most dreadful cancers, ranked eighth in Europe in 2020 [1]. The poor prognosis is primarily due to a late diagnosis and limited treatment options. Surgical resection is a potentially curative treatment, particularly for early-stage tumours. However, early-stage tumours are usually asymptomatic and only become symptomatic when surrounding tissues or distant organs are invaded. Furthermore, pancreatic cancer has multiple genomics alterations associated with tumour progression and a complex tumour micro-environment [2], promoting treatment resistance. It is critical to identify causative risk factors and develop effective diagnostic tools for early cancer detection in the target population, to improve the prognosis of diagnosed patients and reduce the disease burden of pancreatic cancer.

Artificial intelligence (AI) transforms medicine by detecting illnesses earlier, improving prognoses, and providing more effective treatment plans [3]. Algorithms capable of reading pancreatic CT or histopathology slides will manage work lists for doctors, provide decision support for clinicians without speciality training, and power AI-powered telehealth services. Deep Learning, a type of artificial intelligence in which neural networks learn patterns directly from raw data, has shown remarkable success in image classification. Medical AI research has expanded in areas such as radiology, pathology, and ophthalmology, which rely heavily on image interpretation [4]. AI tools have progressed from testing to deployment, overcoming regulatory challenges and gaining administrative approval.

The Centres for Medicare and Medicaid Services [5] have enabled the use of AI in clinical settings. However, there still needs to be a significant gap between the number of deep learning algorithms that

have been successful at medical image interpretation and the number of deep learning algorithms that have been translated into clinical practice [6].

Deep Learning's recent breakthrough has made tremendous progress in computer vision and medical image processing, becoming a key artificial intelligence (AI) technique [7]. Deep Learning has opened up new possibilities for intelligent healthcare systems, such as disease screening, cancer diagnosis, prognosis prediction, and pathology. However, obstacles still need to be overcome to achieve satisfactory results in a real-world scenario due to the shortcomings of current deep learning algorithms, as medical images differ from natural images in various ways. Furthermore, prior medical knowledge is required, as is the need to design and implement robust AI systems for medical image analysis [8].

Medical imaging technologies such as Computed Tomography (CT), Magnetic Resonance Imaging (MR), Positron Emission Tomography (PET), mammography, ultrasound, and X-ray have become critical for disease detection, diagnosis, and treatment. Human experts have traditionally performed medical image interpretation, but the process is often time-consuming and prone to human error. Deep learning tools, enabled by artificial intelligence and machine learning, have significantly improved accuracy and speed, allowing them to be used in clinical practice. However, it remains difficult due to the need for labelled data and the wide anatomical variation among patients. This dissertation examines various challenges and suggests new approaches to dealing with them [9].

The pancreas is an organ in the upper abdomen comprising two types of cells: endocrine glands and exocrine glands. The pancreas produces digestive enzymes mixed with pancreatic juice and released into the duodenum by the endocrine pancreas. It comprises small clusters of cells that produce and release essential hormones into the bloodstream. Exocrine pancreatic neoplasms account for approximately 95% of all malignant neoplasms originating in the ductal epithelium, acinar cells, or connective tissues. The most common subtype is pancreatic ductal adenocarcinoma (PDAC) [10], whereas neuroendocrine tumours are frequently benign. The remaining subtypes are colloid carcinomas, Acinar cell carcinoma, solid-pseudopapillary neoplasm, pancreatoblastoma, and other undifferentiated carcinomas [11].

Macroscopic cystic (IPL), mucinous cystic neoplasm (MCN), or microscopic precursor lesions can all cause PDAC (pancreatic intraductal neoplasia, PanIN). Due to the accumulation of genetic and epigenetic alterations, the normal ductal epithelium can progress to a series of precursor lesions and invasive and metastatic carcinoma. Founder mutations found in parental clone cells are thought to accumulate within PanIN and cause invasive cancer, whereas progressive mutations are only found in a subset of metastatic deposits [12]. Pancreatic cancer's genetic heterogeneity makes early detection

and treatment difficult. However, molecular profiling-based classification can define sub-types with distinct molecular characteristics, biological features, and clinical implications, providing a road map for precision medicine [13].

1. 2 Motivation

The global prevalence of pancreatic cancer varies. In 2020, Europe, North America, and Australia/New Zealand had the highest age-standardised incidence rate (ASIR) of pancreatic cancer, while Africa and South-Central Asia had the lowest [14]. Pancreatic cancer is more common in men than women, with a median age of diagnosis of around 70 years. The ASIR of pancreatic cancer increased slightly over time, from 5.0 to 5.7 per 100,000 person-years between 1990 and 2017. The disease burden of pancreatic cancer is projected to increase further in the next two decades due to population growth and aging. Because of the poor prognosis, mortality rates closely parallel incidence rates. Western Europe is expected to have the highest mortality rate in 2020. In both sexes, the mortality rate rises, and nearly 90% of deaths occur after age 55. The cancer statistics are shown in Figure: 1.1.

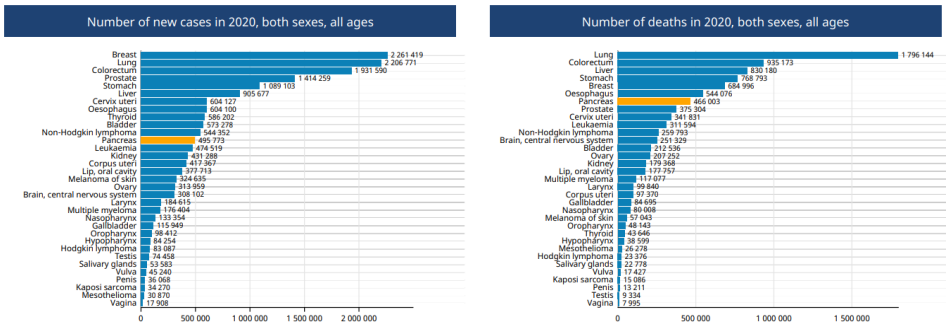
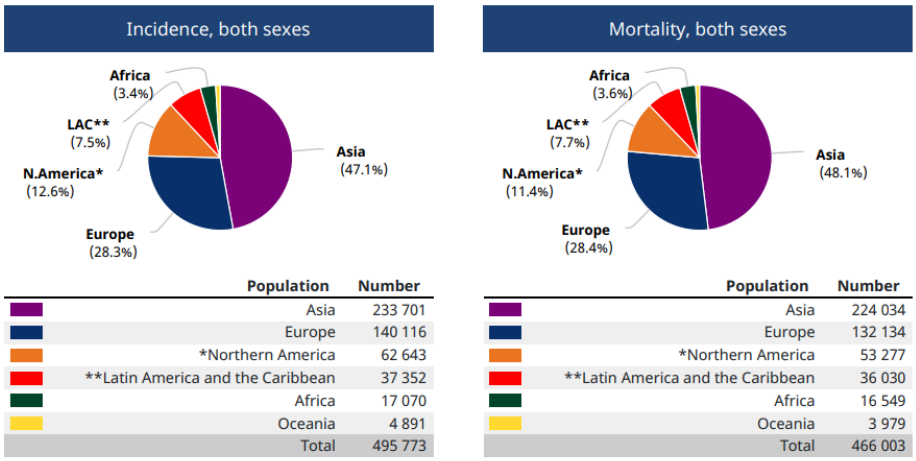


Fig. 1.1: Cancer incidence and mortality statistics worldwide for both men and women.

Source: Globocan 2020

Pancreatic cancer was predicted to kill 155,000 people in Europe (27 European Union countries and the United Kingdom) in 2039 [15]. Compared to other common cancers, pancreatic cancer has the lowest five-year survival rate, which varies little between developed and developing countries. Pancreatic cancer is one of the few cancers in the United States, with a five-year survival rate of 20% or less, Figure: 1.2. Pancreatic cancer survival has improved little since the 1970s compared to most common cancers.



Population	Number
Asia	233 701
Europe	140 116
*Northern America	62 643
**Latin America and the Caribbean	37 352
Africa	17 070
Oceania	4 891
Total	495 773

Population	Number
Asia	224 034
Europe	132 134
*Northern America	53 277
**Latin America and the Caribbean	36 030
Africa	16 549
Oceania	3 979
Total	466 003

Fig. 1.2: Cancer incidence and mortality statistics worldwide and by region.

Source: The Global Cancer Observatory - All Rights Reserved, December 2020

The symptoms of pancreatic cancer are non-specific and largely dependent on tumour location and stage. Because tumours in the head and tail can obstruct the common bile duct and/or the pancreatic duct, they are more likely to be diagnosed earlier than those in the body and tail. Abdominal pain, unexplained weight loss, nausea, vomiting, and new-onset diabetes are common presenting symptoms [16]. The difficulty in imaging, lack of circulating biomarkers, inaccessibility of the pancreas to biopsy, and difficulty in defining sufficiently high-risk populations impede early-stage detection of pancreatic cancer.

1. 3 Problem Statement

The lack of good screening techniques to identify such cancers in their earliest stages is a major challenge in treating pancreatic cancer. The pancreas is in an anatomically unfavourable position for early diagnosis, towards the back of the abdomen. Pancreatic cancer has a low survival rate of 9% or less when diagnosed at an advanced stage. Around 40% of patients' tumours have progressed to the point where they attach to or encompass local structures. Subdivided into borderline tumours, which, while technically treatable, necessitate chemotherapy and radiation therapy before surgery to ensure complete removal [17].

Locally advanced tumours, which surround critical blood vessels or infiltrate adjacent critical organs, cannot be surgically removed in most cases. The remaining pancreatic cancers are already metastatic, spreading to other body parts. Almost all long-term pancreatic cancer survivors are diagnosed when their cancer is surgically removable or can be made so. In contrast, very few five-year survivors present with Stage IV disease due to limited treatment options and inherent treatment resistance [18].

When pancreatic adenocarcinoma is suspected, it has grown to the point where surgical removal is difficult. Critical vascular and other structures complicate surgical excision. Alternatively, it has spread to distant locations.

Furthermore, long before a doctor diagnoses pancreatic cancer, there is often the development of what is known as "microscopic metastatic disease", which implies that cancer cells have already spread to other parts of the body. Chemotherapy and radiation [19] are used before and after surgery to kill such stealth tumour cells. Despite these treatments, most patients whose tumours are surgically removed will die due to recurrence caused by these remaining tumour cells.

1. 4 Thesis Objectives

The main aim of the research is to improve and analyse the survival prediction of pancreatic cancer cases through accurate semantic segmentation of the pancreas and Deep Learning techniques. The survival analysis has three primary goals:

1. estimating and interpreting survival and hazard functions from survival data
2. comparing survival or hazard functions
3. assessing the relationship of explanatory variables to survival time

Survival analysis is a powerful tool for analysing time-to-event data, which is common in clinical trials. Researchers hesitate to use it because they need more confidence in the theory underlying its application and interpretation.

Survival prediction models are frequently used to answer questions such as "What is the probability that this patient will be alive in 5 years, given their baseline co-variate information?" at some baseline time, such as the time of diagnosis or treatment. Clinicians can then use this predicted probability to make important patient care decisions, such as increasing monitoring frequency or implementing specific therapies [20].

The specific research objectives of this thesis are:

1. To explore the significance of early cancer detection for improving survival rates, particularly focusing on pancreatic ductal adenocarcinoma cancer (PDAC).
2. To model a Deep Learning-based framework for accurate prediction of the overall survival of pancreatic ductal adenocarcinoma cancer patients.

3. To enhance the understanding of cancer progression and aid pathologists in patient management by implementing various PDAC Computed Tomography (CT) image segmentation models.
4. To model a hybrid framework for accurate pancreatic cancer classification using VGG16 as the backbone feature extractor and Extreme Gradient Boosting (XGBoost) as the classifier.
5. To advance cancer research by employing cutting-edge techniques to enhance early detection, improve survival rates, and provide accurate predictions for PDAC patients based on CT images.

1. 5 Thesis Contribution

Ensemble modelling is predicting an outcome using multiple diverse base models. The goal of using ensemble models is to reduce prediction generalisation errors. When the ensemble approach is used, the prediction error decreases as long as the base models are diverse and independent. In making a prediction, the approach seeks the wisdom of the crowds. Even though the ensemble model contains multiple base models, it acts and performs as a single model. Ensemble learning is a broad machine learning meta-approach that combines predictions from multiple models to improve predictive performance [21].

Ensemble learning techniques have been shown to improve machine learning performance. These methods can be applied to both regression and classification problems. The final prediction is obtained by combining the results of several base models. Some methods for combining the results are averaging, voting, and stacking [22]. Ensemble learning uses multiple machine learning models to solve a single problem. These models are referred to as "poor learners". The assumption is that if several weak learners are combined, they can become strong learners. Each weak learner is fitted to the training set and provides the results of the predictions. The final prediction result is calculated by aggregating the results of all the weak learners.

Ensemble models solve the technical challenges of developing a single estimator [23]. The following are the technical challenges associated with developing a single estimator:

1. High variance: The model is extremely sensitive to the inputs used to learn the features.
2. Low accuracy: Using a single model (or algorithm) to fit all of the training data may not provide you with the nuance you require for your project.
3. Noise and bias in features: The model makes a prediction based on far too few features.

A single algorithm may not produce the best prediction for a given data set. Machine learning algorithms have limitations, and producing a high-accuracy model is difficult [24]. We can improve

overall accuracy by creating and combining multiple models.

The combination of models is then put into action by aggregating the output of each model with two goals in mind:

1. Minimising model error;
2. Keeping the model's generalisation intact;
3. Improve predictions - Rather than using a single model, it will achieve better predictive skills.
4. Improve robustness - rather than using a single model, it will achieve more stable predictions.

When employing ensemble techniques, the overall goal should be to reduce prediction generalisation error. As a result, using a diverse set of base models will reduce the prediction error automatically [25].

Contributions made include:

1. Enhanced early PDAC detection: Presents state-of-the-art techniques for predicting the overall survival of PDAC patients using deep learning (DL) algorithms. This could lead to earlier diagnosis and treatment, improving patient outcomes.
2. Improved interpretability: Use of Light Gradient-boosting Machine (LGBM) with cropping to segment PDAC images. This makes the model more interpretable, which can help radiologists understand how it makes its predictions.
3. Use of UNet, a fully convolutional neural network based on ResNet34, as an encoder and turning point for boundary class separation using Watershed, in PDAC image segmentation for biomedical applications.
4. Model Analyser (MA) and Intelligent Automation: The Model Analyser (MA) and the Intelligent Automation Model in the Inference Engine assist in choosing the best model and evaluating the selected models. By leveraging these tools, we can efficiently analyze the performance of each model based on the defined factors, facilitating the decision-making process.
5. Hybrid model for PDAC diagnosis: Proposes a novel approach for PDAC diagnosis, combining the VGG16 deep learning architecture with the XGBoost classifier.
6. SEDL an ensemble learning pipeline for pancreas CT image classification: Proposed a novel ensemble learning pipeline for pancreas CT image classification, which combines the Inception V3, VGG16, and ResNet34 deep learning architectures with the XGBoost classifier.

1. 6 Thesis Outline

The remainder of this thesis is organised as follows: Chapter 2 presents the anatomy of the pancreas, causes, screening, and treatment of PDAC. We also looked at ensemble methods that can be implemented in medical imaging. Chapter 3 examines the literature review presented in the state-of-the-art pancreatic cancer survival prediction paper. Chapter 4 investigates the benefits of applying UNet and Watershed combined and investigates Intelligent Automated (IA) pancreas segmentation using U-Net model variants. Chapter 5 looks at implementing VGG16 as a feature extractor for XGBoost, a classifier for pancreatic cancer prediction, and a stacked ensemble DL for improved pancreatic cancer classification using XGBoost. The results discussion in Chapter 6, which reviews some of the technologies in use, concludes this study and proposes possible areas for extending this work.

1. 7 Conclusion

Pancreatic cancer is a devastating disease with a poor prognosis. Early diagnosis and treatment are key to improving survival rates, but early detection is often challenging due to the lack of symptoms in early-stage disease. Artificial intelligence (AI) can potentially revolutionise the early detection and treatment of pancreatic cancer. AI algorithms can be trained to identify pancreatic cancer on imaging scans and tissue samples with high accuracy, even in early-stage disease. This could lead to earlier diagnosis and treatment, improving survival rates. In addition to early detection, AI can also be used to improve the treatment of pancreatic cancer. AI algorithms can personalise treatment plans based on a patient's tumour genetics and characteristics. This could lead to more effective and targeted treatment, improving survival rates.

Deep learning (DL), a type of AI, has shown remarkable success in image classification, which is particularly relevant in medical fields such as radiology, pathology, and ophthalmology. The use of DL in pancreatic cancer is still in its early stages, but the potential benefits are significant. DL can potentially revolutionise the early detection and treatment of pancreatic cancer, improving patient survival rates.

Chapter 2

Anatomy of the Pancreas

2. 1 Introduction

After the stomach, the pancreas is located in the upper abdomen. Across the back of the stomach, behind the stomach, is where the pancreas is found [26]. It is an extended, tapered organ. In addition to producing and secreting hormones into the blood to regulate energy consumption and storage throughout the body, the pancreas is an endocrine organ that produces and secretes digestive enzymes into the intestine as part of the digestive system. Because it plays both exocrine and endocrine roles, the pancreas is a special organ [27].

2. 1.1 Exocrine pancreas

A part of the small intestine that produces and secretes digestive enzymes contains connective tissue, blood vessels, and nerves that are present together with the acinar and duct cells. Proteins, lipids, and carbs can all be broken down by the exocrine gland in the pancreas's secretions of enzymes. In an inactive state, these enzymes pass from the pancreatic duct into the bile duct. They become operational when they enter the duodenum. Moreover, bicarbonate is secreted by the exocrine tissue. Inside the duodenum, it balances stomach acid. A system of ducts connecting to the main pancreatic duct receives these enzymes as secretions. The entire pancreas is covered by it. About all of the pancreatic mass comprises exocrine components [28].

2. 1.2 Endocrine pancreas

The islets of Langerhans generate and secrete the hormones insulin, glucagon, somatostatin, and pancreatic polypeptide, which control the metabolism of glucose, lipids, and proteins in the blood

[29]. The other hormones' levels are managed in part by somatostatin. Of the total pancreatic mass, islets make up 1-2%.

2. 2 Anatomical Structure

The pancreas is typically divided into five parts [30]:

- The head of the pancreas is the broadest portion on its right side. It is surrounded by connective tissue and situated within the duodenum's C-shaped curve.
- The uncinate process is a medially extended protrusion that emerges from the lower half of the head and lies beneath the pancreatic body. The superior mesenteric vessels it is located behind them.
- The neck of the pancreas is situated between its head and body. It covers the superior mesenteric vessels, visible in a groove on its backside.
- Body - The tapered left side of the body crosses the midline, lying below the stomach and to the left of the superior mesenteric arteries, and it is somewhat angled upward in the centre
- The left pancreatic tail, located close to the spleen's hilum, is called the tail. Along with the splenic vessels, it is housed in the splenorenal ligament. The whole pancreas is not intraperitoneal except for this one portion.

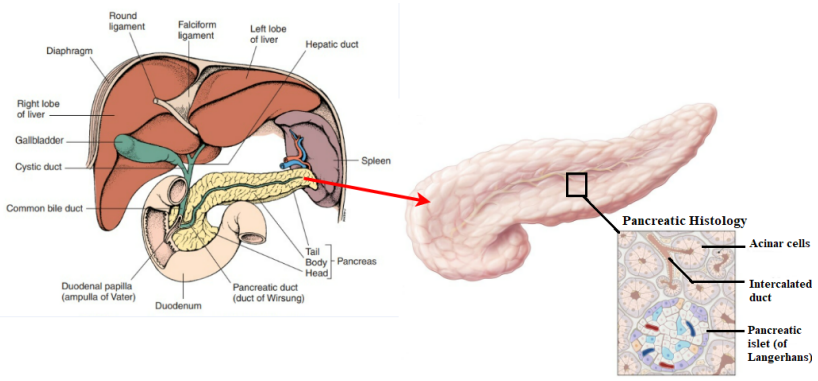


Fig. 2.1: Pancreas location and histology

The pancreatic arteries, which originate from the superior mesenteric, gastroduodenal, and splenic arteries, supply the pancreas' body and tail. Blood that has lost its oxygen is removed from the pancreas through pancreatic veins [31]. Unlike the posterior variant, which empties into the hepatic portal vein, the anterior superior pancreaticoduodenal vein drains into the superior mesenteric vein. The major contributor is the splenic artery. The pancreatic veins drain venous blood from the body,

and the tail discharges into the splenic vein. In contrast, the anterior and posterior inferior pancreaticoduodenal veins drain into the superior mesenteric vein.

2. 3 Duct System

A lobulated, serous gland that creates precursors to digesting enzymes, the exocrine pancreas is so named. It comprises a million acini, or short intercalated ducts, connecting about one million 'berry-like' cell clusters. The intercalated ducts join with those draining neighbouring lobules to form a network of intralobular collecting ducts, which ultimately drain into the main pancreatic duct [32]. The hepatopancreatic ampulla of Vater comprises the common bile duct and pancreatic duct, which are connected along the length of the pancreas. After that, the main duodenal papilla on this structure allows access to the duodenum. The sphincter of Oddi, a muscle valve, regulates the flow of secretions into the duodenum. Acting as a valve, it encircles the ampulla of Vater.

2. 4 Pancreatic ductal adenocarcinoma (PDAC)

The pancreas is prone to malignant and non-cancerous tumours, among other growths. The cells that line the ducts that remove digestive enzymes from the pancreas are the first to become the most prevalent type of pancreatic cancer (pancreatic ductal adenocarcinoma) [33]. Exocrine tumours make up the majority of pancreatic malignancies. These tumours manifest themselves in the ducts and glands that manufacture and transport the digestive enzymes from the pancreas to the intestines.

Adenocarcinoma represents the classification for approximately 95% of exocrine pancreatic tumours. PDAC is the name given to most of these malignancies, which form in the linings of ducts. Aggressiveness is a PDAC trait. The neurological system, blood vessels, liver, and gastrointestinal tract are frequently affected [34].

The exocrine region of the pancreas hosts around 95% of malignant pancreatic neoplasms, with the most prevalent sub-type (90%) beginning in the ductal epithelium, acinar cells, or connective tissues. Neuroendocrine tumours, mostly benign (5%), are the second most frequent neoplasms. The remaining subtypes are pancreatoblastoma, colloid carcinomas, acinar cell carcinomas, solid pseudopapillary neoplasms, and various undifferentiated carcinomas [35].

PDAC is a condition that may arise from microscopic or macrocytic (IPN) antecedent lesions (MCN). The accumulation of genetic and epigenetic changes in normal ductal epithelium can lead to a sequence of precursor lesions, invasive carcinoma, and metastatic carcinoma. Mutations in parental clone cells, such as KRAS, p16/CDKN2A, TP53, and SMAD4, are thought to build up inside PanIN

and result in invasive cancer. Due to genomic abnormalities in tumour tissues and features of the stromal milieu, the genetic variability of pancreatic cancer presents difficulties for early detection and efficient treatment [36]. Precision medicine can be mapped out using molecular profiling-based classification, which can be used to establish sub-types with unique molecular traits, biological traits, and clinical implications.

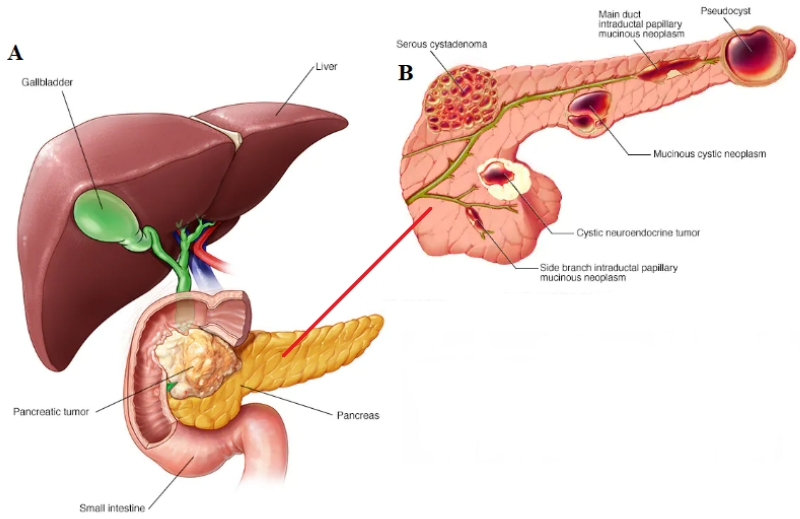


Fig. 2.2: Pancreas location: A - pancreas tumour and B - different tumour locations.

2. 4.1 Symptoms

Depending on the location and stage of the tumour, pancreatic cancer symptoms might vary widely. Due to their potential to clog the common bile duct or pancreatic duct and create obvious symptoms such as jaundice and obstructive cholestasis, tumours in the head (60%–70%) are more likely to be discovered earlier than those in the body and tail (20%–25%).

Often, pancreatic cancer does not show any early warning signs and symptoms until it has spread to more serious stages [37]. They may include:

- Abdominal pain that radiates to the back
- Appetite loss or unintentional weight loss
- Jaundice, which is a yellowing of the skin and whites of the eyes
- Light-coloured faeces
- Dark-coloured urine
- Itchy skin

- Blood clots
- Fatigue
- A new diabetes diagnosis or diabetes that is already present but becomes harder to control

2. 4.1.1 Complications

As pancreatic cancer progresses, it can cause complications [38] such as:

1. **Losing weight:** Several things could contribute to weight loss when a person has pancreatic cancer. Since the malignancy uses up the body's energy, weight loss may occur. Eating may be challenging if experiencing nausea and vomiting due to cancer therapies or a tumour pressing on the stomach. Alternatively, if the pancreas is not producing enough digestive juices, the body may have trouble breaking down the nutrients from food.
2. **Jaundice:** Jaundice can result from pancreatic cancer that restricts the liver's bile duct. The yellow complexion, eyes, dark urine, and light faeces are warning signs. Abdominal pain rarely coexists with jaundice.
3. **Pain:** The abdomen's nerves may become compressed when a tumour grows, resulting in potentially life-threatening agony. Radiation and chemotherapy treatments may reduce discomfort and inhibit the growth of the tumour.
4. **Obstruction of the bowels:** The movement of digested food from the stomach into the intestines can be obstructed by pancreatic cancer that invades or presses on the duodenum, the first segment of the small intestine.

2. 4.2 Risk Factors

Understanding the causes of pancreatic cancer, preventing it, and creating affordable early diagnostic techniques for high-risk people depend on identifying risk factors. Age, male sex, family history of pancreatic cancer, smoking, chronic pancreatitis, diabetes, and obesity are a few well-established risk factors. Even though the aetiology of pancreatic cancer has been thoroughly researched, its causes remain poorly understood [39].

Cases of pancreatic cancer are infrequent and often run in families [40]. Hereditary cancer or chronic inflammatory disorders account for about 3% of all cases. A susceptibility to develop pancreatic cancer that can be inherited by having three or more relatives of any degree or at least two first-degree relatives with the disease is present in 7% of cases.

The most well-known risk factor for pancreatic cancer is tobacco use [41], and the likelihood of developing the disease rises with higher smoking frequency, duration, and cumulative dosage. Both somatic mutations (such as those in KRAS and TP53) and systemic inflammation are brought on by tobacco-related carcinogens, such as nitrosamines and polycyclic aromatic hydrocarbons.

Chronic pancreatitis is a chronic pancreatic inflammatory condition resulting in fibrosis, the loss of islet and acinar cells, DNA damage, and the emergence of PanINs [42]. Chronic pancreatitis, which has an incidence of four to fourteen per 100,000 person-years, is primarily brought on by heavy alcohol usage. Diabetes is a risk factor for pancreatic cancer and a side effect of the disease, according to studies that have established a substantial correlation between chronic pancreatitis and pancreatic cancer. A high body mass index and a large waist-to-hip ratio are associated with an increased risk of PANIC cancer in those who are obese.

2. 4.3 Diagnosis and Screening

Pancreatic cancer is symptomless up to an advanced stage, making treating and prolonging lives possible. Today, a diagnosis involves a comprehensive physical examination, imaging, lab tests, and experimental procedures. Starting with a primary care physician, pancreatic cancer is diagnosed. The procedures' foundation will be a physical examination and considering health and family history. A bloated abdomen or jaundice are two symptoms they will look for as possible indicators of pancreatic cancer [43].

They could advise blood and urine testing to assess general health and seek out cancer-related symptoms. If cancer is present, the doctor can recommend imaging tests from a specialist. Pancreatic cancer can be identified using a variety of imaging tests [44]. They include:

1. Computed Tomography (CT) scans
2. Magnetic Resonance Imaging (MRI).
3. Positron Emission Tomography (PET) scans use positron emission tomography.
4. Abdominal X-rays
5. Laparoscopy, in which a camera is introduced through a cut in the abdomen
6. Endoscopic Ultrasound
7. Endoscopic Retrograde Cholangiopancreatography

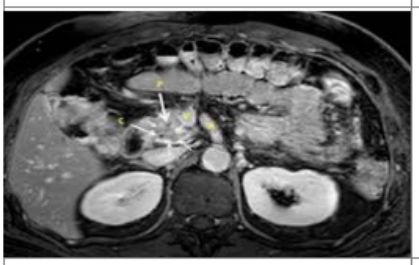
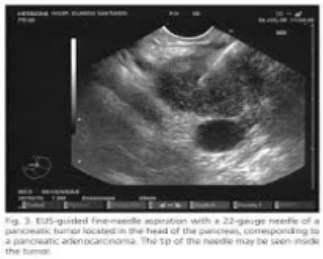
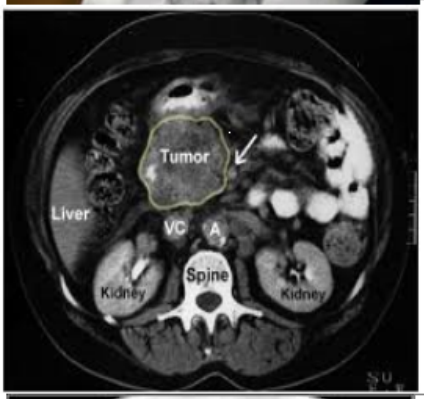
IMAGE	EXPLANATION
	**Nuclear medicine:** covers both diagnostic imaging and treatment of disease, and may also be referred to as molecular medicine or molecular imaging & therapeutics. Nuclear medicine uses certain properties of isotopes and the energetic particles emitted from radioactive material to diagnose or treat various pathology
	**Ultrasound:** uses sound waves rather than ionizing radiation. High-frequency sound waves are transmitted from the probe to the body via the conducting gel, those waves then bounce back when they hit the different structures within the body and that is used to create an image for diagnosis
	**X-ray imaging:** {the most frequently used imaging type. X-rays are a form of electromagnetic radiation. Works on a wavelength and frequency that we're unable to see with the naked, human eye, but can penetrate through the skin to create a picture of what's going on beneath. X-rays can be used to detect cancer through mammography and digestive issues through barium swallows and enemas
	**Computer tomography (CT) scans:** are a form of X-ray that creates a 3D picture for diagnosis and uses X-rays to produce cross-sectional images of the body. The CT scanner has a large circular opening. the patient to lie on a motorised table. The X-ray source and a detector then rotate around the patient producing a narrow beam of X-rays that passes through a section of the patient's body to create a snapshot which are ordered into one, or multiple images of the internal organs and issues. CT scans provide greater lucidity of the internal organs within the body
	**Magnetic Resonance Imaging (MRI):** uses a strong magnetic fields and radio waves to generate images of the body Commonly used to examine internal body structures to diagnose strokes, tumours, spinal cord injuries, aneurysms and brain function. An MRI scanner uses a strong magnetic field to ordinate hydrogen nucleus (proton) from water molecule in the body, and a radio frequency is then utilized

Fig. 2.3: Different modalities used in medical imaging. Image sources: [45]–[49] respectively.

2. 4.4 Treatment

At its earliest stages, when it is most treatable, pancreatic cancer is rarely found since it frequently does not manifest symptoms after progressing to other organs. Options for treating pancreatic cancer are based on age, overall health, and how far PDAC has advanced. Radiation therapy, chemotherapy, surgery, or a combination of these options [50].

Surgery is the best action when cancer is limited to the pancreas or has not migrated too far outside. Chemotherapy is typically administered next. The percentage of PDAC patients who qualify for surgery is only 15% to 20%.

Prevention is better than cure. Pancreatic cancer risks may be reduced by: Quit using tobacco. Make an effort to quit smoking. Discuss stopping methods with a doctor, like support groups, medications, and nicotine replacement therapy [51].

1. Keep weight in check. Maintain weight if it is at a healthy level. Aim for a 1–2 pound (0.5–1 kilogram) weekly weight loss if you need to lose weight. For optimum weight loss, combine regular exercise with a diet high in fruits, vegetables, and whole grains, all of which should be consumed in smaller portions.
2. Make a point of eating well. A diet rich in whole grains, colourful produce, and cancer-preventing foods may lower the risk of developing the disease.

2. 5 Conclusion

Although PDAC often has a relatively bad prognosis compared to other tumours, experts are still trying to determine the best course of action. Over time, survival rates have sluggishly increased. Just 10% of people survive at least five years after their diagnosis. Early detection and surgical removal of PDAC provide the best prognosis. Nevertheless, only 2% of malignancies (Trusted Source) are found to be at stages 0 or 1. According to research, those undergoing surgery for stage 1 PDAC had a 5-year survival rate of 38.2% as opposed to 2.9% for those who did not.

Nearly 50% of PDAC patients die within 10 to 12 months of diagnosis. If cancer has spread to distant body parts, less than 20% of patients survive for a year. According to a trusted source, the head of the pancreas is the location of PDAC in 60% to 70% instances. They often have a better prognosis than tumours in the body or tail and are typically discovered early.

Chapter 3

Literature Review

3. 1 Introduction

In recent years, medical research and healthcare have witnessed a remarkable surge in the application of advanced technologies and data-driven methodologies to enhance the accuracy of disease prognosis and patient outcomes. Predicting pancreatic cancer survival is a particularly challenging yet pivotal endeavour among the many critical focus areas. Pancreatic cancer, characterised by its aggressive nature and often late-stage diagnosis, presents a formidable challenge to patients and healthcare practitioners. In this context, this comprehensive survey titled "Pancreatic Cancer Survival Prediction: A Survey of the State-of-the-Art" delves into the cutting-edge advancements, methodologies, and approaches developed to predict survival rates in individuals afflicted by pancreatic cancer. By exploring the latest innovations in medical imaging, machine learning, biomarker analysis, and other interdisciplinary fields, this survey aims to provide a holistic overview of the current landscape, shedding light on the strides made and the challenges that lie ahead in the pursuit of improving prognostic accuracy and, ultimately, patient outcomes in pancreatic cancer.

3. 2 Pancreatic Cancer Survival Prediction: A Survey of the State-of-the-Art

This chapter analyses the state-of-the-art techniques used in pancreatic cancer survival prediction. The research examines statistical and deep learning (DL) machine learning strategies used in pancreatic cancer patients' overall survival prediction. Genomic and proteomic information, clinical variables, and pathological images are among the features employed for prediction. It is also discussed how

important it is to spot potential flaws in experimental design, data collection from reliable sources, and analysis and validation outcomes because they impact the prognosis of clinical judgements. A model that might aid in calculating the individualised survival prognosis and offer particular treatment options is required.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 1188414, 17 pages
https://doi.org/10.1155/2021/1188414

Review Article

Pancreatic Cancer Survival Prediction: A Survey of the State-of-the-Art

Wilson Bakasa and Serestina Viriri

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Correspondence should be addressed to Serestina Viriri; viriris@ukzn.ac.za

Received 28 May 2021; Revised 24 August 2021; Accepted 18 September 2021; Published 30 September 2021

Academic Editor: Huiling Chen

Cancer early detection increases the chances of survival. Some cancer types, like pancreatic cancer, are challenging to diagnose or detect early, and the stages have a fast progression rate. This paper presents the state-of-the-art techniques used in cancer survival prediction, suggesting how these techniques can be implemented in predicting the overall survival of pancreatic ductal adenocarcinoma cancer (pdac) patients. Because of bewildering and high volumes of data, the recent studies highlight the importance of machine learning (ML) algorithms like support vector machines and convolutional neural networks. Studies predict pancreatic ductal adenocarcinoma cancer (pdac) survival is within the limits of 41.7% at one year, 8.7% at three years, and 1.9% at five years. There is no significant correlation found between the disease stages and the overall survival rate. The implementation of ML algorithms can improve our understanding of cancer progression. ML methods need an appropriate level of validation to be considered in everyday clinical practice. The objective of these techniques is to perform classification, prediction, and estimation. Accurate predictions give pathologists information on the patient's state, surgical treatment to be done, optimal use of resources, individualized therapy, drugs to prescribe, and better patient management.

1. Introduction

Radiologists detect pdac tumours by symptoms of the disease and patient history but not image processing. This paper studies some ML techniques used in image processing to predict pdac overall survival. The anatomy of the pancreas is shown in Figure 1.

Cancer on the pancreas is a challenge because it is difficult to detect. As there are no clear signs in the early stages, it metastasises rapidly, having a poor forecast of the future course of the disease. The early stages of the diseases show no symptoms. In the later stage, the patient will have a lack of appetite and weight loss.

Medical experts recommend knowing the tumour type as each has a different treatment and behaves in different ways. Researchers estimated that 94% of exocrine tumour patients suffer from pdac type [3]. Figure 2 shows a pdac [4]. Tumour size has a significant influence on overall survival rates. A larger tumour is difficult to treat through resection.

Accurate predictions are crucial as they give pathologists information on state of the patient, surgical treatment to be done, optimal use of resources, provision of individualized treatment, drugs to prescribe, and better management of the patient. Features used for prediction include genomics and proteomic data, clinical factors, and pathological images. The need to identify possible weaknesses including experimental design, data collection from valid sources, and the analysis and result validation is a vital step as it affects the prediction of clinical decisions [5, 6].

ML techniques classify pdac patients and learn to predict the best survival period for each patient. Classify pdac stages into low, medium, or high-risk groups. ML techniques have been used to model the progression and treatment of pdac conditions. ML tools are powerful in identifying features from large datasets, which makes them of great use [7].

Big data with the increased patient will stress doctors, making them more error-prone when making critical decisions concerning human life. Machines can manage these

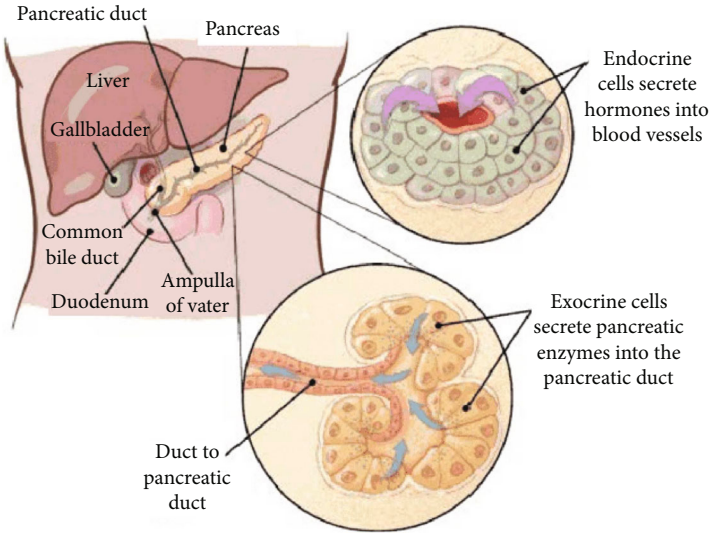


Figure 1: Pancreas [1]. Pancreatic cancer tumours are of two types [2], exocrine and endocrine tumours. The position of origination determines this. The pdac originates from the lining of the ducts in the pancreas.

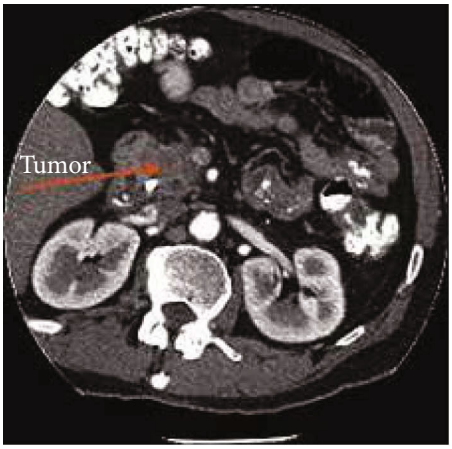


Figure 2: Pancreatic cancer tumour [4].

large sets of imaging data with a lower error rate, attested by the unfitness of medical personnel to colligate and see the big picture from imaging data. Machines can help them by assessing large numbers of image datasets and determine whether there are any patterns suspicious to be cancerous. Machines also can assist by superseding doctors or specialist at times of their absence and provide the diagnosis in even critical cases [8].

Predictive technique models such as ML (statistical multivariate regression and deep learning (DL)) can be used for pdac survival rate prognosis. The techniques have to be reinforced for best predictive performance. Figure 3 shows the steps to follow during the prediction process. This paper will compare imaging techniques used in the medical world, including magnetic resolution imaging (MRI) and computed tomography (CT). These imaging techniques produce images that need to be segmented into pixel classes. In a review of segmentation methods, including those used in deep learning, the current technology is mostly used in overall survival prediction for pancreatic cancer patients. The feature extraction techniques are then looked at, leading to classification after the necessary features to be extracted. Deep learning techniques implemented in feature extraction and classification are then summarised.

2. Medical Imaging Techniques

Medical imaging is a technique and process used to look at the human body, diagnose, monitor, or treat medical statuses, including a visual representation of the functions done by some organs or tissues. Imaging techniques help screen for hidden features before visible symptoms, diagnose the conditions that would have developed into the now visible symptoms, and manage the disease stages or reaction to possible treatment. This help to predict the overall survival of pdac patients. Common imaging techniques are computed tomography (CT) scan and magnetic resonance imaging (MRI).

Blood and other laboratory tests usually determine the existence of pdac. In imaging the pancreas (Figure 4), computed tomography (CT) and magnetic resonance imaging (MRI) were used to help determine if the condition exists. If detected, then determine the stages and reaction to treatment, making it possible to predict the patient's overall survival rate.

3. Segmentation

Segmentation is an image processing proficiency for bisecting an image into multiple regions. The process of segmentation well define semantic entity boundaries in an image, as shown in Figure 5. Pixel in the image is allocated labels to match affiliation according to their semantic properties.

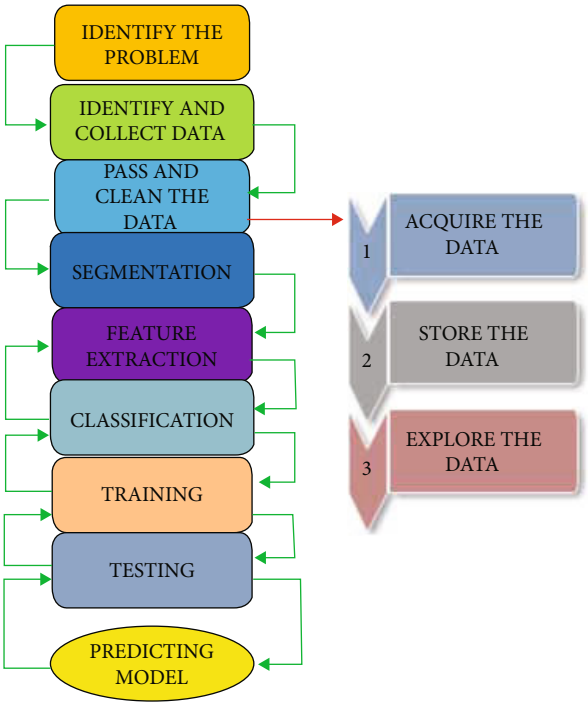


Figure 3: Prediction model process—started by identifying the problem and collecting related data, cleaning, and processing. The output data is passed as input through DL techniques, and a prediction is made.

Segmentation can be defined as a map of the greyscale into the binary set $\{0, 1\}$.

$$s(x,y) = \begin{cases} 0 \text{ if } g(x,y) < T(x,y), \\ 1 \text{ if } g(x,y) \geq T(x,y). \end{cases} \quad (1)$$

Images are segmented using a graph cut with some cost, formed from image pixels. Image pixels of similarity and nearby are put in the same zone [13, 14]. Graph cut T cost (where T is a set of edges) is the sum of the edge weights of the cuts.

$$X_{\text{cut}} = \sum_{(i,j \in T) Z_{ij}}, \quad (2)$$

where Z_{ij} represent edge weight i, j from node i to node j in the graph and cut and T represents all the edges' sum. In graph cut segmentation, a graph image section is zoned such that the cut cost X_{cut} is reduced.

Segmentation of an image M_o is a pair $(\partial\Omega; M)$, where M is some approximation of M_o, where M_o is defined in Ω . The energy associated with a segmentation $(\partial\Omega; M)$ is the sum of three terms:

$$Z(\partial\Omega, M) = \alpha \text{ int} \left(\frac{\Omega}{\partial\Omega} |\nabla M|^2 dx + \beta \text{length}(\partial\Omega) + \text{int} \left(\frac{\Omega}{\partial\Omega} (M - M_o)^2 dx\right). \quad (3)$$

If M is imposed to be constant within each region,

$$Z(\partial\Omega, M) = \alpha \text{length}(\partial\Omega) + \text{int} \left(\frac{\Omega}{\partial\Omega} (M - M_o)^2 dx\right). \quad (4)$$

Gradient operator and Laplacian operator are defined as follows:

(1) *Gradient Operators*. The gradient of an image, $f(x; y)$, at a location $(x; y)$ is defined as the vector.

$$\nabla f(x,y) = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \begin{pmatrix} \frac{\partial t}{\partial x} \\ \frac{\partial t}{\partial y} \end{pmatrix}. \quad (5)$$

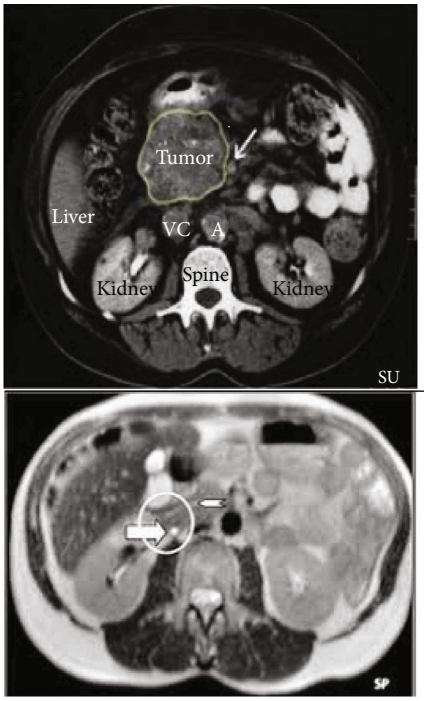
(2) *Laplacian Operator*. The Laplacian of an image $f(x : y)$ is defined as follows:

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2}. \quad (6)$$

Medical images from CT and MRI modalities require segmentation to fix various abnormalities like tumours and cancerous elements. Most methods that use feature-based approaches currently rely on the attributes of features extracted by a human specialist. This poses some challenges as humans are prone to making errors and miss potential features for the segmentation of the image. DL addresses the issue by providing automated feature learning techniques. The mostly used technique in computer vision is convolutional networks [15, 16]. The other used algorithms in computer vision include generative adversarial networks (GNAs), and variational autoencoders (VAEs) are used to solve challenges like the generation of images. The scale-invariant feature transformation (SIFT) is another object detection algorithm used for tumour detection within images regardless of the image rotation, orientation, or scale [17].

Texture features can be used according to [18] and implemented with other methods like the Bacterial Foraging Algorithm (BFA) used for classification. Studies have shown that CAD system, Haar wavelet transform, and clustering methods can be implemented for accurate diagnosis and processing of pdac images obtained from MRI and CT modalities, as shown in Table 1.

Different images call for different segmentation techniques as they have distinct feature and properties as envisioned in terms of colour (greyscale images, binary images, and colour images) and texture (texture images and nontextured images) [23]. Table 2 shows the advantages and disadvantages of some different segmentation techniques used.



Computer tomography (CT) scans: are a form of X-ray that creates a 3D picture for diagnosis and uses X-rays to produce cross-sectional images of the body. The CT scanner has a large circular opening, the patient lie on a motorised table. The X-ray source and a detector then rotate around the patient producing a narrow beam of X-rays that passess through a section of the patient's body to create a snapshot which are ordered into one, or multiple images of the internal organs and issues. CT scans provide greater lucidity of the internal organs within the body

Magnetic resonance image (MRI): uses a stong magnetic fields and radio waves to generate images of the body commonly used to examine internal body structures to diagnose stokes, turmours, spinal cord injuries, aneurysms and brain function. An MRI scanner uses a stong magnetic field to ordinate hydrogen nucleus (proton) from water molecule in the body, and a radio frequency is then utilized

Figure 4: Medical imaging techniques [9–12]: images from CT and MRI modalities.

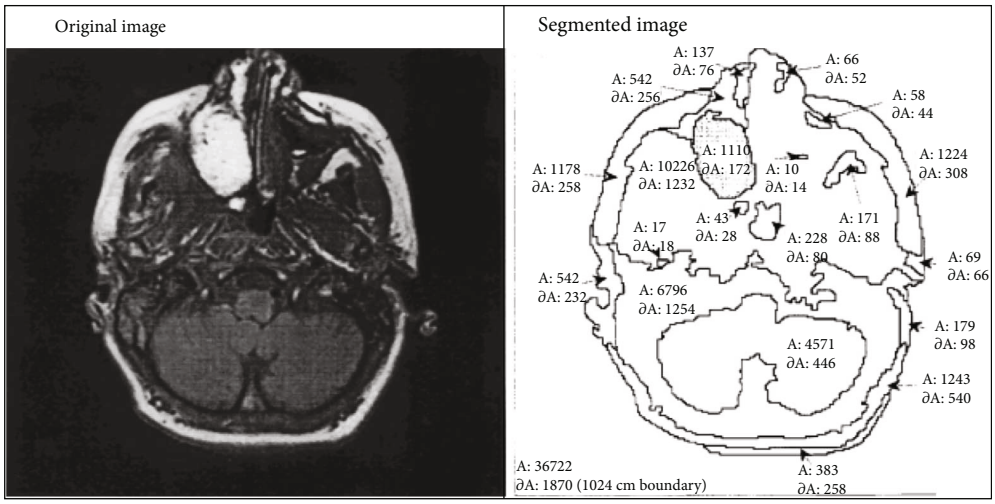


Figure 5: Image bisection [13, 14].

4. Segmentation Using Deep Learning Techniques

Segmentation techniques continue to be developed and improved to solve challenges differently. The evolution of convolutional neural networks has seen most techniques developed, as shown in Tables 3 and 4. Based on architecture or framework, these techniques are grouped into an instance or semantic segmentation [30].

In the semantic segmentation process, a label is assigned to every pixel in the image, which is different from classification, assigning a single label to the entire picture. Semantic segmentation treats multiple objects of the same class as a single entity and involves detecting objects within an image and grouping them based on defined categories.

In the instance segmentation process, multiple objects of the same class are treated as distinct individual objects or instances. It involves detecting objects within defined categories, and it is more complicated than semantic segmentation.

4.1. GoogLeNet [31]. CNN layers are stacked as networks in networks, based on the concept of [32]. GoogLeNet is built from many inception modules of various filter sizes applied to the input and the concatenated outputs. These inception modules and filters are contained within the stem, a standard architecture that simultaneously provides the extraction of

Table 1: pdac tumour segmentation algorithms in medical imaging.

Study	Medical modality	Proposed segmentation method	Results
Tam et al. [19]	CT	Region growing algorithm	Efficient: Jaccard index of 73.37–86.97
Balakrishna et al. [20]	CT	MATLAB	Detecting edges, corners, and points differentiating images
Sindhu et al. [18]	MRI	Texture extraction with BFA	Texture feature extraction with BFA has an accuracy of 89%.
Farag et al. [21]	CT	Cascaded superpixel segmentation	Superpixels preserve more boundaries. Dice coefficient of 70.7% and Jaccard index of 57.9%
Reddy et al. [22]	MRI and CT	K-means clustering and Haar wavelet transform	Based on threshold value with a mean threshold of 8.88

Table 2: Advantages and disadvantages of different segmentation techniques.

Segmentation technique	Advantage	Disadvantage
Region-based [16, 23, 24]	(i) High speed of operation (ii) Efficient for object and background with high contrast (iii) Easier to classify and implement (iv) Best when easy to define region similarities (v) Less sensitive to noise compared to edge detection	(i) Poor segments if there is low greyscale (ii) Are by nature sequential and quite expensive both in computational time and memory (iii) Region growing has inherent dependence on the selection of seed region and the order in which pixels and regions are examined
Fuzzy theory-based [25–27]	(i) The single fuzzy rule applied to stress the importance attached to feature-based and spatial information in the image (ii) Structure of the membership functions and associated parameters automatically derived	(i) Sensitive to noise (ii) Computationally expensive (iii) The determination of fuzzy membership is not very easy
Artificial network-based [15]	(i) Simple programming (ii) Make use of neural net parallelism	(i) Long training time (ii) Initialization could influence the outcome
Generalized PCA (principle component analysis) [28, 29]	(i) Low noise sensitivity (ii) Lack of redundancy of data (iii) Increased efficiency (iv) Reduced overfitting	(i) Independent variables become less interpretable (ii) Data standardization is a must before PCA (iii) Information loss

image features at varying levels of detail. GoogLeNet does not use fully connected layers, instead uses global average pooling, thus lessening model parameters required.

4.2. ResNet [31]. Train much deeper networks making use of skip connections. The authors trained a 152-layer deep ResNet, and they also successfully trained a version with 1001 layers. Combining the power of skip connections in addition to the standard pathway allows the network to copy activations from every ResNet block or layer, maintaining information during processing through layers. Various features are finest assembled in external networks, as some need extra depth. The skip connections facilitate both simultaneously, improving the network's flexibility when given input data. Skip connections allow the network to learn residuals, thereby giving ResNets an advantage to perform a boosting.

4.3. U-Net [33, 34]. U-Net is a very popular and successful network for segmentation in 2D images. When fed an input image, it is first downsampled through a "traditional" CNN before being upsampled using transpose convolutions until it reaches its original size. In addition, based on the ideas of ResNet, there are skip connections that concatenate features from the downsampling to the upsampling paths. It is a fully convolutional network, using the concepts first introduced in [35].

4.4. V-Net [34]. V-Net is a three-dimensional version of U-Net with volumetric convolutions and skip connections as in ResNet.

5. Feature Extraction Techniques

A technique of transmutation to exhume common characteristics and shape in an image, extracting the image properties that differentiate it from a range of images. CT and MRI images are complicated, and methods of feature extraction application on them are limited. Image features can be considered the fundamental attributes or prominent features for realizing the image. As in Table 5, feature extraction methods use four image features, texture features, colour image features, shape features, and spatial relations [46–48]. In [49], the author classified feature extraction into topological and

Table 3: Segmentation techniques: dice coefficient (DC), image dimension (ID), application (App), reference (Ref), convolutional neural network- (CNN-) conditional random field (CRF), named entity recognition (NER), maximum entropy Markov models (MEMM), and hidden Markov model (HMM).

Method	Advantages	Disadvantages	Dataset	DC	ID	Modality	App	Ref
CNN-CRF [36, 37]	(i) Flexible enough in terms of feature selection (ii) Better for NER than MEMM and HMM	(i) High computational complexity of the algorithm training stage algorithm (ii) Difficulty in retraining the model when new training data is available	Data in hospitals	86.0	2D	MRI	Brain tumour	Feng et al.
U-Net [33, 38]	(i) Provides pixel-accurate semantic segmentation (ii) It is fast to compute (iii) Its architecture is easy to understand	(i) Significant memory requirement as lower level features have to be stored for further concatenation in the upsampling phase	1245 CT images	73.6	3D	CT	Pulmonary nodules	Tong et al., Du et al.
E-Net [39]	(i) Significantly faster (ii) Provides a high frame rate for real-time applications (iii) Small storage requirements alleviating the need for model compression	(i) The use of convolutional layer factorization increases the number of kernel calls making each of them smaller	Achieva scanner (Philips Healthcare, Best, The Netherlands) with a pelvic phased-array coil (8 channel HD Torso XL)	90.9	3D	MRI	Prostate	Comelli et al.
V-Net [40]	(i) Fully CNN and suitable for volumetric medical image segmentation (ii) The residual function is learnt	(i) Location information is lost in the compression path	PROMISE 2012 Challenge	82.39	3D	MRI	Prostate	Milletary et al.

Table 4: Segmentation techniques: DeepLab V3, SegNet, and Fast Convolutional Network (FCN).

Method	Advantages	Disadvantages	Dataset	DC	ID	Modality	App	Ref
DeepLab V3 [41, 42]	(i) Allows us to enlarge the fields of view of filters to incorporate large context (ii) Preserves spatial information	(i) Has no postprocessing step conditional random fields (ii) Does not scale well for large or deeper layers if GPU memory is limited	CHAOS	81.0	3D	CT and MRI	Kidney	Guo et al.
SegNet [43]	(i) Low memory requirement during both training and testing (ii) Improved boundary delineation (iii) Reduced number of parameters enabling end to end training	(i) Both input image and output segmentation have fixed resolution	OASIS	91.47	3D	MRI	Brain	Khagi et al.
FCN [44, 45]	(i) Ability to make predictions on arbitrarily sized inputs (ii) End to end trainable fast and improved performance	(i) Direct predictions are typically in low resolution resulting in fuzzy object boundaries (ii) Suitable mainly for object detection, not object classification (used for local rather than global tasks)	DRIVE	95.33	3D	Funduscopy	Retinal vessels	Cai et al.

geometric features, statistical features, and series expansion features and global transformation.

Many researchers have studied the detection of skin cancer, breast cancer, and brain cancer. Different algorithms have been successfully applied for the early diagnosis of these tumours. Artificial neural networks (ANNs) have been implemented for detecting skin cancer [50]. They preprocessed the image to remove noise and enhance the image. The wavelet transform was used for feature extraction, and then, the tumour was classified using backpropagation neural networks.

The features must depict the objects correctly for classification. Shape features and feature extraction methods can both be based on either boundary characteristics or regional characteristics. The methods of shape feature extraction are boundary-based and region-based. Shape feature extraction lacks a mathematical model, and if changed, the result is not reliable, and accuracy depends on the presegmentation effects [49, 51–55].

Wavelet decomposition is used for brain tumour detection. The grey-level cooccurrence matrix (GLCM) is used for feature extraction, while probabilistic neural networks are used for further classification [56]. Another algorithm for brain tumour detection is using an artificial neural network fuzzy inference system [57]. Various supervised learning techniques are used to detect breast cancer. These include principal component analysis (PCA) for feature extraction and support vector machines (SVMs) and k-nearest neighbour for classification [5]. As seen above, the various parametric and nonparametric classifiers can be used to classify and detect different tumours based on the tumour's features.

Dimension reduction is used in machine learning and statistics to reduce the number of feature set, categorized into feature selection and feature extraction [58, 59]. Feature extraction is referred to as dimensionality reduction [60]. The input matrix W, of dimension $A \times B$, is

$$\begin{bmatrix} W_{11} & \cdots & W_{1B} \\ \vdots & W & \vdots \\ W_{A1} & \cdots & W_{AB} \end{bmatrix}, \tag{7}$$

where we have a sample representation by rows and variables represented by columns. The feature extraction algorithm will learn from this transformation the prerequisite features needed for extraction.

Image texture analysis is currently used to predict tumour heterogeneity. However, only a few studies have reported texture analyses of tumour heterogeneity in pdac [61].

Edge detection highlights the contrast or difference in the intensity of an image. This detection draws attention to the boundaries of features within an image, the same way a human vision can perceive the perimeter of an object, which is of different contrast to its environment. Fundamentally it is the boundary of an image that varies in intensity levels or the contrast [13, 14, 62]. The edge is at the location of the variation and is detected by differentiation of first order. Intensity variation is demonstrated by varying adjacent pixels. If computed on image M, the horizontal detector edge creates an intensity variation between two horizontally adjacent points, as such detecting the vertical edges, $\mathrm{VE}(x)$, as follows:

Table 5: Feature extraction advantages and disadvantages.

(a)

Method	Description (texture features)	Advantage	Disadvantage
Gabor wavelet transform [52, 55, 58]	In information theory applications, Dennis Gabor used complex functions to build wavelets forming a basis for Fourier transforms.	Multiscale robust	Incomplete cover of spectrum plane needs rotation normalization
GLCM-based method [48]	Find the frequency of a set of pixel and its spatial relationship in an image to characterize its texture.	Easy to use compact robust	High computation cost partial description of texture

(b)

Method	Description (colour features)	Advantage	Disadvantage
Scalable Colour Descriptor [48]	Defined in the hue-saturation-value (HSV) colour space with fixed colour space quantisation and uses a novel, Haar transform encoding	Compact and robust never changing and uninterrupted	Needs postprocessing for spatial information
Colour histogram [47, 48, 53]	A histogram represents the dispersion of colours in an image. It can be visualised as a graph that gives a high level of suspicion regarding the pixel value distribution	Simple to compute, easy to use and understand	High dimension sensitive to information

(c)

Method	Description (shape features)	Advantage	Disadvantage
Wavelet transform [47]	Mathematical means for performing signal analysis when the signal frequency varies over time.	Translation and scale invariant good affine transformation good noise resistance	Average occultation resistance
Zernike moments [48]	A set of rotation invariant features is introduced. They are the magnitude of a set of orthogonal complex moments of the image.	Good noise resistance	High computational complexity bad affine transform

$$\mathrm{VE}_{X_{xy}} = \left|M_{xy} - M_{x+1,y}\right| \forall x \in 1, N-1\,; y \in 1, N. \quad (8)$$

A vertical edge detector that varies adjacent vertical points is needed to detect horizontal edge using vertical edge detectors, $\mathrm{VE}(y)$, as follows:

$$\mathrm{VE}_{X_{xy}} = \left|M_{xy} - M_{x,y+1}\right| \forall x \in 1, N\,; y \in 1, N-1. \quad (9)$$

We can then combine the two forming an operator $\mathrm{VE}(y)$ that detect both vertical and horizontal edges. That is,

$$\mathrm{VE}_{X_{xy}} = \left|M_{xy} - M_{x+1,y} + M_{xy} - M_{x,y+1}\right| \forall x, y \in 1, N-1, \quad (10)$$

which gives

$$\mathrm{VE}_{X_{xy}} = \left|2 \times M_{xy} - M_{x+1,y} - M_{x,y+1}\right| \forall x, y \in 1, N-1, \quad (11)$$

given the coefficients of a varying template which can be twisted together with an image to detect all the edge points.

Grey-level cooccurrence matrix (GLCM) and histogram analyses are methods employed by texture analyses to classify objects according to their texture. Evaluation of grey level placement, regularity, and coarseness within the damaged or abnormal change in the pancreatic tissues, computer vision, and machine learning algorithms can analyze CT and MRI images to provide other morphological details related to pdac tumour heterogeneity.

6. Survival Prediction Model Review

Traditional diagnosis models like the popular American Joint Committee on Cancer (AJCC) tumour-node-metastasis (TNM) model is used in cancer diagnosis. According to [7, 8], the challenge is that they do not predict the cancer patient's overall survival. The study explores different ML techniques used in pdac survival rate prediction.

ML techniques classify pdac patients and learn to predict the best survival period for each patient. Classify pdac stages into low, medium, or high-risk groups. ML techniques have been used to model the progression and treatment of pdac conditions. ML tools are powerful in identifying features from large datasets, which makes them of great use.

Big data with the increased patient will stress doctors, making them more error-prone when making critical decisions concerning human life. Machines can manage these large sets of imaging data with a lower error rate, attested by the unfitness of medical personnel to colligate and see the big picture from imaging data. Machines can help them by assessing large numbers of image datasets and determine whether there are any patterns suspicious to be cancerous. Machines also can assist by superseding doctors or specialist

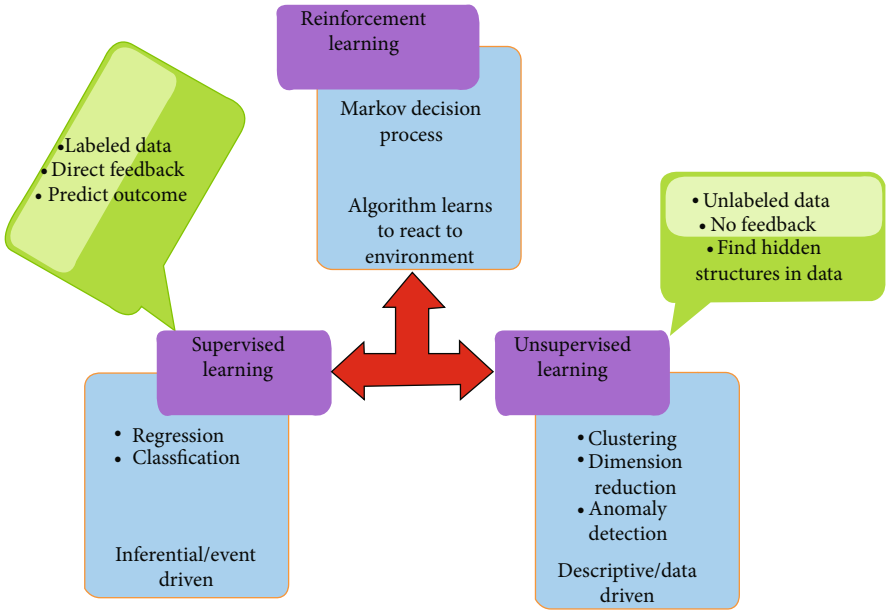


Figure 6: Types of machine learning: supervised learning, unsupervised learning, and reinforcement learning (a hybrid of supervised and unsupervised learning).

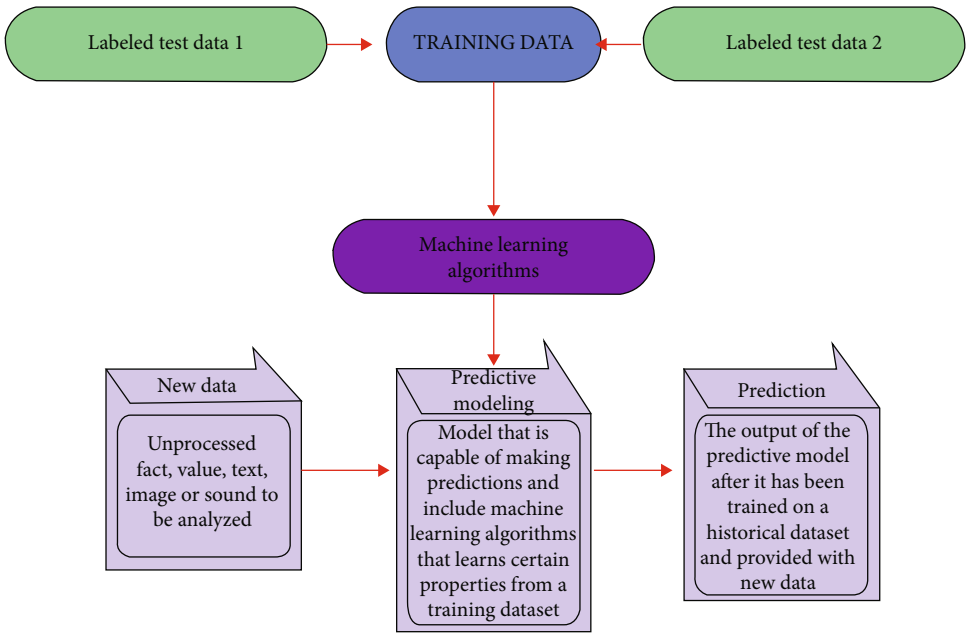


Figure 7: Supervised machine learning: accept training data in labelled test data (CT and MRI images). The trained ML algorithm output plus new data is taken into the predictive model to make pdac patients' survival prediction.

at times of their absence and provide the diagnosis in even critical cases. ML can be either supervised learning or unsupervised learning, as shown in Figure 6.

Supervised learning (Figure 7) is typically implemented in the context of classification, mapping the input to output labels, or regression, for the mapping of input to continuous output.

For both regression and classification, the goal is to find specific relationships or structure in the input clinical data for the correct generation of output patient clinical data, determined entirely from the clinical training data. When carrying on supervised learning, the principal circumstances are model complexness and the bias-variance tradeoff.

Model complexity refers to the complexness of the function sought to learn. Figure 8 shows the data analytics types, indicating the complexity of each type. The proper level of model complexity is determined by the nature of patients' clinical training data, the smallness of data, and distribution throughout different possible assumptions.

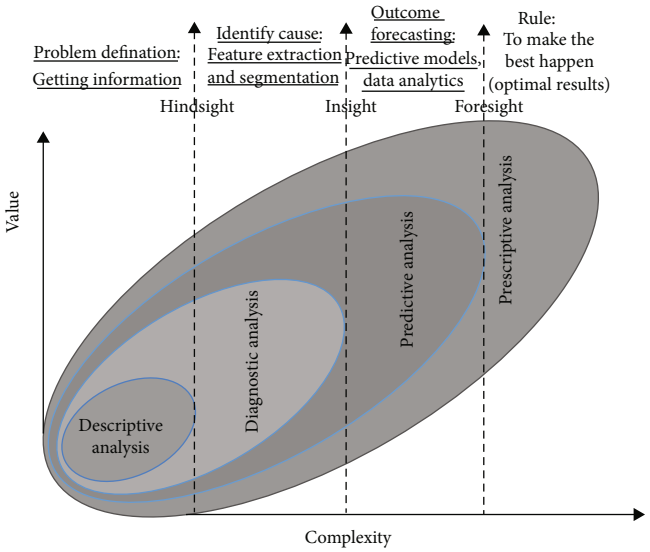


Figure 8: Data analytics types: four types of data analytics—descriptive analysis (problem definition), diagnostic analysis (identify cause), predictive analysis (outcome forecasting), and prescriptive analysis (rules).

Unsupervised learning requires representation learning, density estimation, and clustering to be carried out. We need the inherent structure of our patients' clinical data without using explicitly provided labels. Since no labels are provided, there is no specific way to compare model performance in most unsupervised learning methods. Exploratory analysis and dimensionality reduction are the common use cases for unsupervised learning. Unsupervised learning is powerful for the identification of hidden structures in patients' clinical data. Over the years, survival prediction models in cancer were developed implementing techniques like Decision Trees (DTs), support vector machines (SVMs), Bayesian Networks, and artificial neural networks (ANNs). There is a need to validate these machine learning methods used in progression analysis of cancer for them to be used in medical practices.

ML has been implemented in cancer prognosis or overall survival prediction [63]. Most of the studies proposed for the last years focused on developing predictive models using supervised ML methods to predict disease outcomes [64] accurately. Based on the analysis of their results, it can be concluded that the integration of multidimensional heterogeneous data, with the application of different models for feature selection and patients' data classification, can help to model useful tools for overall survival prediction for pdac patients [65].

Many evolutionary algorithms have been implemented into resolving challenges to do with feature selection and classification to analyze gene expression data. Genetic algorithms [65–67] are implemented for creating selectors with each allele is linked to one gene and having a state to tell if selected or not. Genetic programming is practically applicable to find hidden features in complex datasets. It is good for identifying rule-based classifiers and gene expression profiling from medical data.

A random survival forests method for the analysis of right-censored survival data was suggested by [68]. New survival splitting rules for growing survival trees and a conservation-of-event principle for survival forests are proposed to define an interpretable measure of mortality used as a predicted outcome.

Some papers proposed the use of Feed Forward Neural Networks. ANNs [69] were implemented for two years on patients and the network predictions for twelve months of death compared to surgical doctors. ANNs achieved a 95% prediction accuracy [32].

The techniques can detect features from complex datasets. According to [66], genetic programming can make an automatic feature selection. They showed that genetic programming performs substantively better than SVM, multilayered perceptron, and random forests in classifying. As reported by [70], the result indicated frequency usage of ANN with high accuracy in survival prediction of any malignancy. However, they suggested combining ANN and fuzzy logic with 93% accuracy as superior and powerful.

A short review of current algorithms being used in ML such as SVM [71–73], Naïve Bayes [74], Logistic Regression [75], genetic algorithms, Decision Tree, ANN, and KNN algorithm is done. A graph-based semisupervised learning paradigm that takes advantage of both unlabelled and labelled data when training a model is used. Other semisupervised learning (SSL) are self-training, cotraining, and transductive SVM.

According to the study [76], for lung cancer, the study chooses Linear Regression, Decision Trees, Ensemble Learning algorithm, random forest, and Gradient Boosting Machines as logistic-based methods. SVM was then used to sum the predictions of each of the five models into final predictions. The study by [77] concluded that classification and regression trees (CARTs) are also used compared to ANN as prediction modes. ANN proves significantly more accurate than the CART model. The comparison of three different

techniques is in [78], SVM, Decision Trees, and k-nearest neighbour. SVM proved to be the best performer.

However, in [79], the extracted features are used as inputs for Back Propagation NN and Logistic Regression (LR) and the two algorithms are compared for accuracy. LR was the best given a higher number of features. Random forest [80] was used as a biomedical classifier that has led to the proposal of two variants, namely, Forest-RK and dynamic random forests [81].

In survival prediction for cancer [35, 82–84], they used convolutional neural networks (CNNs) for classification and feature extraction with the aid of computer-aided diagnosis (CAD) based on pathological images. Tumours can be characterized through the implementation of supervised ML [85] for labelled data, and usage of SVM as state-of-the-art classification algorithms. Experiments to prove state-of-the-art ML techniques can be efficiently or equally better than traditional techniques (Linear and Logistic Regression) were done [86]. All algorithms executed using Weka ML workbench. ML algorithms used are Linear Regression, ZeroR, and Logistic Regression. For classification, the algorithms they used are Naive Bayes, J4.8 (C4.5 learning algorithm), k-nearest neighbor, OneR, Locally Weighted Learning, and Bayesian Nets.

Colon cancer predictions [87] implemented a supervised classification technique. Synthetic Minority Oversampling Technique (SMOTE) was used to balance the survival and nonsurvival classes. Ensemble Voting of three classifiers was found to result in the best prediction performance in prediction accuracy and area under the receiver operating characteristic (ROC) curve. The study by [88] proposed a semisupervised model in trying to address data scarcity in cancer datasets. Ensemble classifiers used to learn unlabelled data. CART tools helped to deal with missing attribute values by using the surrogate splitting technique. ML techniques can be categorized into the following:

(1) Statistical techniques

(2) Deep learning (DL)

6.1. Statistical Techniques. According to [3, 63], Weibull distribution is implemented to estimate the overall survival. It is the widely used technique for statistical cases involving data with periods, which considers life behaviour. The author states that alpha is the rendering probability of 63.2% prediction that an event occurs. Beta is used to represent the probability of growing (beta > 1) and to decrease (beta < 1). If nearing one, then the distribution is exponential with an invariable chance rate.

The study [8, 89] used the cox proportional hazard to assess the independent effects of prognosis factors. It was also implemented to assess the correlation between tumour node metastases (TNMs) [63]. Song et al. [8] used a Student's *t*-test or chi-square test to compare disease features. The Kaplan-Meier was used for the assessment of survival factors. The rank test was used to test survival curve variation. They argued that their model is capable of providing quantitative prognosis to individual cancer patients. Song et al. and Hang et al. [8, 89] designed a graphical nomogram using the R statistical package using statistical analysis. The c-index was applied to validate the predictive accuracy of the nomogram.

6.2. Deep Learning. Deep learning is a subset of ML, in which algorithms learn unsupervised from unstructured or unlabelled data. Deep neural networks (Figure 9) are designed based on biological neural networks with many hidden layers to extract features from data using interconnected node matrix to imitate how the human brain works.

A node is the basic unit of an artificial neural network. The input feature set is multiplied by corresponding weights using mathematical functions, passing the output to the next node layer. These outputs are weighed up to known facts. It autoadjusts the weights using errors as feedback to minimise future errors during iterations [90, 91].

Deep learning rose to its prominent position in computer vision when neural networks started outperforming other methods on several high-profile image analysis benchmarks.

The main common characteristic of deep learning methods is their focus on feature learning: automatically learning data representations. Discovering features and performing a task are merged into one problem and therefore both improved during the same training process [31].

Tumour detection in pancreas images is achieved using CNN, which extracts the image features, classifying the tumour based on the extracted features. Various classifiers used for feature extraction in machine learning include Naive Bayes classification, support vector machine, and logistics functions. CNN delivers the information needed using convolution methods in three steps: convolution, pooling, and padding, which are layers in the input and output images. Region Convolution Neural Network (R-CNN) is used to extract image features from a particular region. The CNN will be acting as a feature extractor.

[92] suggested deep learning and described it as in the unsupervised learning category. Segmentation using CNN is mainly used for image processing in machine learning. It requires providing segments of the pancreas image to the convolution neural network as the inputs. CNN labels the pixel and classifies each pixel determining the context to identify the images. A segment represents the tumours in the pancreas image or parts or the superpixels. Analysis of the images is carried out in three levels: classification, tumour detection, and segmentation. Segmentation significantly detects the tumour, classifying them into their different classes.

6.2.1. Deep Learning Frameworks and Image Processing Platforms. Developing complex machine learning models for image processing requires special platforms and frameworks. Some of the popular frameworks are TensorFlow and PyTorch.

(1) TensorFlow [93]. Google developed TensorFlow, which is an open-source framework with support for machine learning and deep learning. TensorFlow facilitates the creation and training of custom deep learning models. The framework has

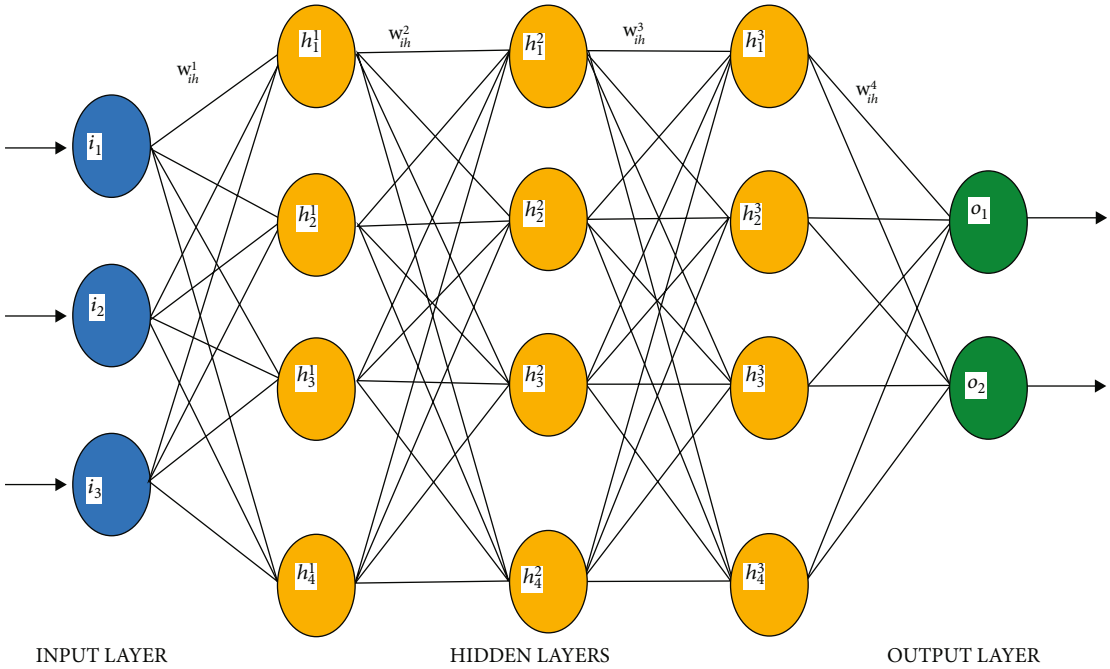


FIGURE 9: A deep neural network basic illustration. New data features (input layer (i)) are examined by many interconnected nodes (hidden layers (h)) that form the hidden layer generating the output (output layer (o)). The weights are represented by w where $n = 4$.

a set of libraries, including for image processing projects and computer vision applications.

(2) PyTorch [94]. PyTorch is an open-source deep learning framework designed by the Facebook AI Research lab (FAIR). This Torch-based framework has Python, C++, and Java interfaces. PyTorch is used for building computer vision and image processing applications.

6.2.2. Neural Networks in Image Processing. Researching on neural networks has been done for many years, which has seen improvements in machine learning, a reason for the magnificent progress in medical imaging and computer vision technology today. Most successful machine learning models for image processing implement neural networks and deep learning. Examples include Mask R-CNN and fully convolutional networks.

(1) Mask R-CNN. Mask R-CNN [95] is a Faster R-CNN-based deep neural network that separates tumours in a processed image. This neural network performs segmentation and generates masks and bounding boxes. The neural network is adjustable, flexible, and efficient as compared to other techniques. Mask R-CNN is poor in real-time processing, as the neural network is massive and the mask layers slow performance, mainly if compared to Faster R-CNN. For instance, segmentation, Mask R-CNN is a very efficient technique.

(2) Fully Convolutional Network (FCN). FCN [40] was developed by University of Berkeley researchers. CNN has a convolutional layer rather than FCN, which has a regular, fully connected layer. This difference enables FCN to manage various input sizes. Also, FCNs use downsampling and upsampling to reduce computational costs for convolution functions.

7. pdac Survival Prediction

There is a possibility that delay in diagnosing pdac can cause concentrated resectable tumours to develop into unresectable by the time of diagnosis [96]. They estimated that it might take just over one year for an average T1-stage pdac to develop into a T4-stage tumour. CT/MRI scans taken by radiologists for other medical purposes but not focused on the pancreas are useful for screening pdac at a reduced cost, time, or radiation exposure [97]. Deep learning has proven to serve as a tool at the disposal of a radiologist in computer-aided diagnosis that points to minor changes on the pancreas that may result in abnormalities that health experts could miss.

A few studies have evaluated the detection of pdac tumours using deep learning techniques [98]. According to [99], they study CT scans using deep learning networks from 303 pdac patients and 136 normal test data. The results showed the detection of pdac had 94.1% sensitivity and 98.5% specificity. The study by [100] used 370 patient CT scans suffering from pdac and 320 test data to study the effectiveness of deep networks in pdac tumour detection. A 98.8% accuracy, 99.3% specificity, and 98.3% sensitivity were achieved. The study proved that a deep network, if implemented in detecting pdac, is more sensitive than radiologists. The results indicated that about 91.7% of tumours missed by radiologists could be correctly classified by deep network

and paying attention to tumours less than 2 cm in size, achieving a 92.1% sensitivity.

The advantage is that deep learning has over traditional methods, as the networks can adapt automatically and modernise features from big data instead of already added features. Deep learning is effective in that it can receive new complex pancreas image feature representation quickly [92].

Convolutional neural networks (CNNs) are a category of deep learning networks developed precisely for image processing. The networks have neurons that mimic neurons in the human brain. CNNs need fewer preprocessing operations than other neuron networks, and the networks learn the required filters and characteristics during training rather than using hand-engineered filters. They are multilayered neural networks with layers organised in three dimensions: weight, height, and depth. They have two components: feature extraction and classification [17, 98, 100].

Fully convolutional network (FCN) is suitably implemented in image segmentation tasks when the neural network splits the processed image into many pixels to be labelled and classified. Popular examples of FCNs for semantic segmentation are RefineNet and DeepLab.

U-Net is a convolutional neural network that allows for fast and precise image segmentation. U-Net was designed particularly for the segmentation of complex tasks in biomedical image processing. U-Net is built with U-shaped architecture with more feature channels in its upsampling part so that the network propagates context information to higher-resolution layers [34].

In feature extraction, CNN runs many convolutions and pooling functions to detect features used for image classification. The network algorithm predicts the tumour in the pancreas image with a calculated probability in the classification component.

Techniques employing deep learning are used in the prediction and prognosis of pdac development. They focus on three major domains: prediction of cancer susceptibility, prediction of pdac relapse, and prediction of pdac survival rate. This paper focuses on the third case, which predicts several possible parameters characterizing pdac development like survival time, life expectancy, and progression. The overall survival rate and the pdac relapse mostly depend on the medical treatment and the quality of the diagnosis [101].

8. Dataset

According to the study [102], they used the Cancer Imaging Archive (TCIA) Public Access dataset, consists of 3D CT scans of 512 × 512-pixel resolution from 53 males and 27 female subjects of the 18-76 age group. Liu et al. [100] implemented CNN to the Taiwanese Centre dataset with contrast-enhanced CT images of 370 patients with pdac and 320 controls. These datasets can be used to study the prediction of pdac overall survival using CT images and ML or deep learning techniques.

9. Period for Prediction

According to [63], they considered one year, three years, and five years, but [87] reduced the three years to two. While in [65], high or low was used. Follow up for five years [32] and predict death within nine, twelve, fifteen, eighteen, twenty-one, and twenty-four months. However, they differ with [7] as they used six, twelve, and twenty-four months. Short periods of survival times like less than six months give better accurate results when developing a prediction model [103]. Some authors [89] used three risk groups with median overall survival of 11.7, 7.0, and 3.7 months. Short- and long-term classes were used by [104, 105]. According to [77], five-year survival was output prediction with zero for dead and one for alive and classified as good, intermediate, or poor.

10. Validation Methods

In the study of Kourou et al. [65], the methods used for evaluating the performance of a classifier are the holdout method, bootstrap, random sampling, and cross-validation. According to [8, 89], they validated their nomogram using discrimination and calibration and using bootstrap resampling. Discrimination between survival probability and actual observation was evaluated using c-index. Calibration plot was constructed to determine the concordance of predicted survival and actual survival.

Tenfold cross-validation methods were employed [76] to measure the unbiased estimate of the prediction models (DTs, ANN, and SVM). Nevertheless, [104] called it k-fold cross-validation, depending on the factor (k) used. According to [105], they used a leave-one-out-cross-validation (LOOCV) protocol.

11. Conclusion

This paper reviews machine learning techniques implemented in overall survival prediction of pancreatic cancer patients, that is, statistical and DL methods. ML methods have significantly proved to be effective if applied to the overall survival prediction of pdac. Features used for prediction include genomics and proteomic data, clinical factors, and pathological images. The need to identify possible weaknesses in experimental design, data collection from good sources, and the analysis and validation results is vital as it affects the prediction of clinical decisions. There is a need for a model that could help in the individualized survival prediction calculation and provide specific treatment decisions. Based on the reviews done on most studies, we have found that for the integration of various feature extraction, segmentation, and classification, DL techniques can provide a useful prediction tool best to make accurate predictions that will assist pathologists in making informed decisions.

Implementing frameworks and platforms like PyTorch and TensorFlow in developing computer vision and image processing significantly improves the overall survival prediction for pdac patients. Various neural networks are deployed to solve various image processing tasks, from binary classification to instance segmentation. Selecting the proper type and architecture of a neural network is important in creating an efficient machine learning-based image processing solution. The most used neural networks in DL are CNN and

FCN. The U-Net model is largely implemented in image processing tasks and has shown a high degree of accuracy.

The authors suggest reinforcing different DL techniques in order to come up with an efficient pdac predictive model. An end-to-end model that will use DL techniques to predict the overall survival rate of pdac patients is proposed.

Data Availability

No data has been used in this review research paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] *What is pancreatic cancer*, 2021, https://www.cancer.org/cancer/pancreatic-cancer/about/what-is-pancreatic-cancer.html.

[2] Q. Zhang, L. Zeng, Y. Chen et al., "Pancreatic cancer epidemiology, detection, and management," *Gastroenterology Research and Practice*, vol. 2016, Article ID 8962321, 10 pages, 2016.

[3] A. C. Lockhart, M. L. Rothenberg, and J. D. Berlin, "Treatment for pancreatic cancer: current therapy and continued progress," *Gastroenterology*, vol. 128, no. 6, pp. 1642–1654, 2005.

[4] J. Shah, S. Surve, and V. Turkar, "Pancreatic tumor detection using image processing," *Procedia Computer Science*, vol. 49, pp. 11–16, 2015.

[5] A. Osareh and B. Shadgar, "Machine learning techniques to diagnose breast cancer," in *2010 5th international symposium on health informatics and bioinformatics*, pp. 114–120, Ankara, Turkey, 2010.

[6] A. Safiyari and R. Javidan, "Predicting lung cancer survivability using ensemble learning methods," in *in 2017 Intelligent Systems Conference (IntelliSys)*, pp. 684–688, London, UK, 2017.

[7] M. H. Osman, "Pancreatic cancer survival prediction using machine learning and comparing its performance with TNM staging system and prognostic nomograms," in *AACR Annual Meeting 2019*, Atlanta, GA, 2019.

[8] W. Song, D.-L. Miao, and L. Chen, "Nomogram for predicting survival in patients with pancreatic cancer," *Oncotargets and Therapy*, vol. 11, pp. 539–545, 2018.

[9] "Modalities," 2020, https://www.medicalimaging.org/about-mita/medical-imaging-primer/.

[10] "Imaging explained," 2020, https://www.nps.org.au/consumers/imaging-explained.

[11] P. Nordbeck, G. Ertl, and O. Ritter, "Magnetic resonance imaging safety in pacemaker and implantable cardioverter defibrillator patients: how far have we come?," *European Heart Journal*, vol. 36, no. 24, pp. 1505–1511, 2015.

[12] *Diagnostic imaging*, 2019, https://www.who.int/diagnostic_imaging/en/.

[13] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer Science & Business Media, 2010.

[14] L. Ross and J. C. Russ, "The image processing Handbook, Sixth Edition, John C. Russ. CRC Press, Boca Raton FL, 2011, 972 pages. ISBN 1-4398-4045-0(Hardcover)," *Microscopy and Microanalysis*, vol. 17, no. 5, p. 843, 2011.

[15] K. K. Gupta, N. Dhanda, and U. Kumar, "A comparative study of medical image segmentation techniques for brain tumor detection," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1–4, Greater Noida, India, 2018.

[16] H. Gao, L. Dou, W. Chen, and G. Xie, "The applications of image segmentation techniques in medical ct images," in *Proceedings of the 30th Chinese Control Conference*, pp. 3296–3299, Yantai, China, 2011.

[17] N. O'Mahony, S. Campbell, A. Carvalho et al., "Deep learning vs. traditional computer vision," in *Science and Information Conference*, pp. 128–144, Springer, 2019.

[18] A. Sindhu and V. Radha, "Pancreatic tumour segmentation in recent medical imaging–an overview," in *International Conference On Computational Vision and Bio Inspired Computing*, pp. 514–522, Springer, 2019.

[19] T. D. Tam and N. T. Binh, "Efficient pancreas segmentation in computed tomography based on region-growing," in *International Conference on Nature of Computation and Communication*, pp. 332–340, Springer, 2014.

[20] R. Balakrishna and R. Anandan, "Soft computing analysis for detection of pancreatic cancer using MATLAB," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 18, pp. 379–392, 2018.

[21] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, and R. M. Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 386–399, 2017.

[22] C. K. K. Reddy, G. Raju, and P. Anisha, "Detection of pancreatic cancer using clustering and wavelet transform techniques," in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 332–336, Jabalpur, India, 2015.

[23] N. Shareef, D. L. Wang, and R. Yagel, "Segmentation of medical images using legion," *IEEE Transactions on Medical Imaging*, vol. 18, no. 1, pp. 74–91, 1999.

[24] A. Kumar, "A method of segmentation in 3d medical image for selection of region of interest (ROI)," in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, pp. 1–5, Bhimtal, India, 2018.

[25] D. Gupta and R. Anand, "A hybrid edge-based segmentation approach for ultrasound medical images," *Biomedical Signal Processing and Control*, vol. 31, pp. 116–126, 2017.

[26] H. Costin, "A fuzzy rules-based segmentation method for medical images analysis," *International Journal of Computers Communications & Control*, vol. 8, no. 2, pp. 196–205, 2013.

[27] R. B. Jeyavathana, R. Balasubramanian, and A. A. Pandian, "A survey: analysis on pre-processing and segmentation techniques for medical images," *International Journal of Research and Scientific Innovation (IJRSI)*, vol. 3, 2016.

[28] Y. Ma, S. Sastry, and R. Vidal, "Generalized principal component analysis," *Interdisciplinary Applied Mathematics*, vol. 43, pp. 22-23, 2015.

[29] C. L. Chowdhary and D. Acharjya, "Segmentation and feature extraction in medical imaging: a systematic review," *Procedia Computer Science*, vol. 167, pp. 26–36, 2020.

[30] J. Ruiz-Santaquiteria, G. Bueno, O. Deniz, N. Vallez, and G. Cristobal, "Semantic versus instance segmentation in

microscopic algae detection," *Engineering Applications of Artificial Intelligence*, vol. 87, article 103271, 2020.

[31] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift four Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.

[32] F. E. Ahmed, "Artificial neural networks for diagnosis and survival prediction in colon cancer," *Molecular Cancer*, vol. 4, no. 1, p. 29, 2005.

[33] Y. Yang, C. Feng, and R. Wang, "Automatic segmentation model combining U-Net and level set method for medical images," *Expert Systems with Applications*, vol. 153, pp. 113419–113419, 2020.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[35] K. Dmitriev, J. Marino, K. Baker, and A. E. Kaufman, "Visual analysis of a computer-aided diagnosis system for pancreatic lesions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, 2019.

[36] N. Feng, X. Geng, and L. Qin, "Study on mri medical image segmentation technology based on cnn-crf model," *IEEE Access*, vol. 8, pp. 60505–60514, 2020.

[37] B. Yu and Z. Fan, "A comprehensive review of conditional random fields: variants, hybrids and applications," *Artificial Intelligence Review*, vol. 53, pp. 4289–4333, 2019.

[38] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoderdecoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[39] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: a deep neural network architecture for real-time semantic segmentation," 2016, https://arxiv.org/abs/1606.02147.

[40] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, Stanford, CA, USA, 2016.

[41] M. Mittal, M. Arora, T. Pandey, and L. M. Goyal, "Image segmentation using deep learning techniques in medical images," in *Advancement of machine intelligence in interactive medical image analysis*, pp. 41–63, Springer, 2020.

[42] C.-W. Xie, H.-Y. Zhou, and J. Wu, "Vortex pooling: improving context representation in semantic segmentation," 2018, https://arxiv.org/abs/1804.06242.

[43] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, and M. A.-M. Salem, "Liver tumor segmentation in CT scans using modified SegNet," *Sensors*, vol. 20, no. 5, p. 1516, 2020.

[44] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International journal of multimedia information retrieval*, vol. 7, no. 2, pp. 87–93, 2018.

[45] T.-y. Ko and S.-h. Lee, "Novel method of semantic segmentation applicable to augmented reality," *Sensors*, vol. 20, no. 6, p. 1737, 2020.

[46] P. L. E. S. Soumya and P. Balan, "Survey on feature extraction techniques in image processing," *International Journal for Research in Applied Science Engineering Technology (IJRASET)*, vol. 6, no. 3, pp. 217–222, 2018.

[47] M. M. Sahu, A. Saxena, and M. Manoria, "Application of feature extraction technique: a review," *International Journal of Computer Science and Information Technologies*, vol. 4, pp. 3014–3016, 2015.

[48] N. Gupta and V. A. Athavale, "Comparative study of different low level feature extraction techniques for content based image retrieval," *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, vol. 1, 2011.

[49] D. P. Tian, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385–396, 2013.

[50] J. A. Jaleel, S. Salim, and R. Aswin, "Artificial neural network based detection of skin cancer," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 1, no. 3, 2012.

[51] S. A. Medjahed, "A comparative study of feature extraction methods in images classification," *International Journal of Image, Graphics and Signal Processing*, vol. 7, no. 3, pp. 16–23, 2015.

[52] R. Rouhi, M. Amiri, and B. Irannejad, "A review on feature extraction techniques in face recognition," *Signal & Image Processing*, vol. 3, no. 6, p. 1, 2012.

[53] K. K. Pachouri, "A comparative analysis & survey of various feature extraction techniques," *International Journal of Computer Science and Information Technologies*, vol. 6, pp. 377–379, 2015.

[54] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Advances in Bioinformatics*, vol. 2015, Article ID 198363, 13 pages, 2015.

[55] W. S. H. M. W. Ahmad and M. F. A. Fauzi, "Comparison of different feature extraction techniques in content-based image retrieval for CT brain images," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pp. 503–508, Cairns, QLD, Australia, 2008.

[56] P. John, "Brain tumor classification using wavelet and texture based neural network," *International Journal of Scientific & Engineering Research*, vol. 3, no. 10, pp. 1–7, 2012.

[57] M. Sharma, "Artificial neural network fuzzy inference system (ANFIS) for brain tumor detection," 2012, https://arxiv.org/abs/1212.0059.

[58] J. Babu, S. Rangu, and P. Manogna, "A survey on different feature extraction and classification techniques used in image steganalysis," *Journal of Information Security*, vol. 8, no. 3, pp. 186–202, 2017.

[59] A. Latif, A. Rasheed, U. Sajid et al., "Content-based image retrieval and feature extraction: a comprehensive review," *Mathematical Problems in Engineering*, vol. 2019, Article ID 9658350, 21 pages, 2019.

[60] Y.-H. Liu, C.-H. Chen, and P. C.-P. Chao, "Mathematical methods applied to digital image processing," *Mathematical Problems in Engineering*, vol. 2014, Article ID 480523, 4 pages, 2014.

[61] J. Chakraborty, L. Langdon-Embry, K. M. Cunanan et al., "Preliminary study of tumor heterogeneity in imaging predicts two year survival in pancreatic cancer patients," *PloS One*, vol. 12, no. 12, article e0188022, 2017.

[62] M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision*, Academic Press, 2020.

[63] D. Timofte, M. Blaj, F. Petrariu, L. Ionescu, and L. Ochiuz, "Survival prediction for romanian patients with pancreatic cancer," *Journal of Surgery [Jurnalul de Chirurgie]*, vol. 13, no. 2, pp. 59–61, 2017.

[64] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, 2006.

[65] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.

[66] L. Vanneschi, A. Farinaccio, G. Mauri, M. Antoniotti, P. Provero, and M. Giacobini, "A comparison of machine learning techniques for survival prediction in breast cancer," *BioData Mining*, vol. 4, no. 1, p. 12, 2011.

[67] M. J. van de Vijver, Y. D. He, L. J. van 't Veer et al., "A gene-expression signature as a predictor of survival in breast cancer," *New England Journal of Medicine*, vol. 347, no. 25, pp. 1999–2009, 2002.

[68] H. Ishwaran and M. Lu, *Random Survival Forests*, Wiley StatsRef: Statistics Reference Online, 2014.

[69] M. H. Hsieh, L.-M. Sun, C.-L. Lin, M.-J. Hsieh, C.-Y. Hsu, and C.-H. Kao, "Development of a prediction model for pancreatic cancer in patients with type 2 diabetes using logistic regression and artificial neural network models," *Cancer Management and Research*, vol. 10, pp. 6317–6324, 2018.

[70] A. Bashiri, M. Ghazisaeedi, R. Safdari, L. Shahmoradi, and H. Ehtesham, "Improving the prediction of survival in cancer patients by using machine learning techniques: experience of gene expression data: a narrative review," *Iranian Journal of Public Health*, vol. 46, no. 2, pp. 165–172, 2017.

[71] X. Zhang, J. Xiao, and F. Gu, "Applying support vector machine to electronic health records for cancer classification," in *2019 Spring Simulation Conference (SpringSim)*, pp. 1–9, Tucson, AZ, USA, 2019.

[72] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing*, vol. 88, article 105946, 2020.

[73] H. Chen, S. Li, A. Asghar Heidari et al., "Efficient multi-population outpost fruit fly-driven optimizers: framework and advances in support vector machines," *Expert Systems with Applications*, vol. 142, article 112999, 2020.

[74] H. Pei, B. Yang, J. Liu, and K. Chang, "Active surveillance via group sparse Bayesian learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2020.

[75] A. Vial, D. Stirling, M. Field et al., "A comparative study of machine learning techniques for the improved prediction of nsclc survival analysis," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, pp. 1-2, Sydney, NSW, Australia, 2018.

[76] C. M. Lynch, B. Abdollahi, J. D. Fuqua et al., "Prediction of lung cancer patient survival via supervised machine learning classification techniques," *International Journal of Medical Informatics*, vol. 108, pp. 1–8, 2017.

[77] C.-M. Chen, C.-Y. Hsu, H.-W. Chiu, and H.-H. Rau, "Prediction of survival in patients with liver cancer using artificial neural networks and classification and regression trees," in *2011 Seventh international conference on natural computation*, vol. 2, pp. 811–815, Shanghai, China, 2011.

[78] S. Kim, M. Kon, and H. Lee, "Some comparisons of gene expression classifiers," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1753–1756, Shenzhen, China, 2016.

[79] R. Al-Hadidi, A. Alarabeyyat, and M. Alhanahnah, "Breast cancer detection using k-nearest neighbor machine learning algorithm," in *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 35–39, Liverpool, UK, 2016.

[80] T. Decaens, C. Barone, E. Assenat et al., "Phase II efficacy and safety data for the MET inhibitor tepotinib in patients (pts) with sorafenib-treated advanced hepatocellular carcinoma (HCC)," *Annals of Oncology*, vol. 29, article viii235, 2018.

[81] L. Heutte, "Keynote 3: random forests for biomedical data classification," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, p. ix, Kuching, Malaysia, 2017.

[82] C. Haarburger, P. Weitz, O. Rippel, and D. Merhof, "Image-based survival prediction for lung cancer patients using CNNS," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1197–1201, Venice, Italy, 2019.

[83] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, "Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1777–1787, 2019.

[84] H. Li, P. Boimel, J. Janopaul-Naylor et al., "Deep convolutional neural networks for imaging data based survival analysis of rectal cancer," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 846–849, Venice, Italy, 2019.

[85] C. Li, X.-Z. Lin, R. Wang, C. Hui, K.-M. Lam, and S. Zhang, "Differentiating pancreatic mucinous cystic neoplasms form serous oligocystic adenomas in spectral ct images using machine learning algorithms: a preliminary study," in *2013 International conference on machine learning and cybernetics*, vol. 1, pp. 271–276, Tianjin, China, 2013.

[86] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Hybrid intelligent systems in survival prediction of breast cancer," in *2012 12th International Conference on Hybrid Intelligent Systems (HIS)*, pp. 555–559, Pune, India, 2012.

[87] R. Al-Bahrani, A. Agrawal, and A. Choudhary, "Colon cancer survival prediction using ensemble data mining on SEER data," in *2013 IEEE international conference on Big Data*, pp. 9–16, Silicon Valley, CA, USA, 2013.

[88] H. R. Hassanzadeh, J. H. Phan, and M. D. Wang, "A semi-supervised method for predicting cancer survival using incomplete clinical data," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 210–213, Milan, Italy, 2015.

[89] J. Hang, L. Wu, L. Zhu et al., "Prediction of overall survival for metastatic pancreatic cancer: development and validation of a prognostic nomogram with data from open clinical trial and real-world study," *Cancer Medicine*, vol. 7, no. 7, pp. 2974–2984, 2018.

[90] G. Chartrand, P. M. Cheng, E. Vorontsov et al., "Deep learning: a primer for radiologists," *Radiographics*, vol. 37, no. 7, pp. 2113–2131, 2017.

[91] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.

[92] G. Hinton, "Deep learning|a technology with the potential to transform health care," *JAMA*, vol. 320, no. 11, pp. 1101-1102, 2018.

[93] J. D. Kothari, "A case study of image classification based on deep learning using TensorFlow," *Jubin Dipakkumar Kothari*

(2018). A Case Study of Image Classification Based on Deep Learning Using Tensorflow. International Journal of Innovative Research in Computer and Communication Engineering, vol. 6, no. 7, pp. 3888–3892, 2018.

[94] A. Paszke, S. Gross, F. Massa et al., "Pytorch: an imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8026–8037, 2019.

[95] F. Lin, B. Li, W. Zhou, H. Li, and Y. Lu, "Single-stage instance segmentation," *Communications, and Applications (TOMM)*, vol. 16, no. 3, pp. 1–19, 2020.

[96] J. Yu, A. L. Blackford, M. Dal Molin, C. L. Wolfgang, and M. Goggins, "Time to progression of pancreatic ductal adenocarcinoma from low-to-high tumour stages," *Gut*, vol. 64, no. 11, pp. 1783–1789, 2015.

[97] S. Jang, P. M. Graffy, T. J. Ziemlewicz, S. J. Lee, R. M. Summers, and P. J. Pickhardt, "Opportunistic osteoporosis screening at routine abdominal and thoracic CT: normative l1 trabecular attenuation values in more than 20 000 adults," *Radiology*, vol. 291, no. 2, pp. 360–367, 2019.

[98] L. C. Chu, S. Park, S. Kawamoto et al., "Application of deep learning to pancreatic cancer detection: lessons learned from our initial experience," *Journal of the American College of Radiology*, vol. 16, no. 9, pp. 1338–1342, 2019.

[99] Z. Zhu, Y. Xia, L. Xie, E. K. Fishman, and A. L. Yuille, "Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pp. 3–12, Springer, 2019.

[100] K.-L. Liu, T. Wu, P.-T. Chen et al., "Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation," *The Lancet Digital Health*, vol. 2, no. 6, pp. e303–e313, 2020.

[101] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[102] M. Nishio, S. Noguchi, and K. Fujimoto, "Automatic pancreas segmentation using coarse-scaled 2d model of deep learning: usefulness of data augmentation and deep U-net," *Applied Sciences*, vol. 10, no. 10, article 3360, 2020.

[103] J. A. Bartholomai and H. B. Frieboes, "Lung cancer survival prediction via machine learning regression, classification, and statistical techniques," in *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 632–637, Louisville, KY, USA, 2018.

[104] R. Jayashanka, C. Wijesinghe, A. Weerasinghe, and D. Pieris, "Machine learning approach to predict the survival time of childhood acute lymphoblastic leukemia patients," in *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pp. 426–432, Colombo, Sri Lanka, 2018.

[105] S. Kim, T. Park, and M. Kon, "Cancer survival classification using integrated data sets and intermediate information," *Artificial Intelligence in Medicine*, vol. 62, no. 1, pp. 23–31, 2014.

3. 3 Conclusion

In conclusion, the comprehensive survey of the state-of-the-art in pancreatic cancer survival prediction underscores the critical advancements made in recent years. The amalgamation of advanced machine learning and deep learning techniques, innovative biomarker identification, and extensive data integration has brought about a paradigm shift in our ability to foresee and comprehend pancreatic cancer prognosis. While challenges such as dataset heterogeneity and limited longitudinal data persist, the remarkable progress showcased in this survey instils optimism for the future of pancreatic cancer management. As research endeavours push the boundaries of prediction accuracy and clinical applicability, the potential to enhance patient outcomes through early intervention and tailored treatment strategies becomes increasingly tangible. This survey not only highlights the strides made but also underscores the imperative of continued collaborative efforts between medical experts, data scientists, and technology developers to refine our understanding and prediction of pancreatic cancer survival.

Chapter 4

Segmentation of CT Pancreas Images

4. 1 Introduction

In medical imaging and image analysis - specifically, the segmentation of CT pancreas images. In the ever-evolving healthcare landscape, non-invasive imaging techniques like Computed Tomography (CT) have become indispensable tools for diagnosing and understanding complex anatomical structures. The pancreas, a vital organ with multifaceted functions, presents unique challenges in medical imaging due to its intricate contours and variable appearances across patients. Through advanced computational methods and cutting-edge technology, the segmentation of CT pancreas images aims to delineate this organ with remarkable precision, fostering improved disease detection, treatment planning, and a deeper comprehension of its physiological dynamics. This study will delve into the methodologies, significance, and potential applications of CT pancreas image segmentation, unravelling its pivotal role in modern healthcare.

4. 2 Light Gradient-boosting Machine Edge Detection with Cropping Layer for Semantic Segmentation of Pancreas

Accurate and efficient organ segmentation is paramount for diagnosis and treatment planning in medical imaging. In particular, pancreas segmentation is crucial in assessing pancreatic diseases. To address this challenge, researchers have explored various deep-learning techniques, and one promising approach is the application of a Light Gradient-boosting Machine (LightGBM) with a cropping layer for semantic segmentation of the pancreas.

Light Gradient-boosting Machine Edge Detection with Cropping Layer is a novel method for semantic

segmentation of the pancreas in medical images. The paper first uses traditional methods to detect the edges and crop the image. The cropped image is then passed to LGBM for a semantic segmentation network to predict the pancreas segmentation.

LGBM has several advantages over other methods for semantic segmentation of the pancreas. First, LGBM is very fast, making it suitable for real-time applications. Second, LGBM is accurate, achieving state-of-the-art results on several benchmark datasets. Third, LGBM is very robust to noise and variations in image quality. LGBM is a powerful gradient-boosting framework with exceptional performance in various machine-learning tasks. By leveraging its ability to handle large-scale datasets efficiently, LGBM can effectively capture complex patterns and relationships in medical images.

Light Gradient-boosting Machine Edge Detection with Cropping Layer for Semantic Segmentation of Pancreas

Wilson Bakasa and Serestina Viriri
School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal, Durban, South Africa

July 20, 2023

Abstract

Anatomical variations in shape and volume metrics make pancreas medical image processing one of the most difficult subjects. Image processing in pancreas Computed Tomography (CT) is required to detect the class of pancreas anomaly accurately, reducing the possibility of a fatal outcome. It is estimated that a misunderstanding of radiology images accounts for 40% to 54% of all negligence claims. Machine learning advancements have created opportunities to train algorithms to learn patterns in pancreas images and use that knowledge to predict the pancreatic ductal adenocarcinoma (PDAC) overall survival rate. This paper performs feature extraction on CT pancreas images using edge detection techniques. Non-Local Means (NLM) are applied to the images to remove noise and smoothen them to highlight the defect. The experiments achieved an Intersection over Union (AOU) of 0.79 and an accuracy of 0.96. The images are then passed through the cropping layer to remove unwanted sections. The new train set is obtained from aggregating features from the edge detection feature extraction techniques and was then pruned to only eight features using the Principal Component Analysis technique. The train set was loaded into Light Gradient-boosting Machine (LGBM) to train the model to segment all the images.

Edge Detection; pancreatic ductal adenocarcinoma; Cropping; Semantic Pancreas Segmentation; LGBM

1 Introduction

Medical imaging is a critical component in many applications [1], [2] in treating PDAC, such as computer-aided diagnosis (CAD) using CT images to detect PDAC. These applications are found throughout the clinical course of events, most notably in the planning, executing, and evaluating procedures before surgical operations and diagnostic settings. Image segmentation, in general, is splitting an image into individual regions to reduce the complexity of analysis by assigning labels to pixels used to isolate homogeneous pixels of pancreas images.

LGBM [3], [4] is a gradient-boosting framework that generates high-quality prediction models using tree-based learning algorithms. Gradient Boosting Decision Tree (GBDT) [5], [6] is a decision tree-based gradient-boosting implementation. Each GBDT iteration adds a weak learner, such as a narrow decision tree, to the decision function model to approximate the objective's current negative gradient. These weak learners are constructed from arbitrary subsamples of the training

set, considering both examples and features. This subsampling reduces overfitting because learners only see a portion of the available information. Gradient boosting [7] is an ensemble technique for producing a single prediction model by combining weak learners stage by stage. Because of its generalisation ability, efficiency and precision are widely used in many machine learning tasks. [8].

Unlike most decision tree learning algorithms, LGBM grows the tree leaf-by-leaf, selecting the leaf with the highest delta loss to grow. Because the leaf-wise algorithm reduces loss by growing deeper trees, this procedure produces better results with fewer trees. To reduce over-fitting, the tree depth must be limited. Furthermore, other approaches that use [9] for decision tree learning, LGBM uses histogram-based algorithms [10]. Histograms divide continuous feature values into discrete bins, which saves memory and speeds up training, allowing LGBM to be used with large datasets.

The main contribution of this work is as follows:

1. A systematic technique for segmenting pancreas PDAC cancer CT images was investigated using LGBM and traditional edge detection feature extractors.

2. The dataset's item segmentation challenges, such as different image resolutions and class imbalance issues, were addressed.

3. A transfer learning architecture was proposed, automatically determining the num leaves parameters (which control the number of decision leaves in a single tree), resulting in improved accuracy.

4. A feature fusion architecture was proposed to improve segmentation accuracy by combining image features extracted by different traditional edge detection feature extractors.

This paper is structured as follows: Section 2 discusses related papers, and Section 3 describes the methodology, dataset, and various architectures investigated in this study. Section 4 summarises the experimental results and analysis, and Section 5 concludes and discusses future work.

2 Related Work

Extracting features from images is necessary because they can contain useful information. The detection of these features is required for image segmentation for diagnostic purposes. Edge detectors work well on clean images but are extremely sensitive to noise. Because of their excellent soft tissue contrast and high spatial resolution, CT image systems have recently become a massively valuable technique [11], [12]. On the other hand, image noise is the primary cause of degeneration, making it more difficult to distinguish fine detail in images during diagnosis examination. Then, because edge detectors are susceptible to noise, extracting features from Magnetic resonance imaging (MRI) images is difficult. As a result of this problem, researchers devised several methods for extracting features from CT images. [13], [14].

Roy et al. [15] proposed using brain MRI to investigate automated brain tumour identification and classification. Brain tumour segmentation was an important technique for extracting information from complex MRI brain images. In [4], Khuntia et al. developed and tested a method for segmenting brain tumours from two-dimensional MRI data. Tumours found in three dimensions are also shown. High-pass filtering, histogram equalisation, thresholding, morphological methods, and linked component labelling were used to detect cancer. The tumour volume was calculated after the 2D images were reconstructed into 3D volumetric data.

Havaei et al. [16] developed a semi-automatic method based on the KNN classifier. The researchers used the well-known BRATS 2013 dataset for both whole and core tests, achieving Dice similarities of 0.85 for the entire tumour region and 0.75 for the core tumour area. Mohsen et al. proposed classifying a dataset containing three distinct brain tumours using a deep learning-based classifier with discrete wavelet transform (DWT) and PCA. Four other deep learning-related studies with similar goals use the same dataset as we did in this study, which is critical for comparing and evaluating the performance outcomes of the proposed model.

3 Materials and Methods

This section discusses the implementation of edge detection techniques as feature extractors and the segmentation of pancreas CT images using LGBM to analyse how the model can augment today's medical imaging segmentation. First, to reduce noise, the NLM technique is used. Feature extraction is done using edge detection techniques, followed by the aggregation of features extracted and then feature pruning is done before the segmentation process.

3.1 Non-Local Means

Non-local means (NLM) is a denoising algorithm implemented in image processing that provides post-filtering clarity and preserves image details compared to local mean algorithms. The NLM [17], [18] fine-tuned individual pixel values with a weighted average of all pixels in the image. In this paper, we performed denoising using NLM to each image changed from DCM to tiff files to work easily with them. The tiff has an alpha channel that reserves discrete pixel transparency and colour information. While preserving the quality of files, the method provides a fast and simple decompression and compression of images. A tiff file is a great choice when high quality is the goal [19], [20]. The technique is based on changing the pixel colour to the average of the colours of nearby pixels [21], [22].

$$NLu(p) = \frac{1}{X(p)} \int f(d(Y(p), Y(q)))u(q)dq \tag{1}$$

where image patches Euclidean distance is given by $d(Y(p), Y(q))$, centred respectively at p and q, $C(p)$ is the normalising factor, and f is a decreasing function.

3.2 Cropping Layer

The definition of the term crop is "to trim" or "to cut back". To remove the outer margins of a digital image during pre-processing. Cropping can be used to adjust the aspect ratio (ratio of the image's length to width) and/or reduce the size of an image (in pixels). Cropping an image typically involves removing an unwanted segment or unimportant detail, changing the image's aspect ratio, or enhancing the composition [23]. A cropping operation is required to centre the pancreas CT images and remove extraneous portions of the image. Different locations within the image may also be used for separate pancreas imaging. By cropping the image and adding pads, we can ensure that nearly all images are in the same area of the overall image.

3.3 Feature Extraction based on Edge Detection

The comprehensive feature extraction process is important in predicting PDAC patient survival days using machine learning algorithms, according to [24]. The model's accuracy improves by removing redundant data from the input data and computing its features. This phase generates a new set of features by combining and transforming the original feature set. We extracted features using traditional edge detection methods such as Sobel, Prewitt, Gaussian, Median, Scharr, Farid, and Canny.

Edge detection techniques [25]–[27] are generally used to find and identify sharp discontinuities in images. For medical image recognition of the pancreatic, edge detection is crucial.

4 Feature Function

The feature functions [28] is written as $Vin(v0, ..., vn)$. Table: 1 shows some examples of feature functions. A new feature function could be created by combining several feature functions. The extracted feature's depth is determined by three factors, which are as follows:

1. The initial raw features extracted from images by feature extractors;
2. The feature function;
3. The number of times feature function operators were used.

We will use the sum function in this formulation to aggregate all features extracted from the edge detection feature extractors. Compositions of different feature functions are an easy way to express more complex feature functions. Feature functions are important not only for transferring but also for interpreting features [29].

Table 1: Feature Functions [29], [28]

Functions	Mathematical Definitions		Representation
Sum	$V\langle G, k\rangle = \sum_{g_j \in G} k_j$	(2)	Sum
Weight.Lp	$V\langle G, k\rangle = \sum_{g_j \in G} \lvert k_i - k_j\rvert^p$	(3)	Diff
Hadamard	$V\langle G, k\rangle = \prod_{g_j \in G} k_j$	(4)	Mul
Mean	$V\langle G, k\rangle = \frac{1}{\lvert G\rvert} \sum_{g_j \in G} k_j$	(5)	Mean

5 Feature Pruning

The edge detection feature extractors' output features are then evaluated. The feature evaluation routine selects the important features from the feature functions. The feature evaluation function is used to select the subset of important features. We chose the following characteristics: We began by determining the evaluation score for each feature. The feature score is calculated by comparing it to the threshold value and the basic feature score. If the feature score exceeds the sum of the two, the basic feature is discarded [30].

5.1 Principal Component Analysis

The computation of features is an important step in the segmentation process. However, not all extracted features contribute equally to the prediction task. As a result, feature reduction is required to select the most weighted features, thereby improving the machine learning algorithm's accuracy. Using the PCA method, traditional edge detection models generate many features that can be reduced to a lower dimension.

PCA [31]–[33] is a widely used dimensionality reduction method for reducing the dimension of large datasets while retaining the majority of the information. PCA is used to eliminate those irrelevant features, resulting in lower computational overhead for the proposed method while retaining the same accuracy.

5.2 Feature Map

A feature map is a function that maps a data vector to a feature space. The main logic in machine learning is to provide data that can be segmented to the learning algorithm. It refers to mapping features from input space to kernel Hilbert space, which is common in very large dimensions [34], [35].

A filter uses traditional methods to measure how closely an input region resembles a feature, which can be achieved after the pre-processing stage. Any pronounced aspect, such as a vertical edge, a horizontal edge, an arch, or a diagonal, can be considered a feature. The function of a filter is to serve as a single template or pattern that, when convolved across the input, identifies correspondences between the stored template and various locations in the input image [20], [36].

Some of the feature maps obtained from our images are shown in Figure: 1.

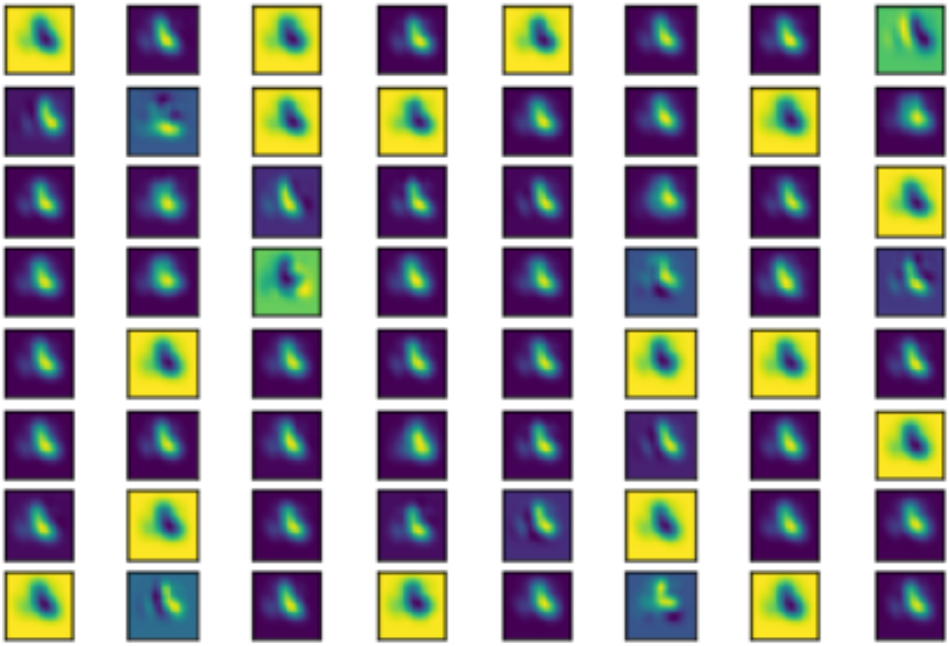


Figure 1: Some selected Feature Maps for the images

6 Implementing LGBM for Segmentation

Compared to deep learning algorithms, traditional approaches are frequently easier to interpret. Certain applications, such as the segmentation of medical images, may require this. When compared to deep learning methods, conventional methods frequently require less processing power. Applications that demand real-time performance may find this to be crucial.

In some circumstances, deep learning approaches or conventional methods alone may not be superior to ensemble methods. For instance, a study [37] demonstrated that, for the segmentation of skin lesions, a combination of deep learning and conventional techniques produced superior results to either methodology used alone. Standard and deep learning methods will be used depending on the particular application. To improve segmentation outcomes, ensemble approaches, however, might be a potent tool.

LGBM [3], [38] is a high-performance gradient-boosting algorithm based on decision trees, similar to Random Forest. It splits the tree using the best-fit algorithm, unlike other boosting algorithms. The algorithm is widely used due to its fast and efficient training, requiring less memory than other algorithms.

First, load the image, convert it to grayscale, and preprocess it to form a blob. Then apply, traditional edge detection methods Sobel, Prewitt, Gaussian, Median, Scharr, Farid, and Canny as feature extractors method to the blob image. The edge features will then be extracted from the image, and a training dataset will be created with the edge features and the ground truth labels.

The LGBM Classifier model will then be trained on the training dataset. Finally, the model will be evaluated on a test dataset.

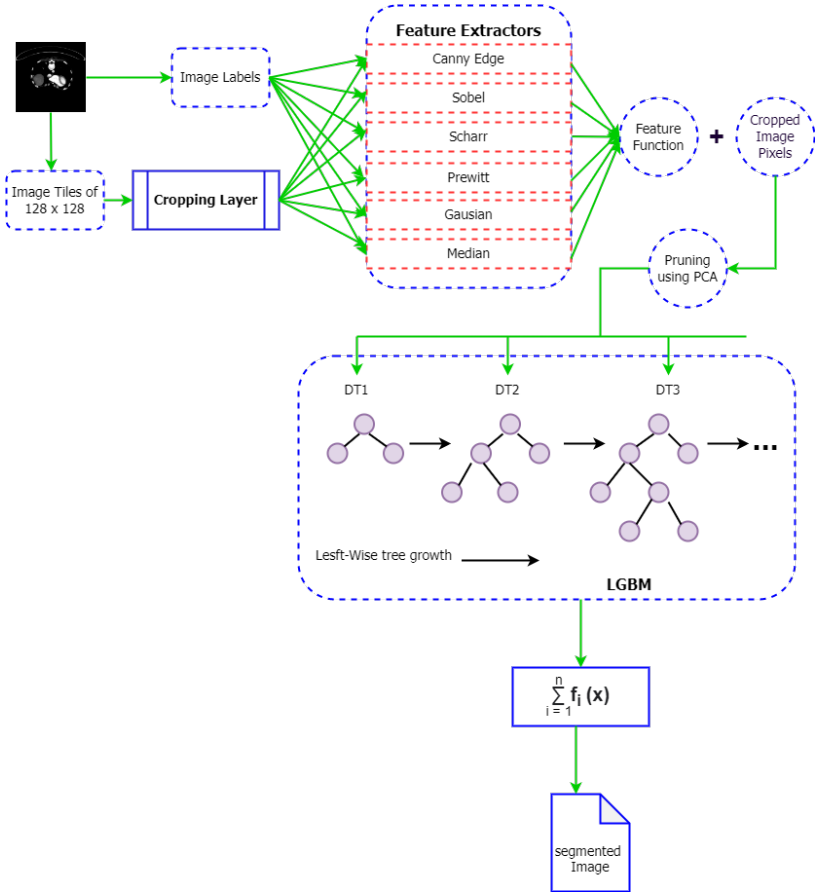


Figure 2: Images pass through the cropping layer for pre-processing, followed by Transfer Learning of features from traditional edge detection methods to the LGBM for segmenting the CT pancreas images. Features from the feature extractors are fused and added to the cropped image pixels. Pruning is then done to reduce the number of features before they are fed into LGBM as a new training set for image segmentation.

Figure: 2 shows the images are pre-processed using the cropping technique in the cropping layer to remove unwanted pixels that might be considered during feature extraction. The cropped images are then passed as input to feature extractors to extract various features combined with the feature function. Results from the feature function and the features from the cropped image are then summed and passed through a pruning stage using the PCA, and the output is sent as input for segmentation using LGBM. Algorithm: 1 shows the processing done by the model.

LGBM can handle massive amounts of data while consuming very little memory. It emphasises outcome accuracy. Scientists use LGBM [39] to create research applications because it supports

Algorithm 1: Model algorithm: The model performs semantic segmentation using the feature extracted using edge detection techniques. Cropping is done first in the cropping layer, with $R^j CropOutValue$ denoting the crop-out value of region R^j, and C signifying the crop. The cropped image is sent to feature extractors as input. The features are fused and then pruned to remain with only eight features which are considered the best for segmenting the PDAC ct images using LGBM

Input: Data: Labeled Images.
Input: P(i): Pixel Values.
/* Cropping Layer. */

$$E_{CropOut} = \sum_{R^j \epsilon C} R^j_{CropOutValue}$$

Input: Data: Cropped Images.
/* All features generated must match how features are generated for training. */
for $P(i) = 0$ **do**
    // Ground truth.
    $Feature1 \leftarrow original\_image\_pixels$
    for $P(i) = 1$ **do**
        // Other pixels not ground truth.
        /* Generate FEATURES and add them to the data frame. */
        $Feature2 \leftarrow canny\_edge$
        $Feature3 \leftarrow Scharr$
        $Feature4 \leftarrow Sobel$
        $Feature5 \leftarrow Prewitt$
        $Feature6 \leftarrow Gaussian(sigma = 3)$
        $Feature7 \leftarrow Gaussian(sigma = 7)$
        $Feature8 \leftarrow Median(sigma = 3)$
        if $A \perp\!\!\!\perp B$ **then**
            $A \equiv independent\_variables.$
        else
            $A \not\equiv independent\_variable$
        end if
        —
    end for
    /* New Train Set from the extracted feature now to be loaded to LGBM */
    Input: Data: New Train Set
    $Data = train + test$ // Test data for accuracy testing
    /* Implement LGBM model: */
    $model = LGBM.train(LGBM\_params, d\_train, 100)$
    $num\_leaves = 90$
    $max\_depth = 9$
    $Boosting\_Type \leftarrow dart$
    $loaded\_model = pickle.load(open(filename,'rb'))$ // Applying trained model to segment multiple files.
end for
return $Segmented\_Images$

GPU learning. LGBM is an ensemble learning algorithm that performs segmentation by combining the outputs of multiple base learner models. The Light GBM, a gradient-boosting framework, employs tree-based learning methods. It is intended to be widely disseminated and efficient, with the following benefits:

- Fast when training and improved accuracy.
- Less memory requirement.
- Very efficient.
- Provide GPU learning Support.
- Capable of handling Large-scale data.

According to [40], LGBM grows trees vertically, whereas other tree-based learning algorithms grow trees horizontally. LGBM grows in a tree, whereas other algorithms grow in the form of levels. It will grow the leaf with the greatest delta loss. When expanding the same leaf, a leaf-wise algorithm can reduce loss more than a level-wise method.

7 Results

Ten decom files were randomly selected from the Cancer Imaging Archive (TCIA) Public Access dataset, consisting of 3D CT scans of 512 $\times$ 512-pixel resolution from 53 males and 27 female subjects 18-76 age group [41] 370 patients with PDAC and 320 controls' contrast-enhanced CT images from the Taiwanese Centre dataset were used to train CNN. Using these datasets, researchers can examine how to use CT imaging to predict PDAC overall survival [42].

7.0.1 Cropped Images

First, the images are passed through the cropping layer to remove unwanted sections. Figure: 3 shows some results for the non-cropped images (A1 - A4) and cropped images (B1 - B4). The first column shows the original image; column two shows images after setting the blob as the input to the network and performing a forward pass to compute the edges; column three shows connected component-based labelled images; and column four shows created a false colour image with the black background and coloured objects.

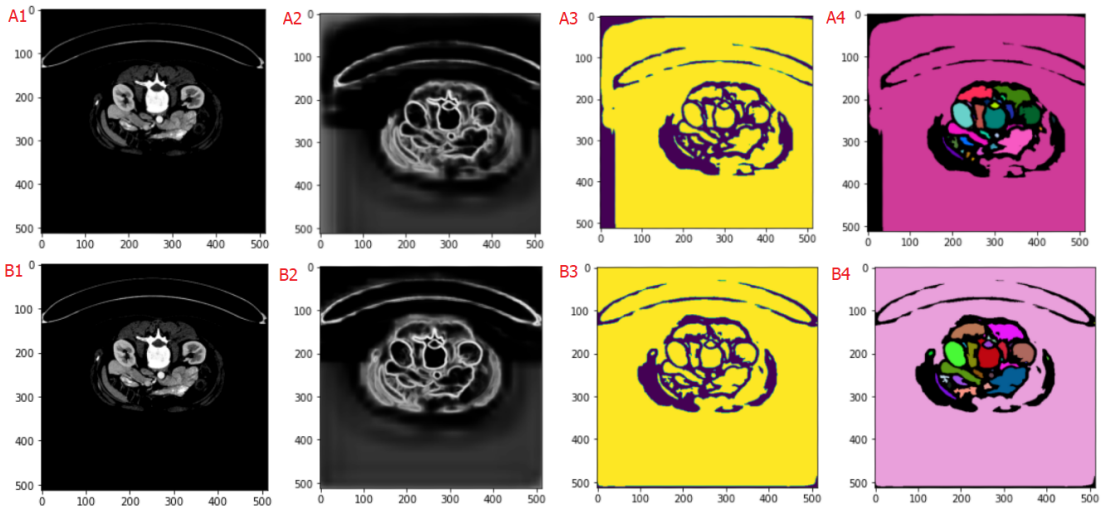


Figure 3: Cropping Layer implementation to the pancreas CT images. A1 (Uncropped) and B1 (Cropped) - Original Image, A2 (Uncropped) and B2 (Cropped) - blob (preprocessed image), A3 (Uncropped) and B3 (Cropped) - Segmented binary image, and A4 (Uncropped) and B4 (Cropped) - False colour image with a black background and coloured objects.

7.0.2 Data Frame

A Data Frame is a data structure that organises information into a two-dimensional table of rows and columns. The two tables, Table: 2 and Table: 2, shows data frame prints before and after implementing cropping. Pandas, the popular Python data analysis library, uses Data Frames as its primary data type [43].

Tables: 2 and 3 show parameters for the data frame before and after implementing cropping to the images. Solidity is calculated as the difference between the number of pixels in the object and the convex hull image. This can measure the object's compactness. The image passed through the cropping layer has a high solidity of 0.97, which is better than that of an image not cropped, 0.94. The area for up to ten (without cropping) and thirteen (with cropping) different labels is calculated, which shows the cropped images had a high area of 2017 at label 5 and non-cropped images had the least area of 65 at label 4. The equivalent diameter is calculated to give the circle's diameter, whose area is the same as the contour area. This shows the highest equivalent diameter for the cropped image of 51.35 at label 5, which has the highest area, and the least equivalent diameter of 9.1 at label 4, which has the least area. The mean intensity at any given pair of coordinates (x,y) is a two-dimensional function f(x,y), where x and y are spatial coordinates. Solidity is also calculated for various labels. The mean intensity is highest at label 9 for the cropped image, which is 249.2, but that will record a low solidity for the same image.

7.0.3 Feature Importance

The node impurity in Scikit-learn is calculated using Gini importance [44], and feature importance is essentially a reduction in a node's impurity weighted by the number of samples reaching

Table 2: Dataframe without cropping

Label	**Area**	**Equivalent diameter**	**Mean intensity-0**	**Mean intensity-1**	**Mean intensity-2**	**Solidity**
2	1702	46.551624	117.602233	117.602233	117.602233	0.792734
3	1849	48.520304	86.833423	86.833423	86.833423	0.724246
4	65	9.097284	101.984615	101.984615	101.984615	0.802469
6	103	11.451798	247.922330	247.922330	247.922330	0.811024
10	175	14.927053	30.697143	30.697143	30.697143	0.935829

Table 3: Dataframe after cropping.

Label	**Area**	**Equivalent diameter**	**Mean intensity-0**	**Mean intensity-1**	**Mean intensity-2**	**Solidity**
4	1718	46.769921	48.939464	48.939464	48.939464	0.852605
5	2071	51.350551	43.259778	43.259778	43.259778	0.828731
9	2012	50.613812	249.214712	249.214712	249.214712	0.784405
12	196	15.797308	41.306122	41.306122	41.306122	0.915888
13	1933	49.610201	138.455251	138.455251	138.455251	0.966983

that node relative to the overall number of samples, which is node probability.

$$nj_i = s_j C_i - s_{left(i)} I_{left(i)} - s_{right(i)} I_{right(i)} \tag{6}$$

Where: nj_i = node i importance, s_i = number of samples weighted, reaching node i, I_i = the impurity value of node i, $left(i)$ = child node on left of node i and $right(i)$ = child node on right of node i

To determine the feature importance for each decision tree, this equation provides us with the importance of a node j. The tree's different branches can each use a particular feature.

$$fj_j = \frac{\sum_{i:node\ i\ splits\ on\ feature\ j} nj_i}{\sum_{i\epsilon\ all\ node} nj_i} \tag{7}$$

The features are divided by the total number of trees in our random forest, and the result is the overall feature relevance. The features are normalised against the sum of all feature values in the tree. With this in mind, you can understand the significance of features in random forests.

Depending on how well a feature may improve the purity of the leaves, each tree in a random forest, LGBM, or XGBoost can determine how important that feature is. The significance of the characteristic increases as the increase in leaf purity increases. This is carried out individually for each tree, averaged across all trees, and then set to one. Using feature selection can aid in better comprehending the problem that has been handled and occasionally result in model enhancements. A higher score indicates that the particular feature will have a greater impact on the model being used to forecast the particular variable.

The feature importance graph is plotted to see which feature extractors did the best to extract features. To avoid ignoring important features, we had to set a threshold which was the original image. Gaussian was the best, and different feature ranges are shown in Figure: 4, 5 and 6. The

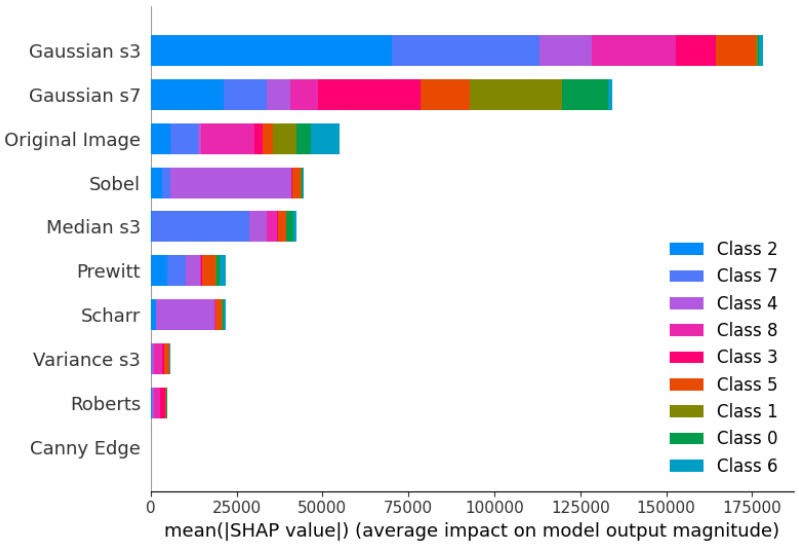


Figure 4: Implementing LGBM for segmentation with different feature extractors: The feature importance graph demonstrates which feature extractor extracted the most important features in predicting the target variable. Gaussian at Sigma 3 was the best, and Canny Edge was the least. The original image is the cropped image.

original image (cropped image) pixel values have also indicated that they are important for feature extraction [45].

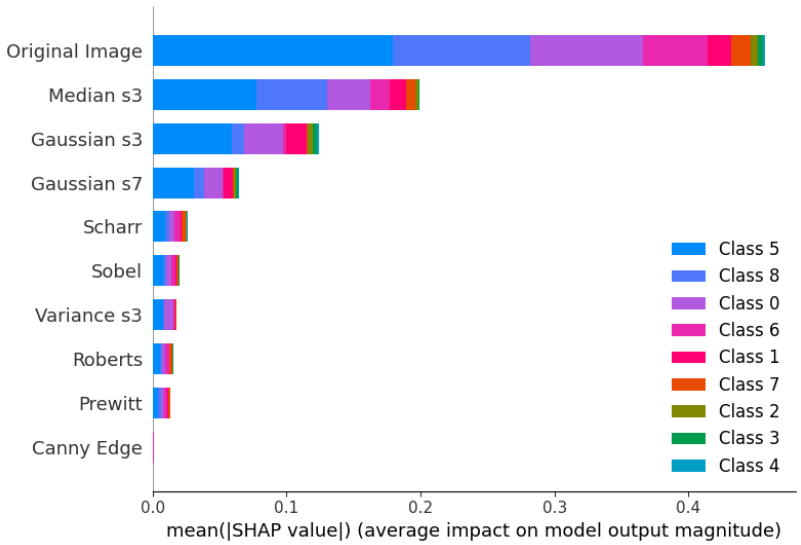


Figure 5: Implementing XGBoost for segmentation with different feature extractors: The feature importance graph demonstrates which feature extractor extracts the most important features in predicting the target variable. The original image was the best, and Canny Edge was the least. The original image is the cropped image. Accuracy can be improved by adding more feature extractors.

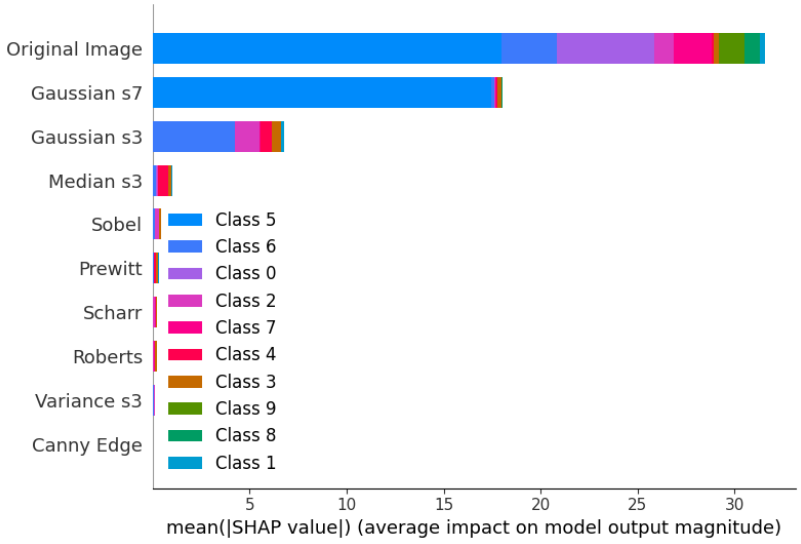


Figure 6: Implementing RandomForest for segmentation with different feature extractors: The feature importance graph demonstrates which feature extractor extracts the most important features in predicting the target variable. The original image was the best, and Canny Edge was the least. The original image is the cropped image. Accuracy can be improved by adding more feature extractors.

Feature importance refers to techniques that assign a score to input features based on their usefulness in predicting a target variable. Statistical correlation scores, coefficients calculated as part of linear models, decision trees, and permutation importance scores are all common examples of feature importance scores. Feature importance scores are important in a predictive modelling project because they provide insight into the data and the model and the foundation for dimensional reduction and feature selection, which can improve the efficiency and effectiveness of a predictive model on the problem [46].

Any features extracted by all those methods below the original image will be discarded. LGBM produced better results by adding important features from Gaussian Sigma 3 and 7. For randomForest and XGBoost, all the methods failed to extract features above the threshold and all features were discarded.

7.1 Improving Segmentation Speed

The PCA was implemented to improve the segmentation algorithm as it would calculate the maximum number of features that can be used. This is to avoid inputting many features, thereby increasing the load on the model without improving accuracy. As shown in Figure: 7, when using thirty features, the cumulative variance will be below 0.9, which is low. It will be slightly above 0.9 when using one hundred features but shows signs of increasing. When using eight hundred features, the graph line started to be constant when the number of components was above two hundred, so two hundred components were used as the graph shows cumulative variance above 0.95.

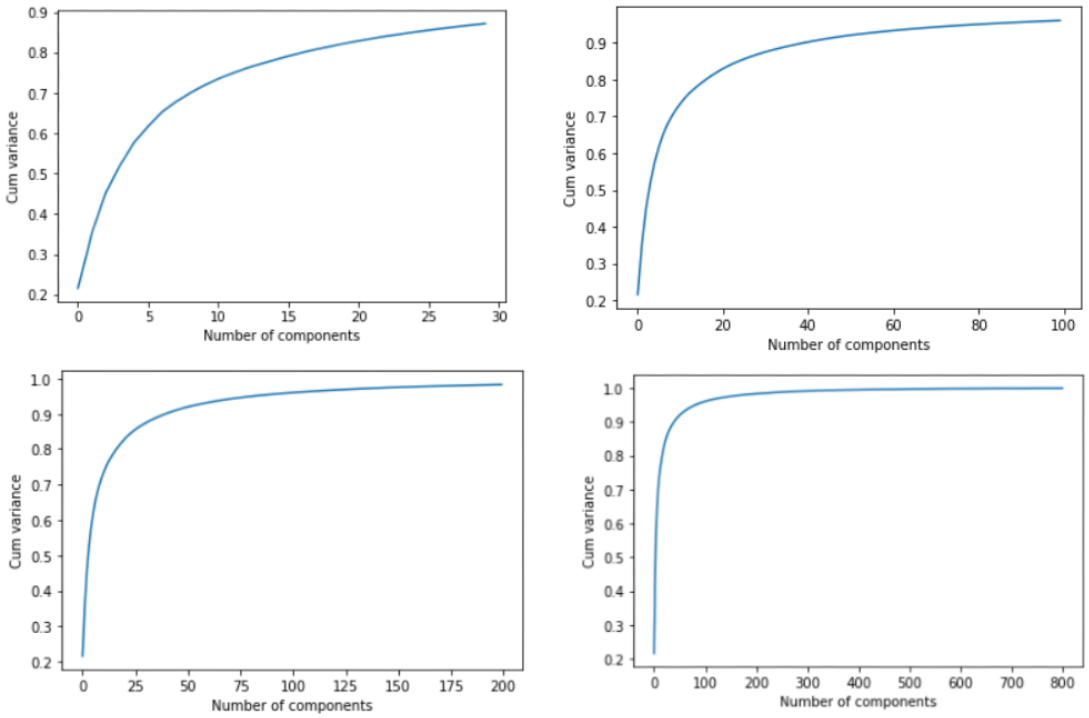


Figure 7: Graph to analyse the best cumulative variance against the number of components using PCA.

7.2 Segmented Images

Figure: 8 depicts a sample of original images segmented using the LGBM, XGBoost and Random Forest, and features extracted using traditional methods from the TCIA dataset. The algorithm first identifies regions in the image most likely to contain the desired objects and then examines these regions at increasing resolutions. The ground truth is the true and accurate segmentation to compare with our segmented images to test the accuracy of automated image analysis processes.

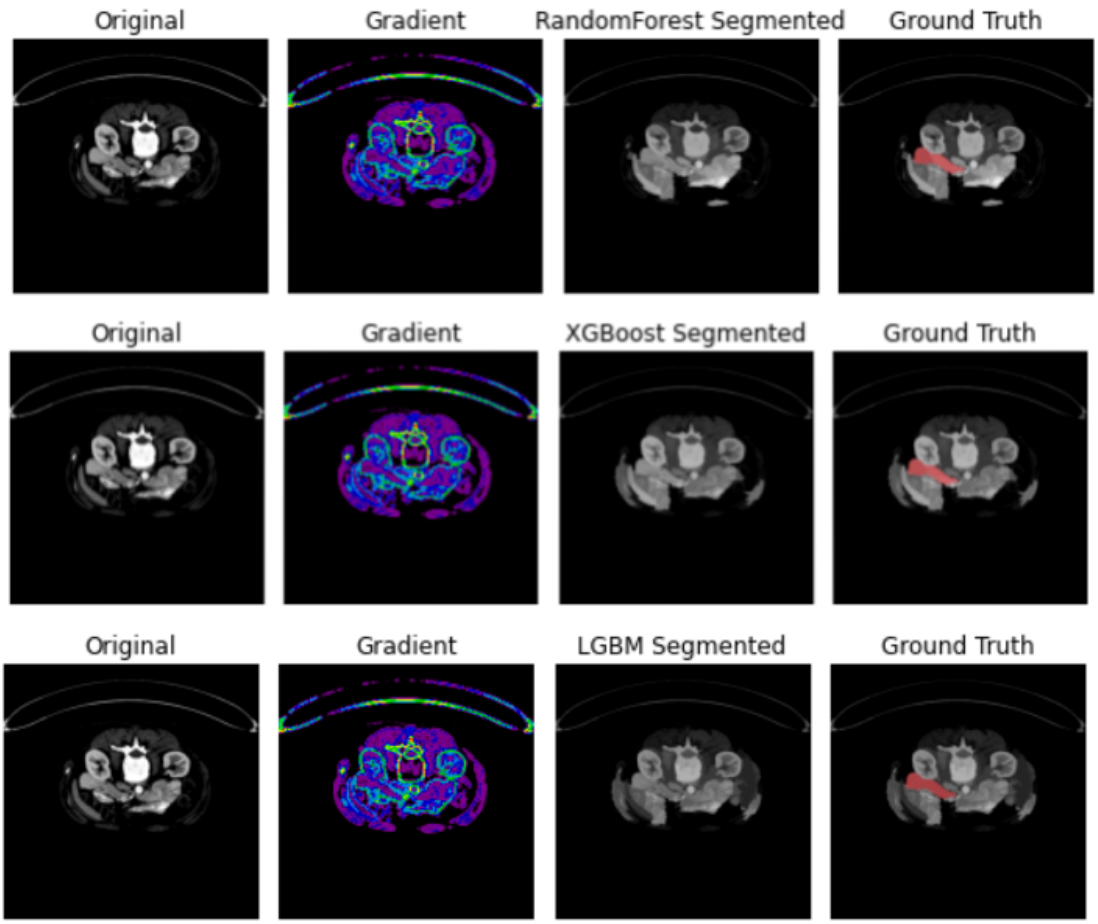


Figure 8: Original images from the dataset, with each row showing the local gradient, segmented and ground truth images for each original image. The pancreas is shown in red for the ground truth, and images are segmented using LGBM, XGBoost and Random Forest.

8 Discussion

Segmentation's primary objective is to assist in identifying and characterising the pancreas and its pathogenic characteristics. The comprehension of the main pancreatic diseases, like the size cyst, may be enhanced by quantitative data obtained utilising segmentation techniques [47]. To decide on surgery and treatment, it is crucial to use segmentation algorithms to obtain the quantitative volumetric measures of the pancreatic lesion.

Publicly accessible datasets are crucial for the research community and evaluating the benefits and drawbacks of segmentation techniques since assessing the precision of various segmentation techniques using the same dataset is more useful. It is necessary to use the same standardised

dataset and to develop methods for reducing the large degree of variation in image parameters between imaging facilities [48]. Additionally, most of the data only includes CT images; MRI images are few.

The absence of publicly accessible datasets containing matching labelled images has hampered the development of pancreatic segmentation algorithms [49]. The NIH datasets, the first to be utilised publically, contain the most thorough hand labelled/annotated atlases; however, they are deficient in basic patient data necessary for preprocessing, such as atlas selection. TCIA is designed to foster increased public availability of high-quality cancer imaging data sets for research. F.A.I.R. (Findable, Accessible, Interoperable, and Reproducible) criteria for data release are strictly followed, making data accessible.

The largest research material studies pancreatic segmentation based on atlas and neural network segmentation methods. The equipment operating speed has increased, and the segmentation accuracy is primarily considered reasonably high. These two approaches are crucial directions for future growth. Future research will also include a significant amount of work on novel methodologies and hybrid technologies [50]. Although the pancreas only takes up a very small portion (e.g., 0.5% of the total CT volume), it is well known that it frequently exhibits substantial variability in form, size, and position. Therefore, segmenting the pancreas is still a complex and challenging task.

The accuracy assessment has generally improved yearly; however, the evaluated papers range greatly from one another, making it difficult to evaluate accuracy using standardised metrics and datasets. The datasets used in the publications are, in fact, highly varied from one another, containing various numbers and varieties of patients and images obtained using various acquisition parameters.

According to [51], statistical shape models effectively use complicated form variability, whereas deep neural networks have seen substantial success in segmenting medical images. a Bayesian model that incorporates the segmentation findings from both deep neural networks and statistical shape models was proposed, and it reported a decent result of 85.32 ± 4.19 of DSC. This model used the rich shape priors gained from statistical shape models to improve the segmentation of deep neural networks [52].

The best dice similarity coefficients for segmenting the pancreas and cysts were 85.1 and 86.7, respectively, and were achieved by [53] using a mix of the random walker and region-expanding approaches. [54] used region growth to mark the pancreatic region after applying histogram equalisation to improve the contrast of computed tomography images in low-contrast conditions. Although the best Jaccard index of 86.97 was used to segment data, the used datasets are not mentioned specifically.

Image segmentation uses a technique called "graph cut," [55] one of those for energy reduction. The pancreas was segmented from CT images 2016 using a graph cut approach and probabilistic atlas. Their data found that the pancreatic segmentation Jaccard index was 0.705. In natural image processing, region growth is a crucial and well-established segmentation approach applied to segment abdominal organs using different methods.

A bottom-up pancreas segmentation representation has been developed using aggregated confidence maps to categorise image areas or super-pixels to the pancreas and non-pancreas label assignments. This representation uses dense binary image patch labelling to label the confidence maps [52]. This technique uses mid-level visual representations of image segments to increase the segmentation accuracy of highly deformable organs, such as the pancreas. A set of multi-scale and multi-level deep convolutional neural networks applied in a sliding window fashion on local image patches to capture the complexity of the pancreas' appearance in CT images was proposed

in previous work, which further improved segmentation performance.

Our proposed method has the advantage of the ability to learn rich representations, be versatile, and increase segmentation accuracy (0.96) by reducing manual intervention. The cropping layer plus a reliable bounding box for more precise segmentation is created by first localising the pancreas within the entire 3D CT image. The next step is to segment the pancreatic tissue using a fully deep-learning approach based on the effective use of holistically-nested convolutional networks on the three orthogonal axial, sagittal, and coronal views. The holistically nested convolutional networks probability maps for per-pixel are fused via pooling to build a 3D bounding box of the pancreas that can maximise recall.

Soon, the primary goal will continue to increase the pancreas segmentation's accuracy and efficiency because the most recent accuracy is below that of other abdominal organs and well below what is needed to meet clinical needs. Additionally, due to the poor contrast in the boundary, the unpredictability in location, shape, and size, and the importance and value of the pancreas and its lesion region in clinical practice [56], automatic segmentation of these structures is vital and difficult.

Segmenting pancreas images simultaneously from various modalities, such as CT and MRI, is also important. Doctors typically analyse images from multiple imaging modalities to more accurately measure and evaluate the state of the abdominal organs.

9 Conclusions

Image segmentation has been widely used in pancreas imaging to separate homogeneous areas. Organ segmentation methods have been successfully implemented in large organs such as the lungs, heart, and kidneys in recent years, but little has been done to small organs like the pancreas. The accuracy of segmentation determines whether the automated analysis procedures succeed or fail. However, the complexity of the anatomical structure and the scarcity of labelled data for PDAC patients make it difficult to achieve an acceptable level of accuracy in pancreas segmentation. These and other factors make pancreas segmentation difficult, and there is a high risk of over or under-segmentation.

The research has proved to be efficient in using edge detection traditional methods to be feature extractors and the features to be aggregated and pruned to come up with only crucial features for segmentation. Pruning is important for reducing features and helps to increase accuracy as only specific features are used rather than a general description.

The fact that there aren't enough labelled imaging datasets readily available to the public is one of the remaining barriers to developing pancreatic segmentation algorithms. Future approaches must be resilient to the wide variations in pancreatic appearance, shape, and size to improve segmentation quality. In diseased circumstances, this will make it possible to capture the characteristics of the pancreas in great detail and may pave the way for automatic segmentation in clinical settings.

References

[1] M. P. Arakeri, S. K. Manvi, *et al.*, "Medical imaging and computer-aided diagnosis," in *Medical Imaging Methods*, CRC Press, 2021, pp. 45–56.

[2] J. Paul and T. Sivarani, "Computer aided diagnosis of brain tumor using novel classification techniques," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 7, pp. 7499–7509, 2021.

[3] R. He, X. Yang, T. Li, *et al.*, "A machine learning-based predictive model of epidermal growth factor mutations in lung adenocarcinomas," *Cancers*, vol. 14, no. 19, p. 4664, 2022.

[4] M. Khuntia, P. K. Sahu, and S. Devi, "Prediction of presence of brain tumor utilizing some state-of-the-art machine learning approaches," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022.

[5] H. Rao, X. Shi, A. K. Rodrigue, *et al.*, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Applied Soft Computing*, vol. 74, pp. 634–642, 2019.

[6] Q. Li, Z. Wen, and B. He, "Practical federated gradient boosting decision trees," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, 2020, pp. 4642–4649.

[7] Z. Zhang, Y. Zhao, A. Canes, D. Steinberg, O. Lyashevska, *et al.*, "Predictive analytics with gradient boosting in clinical medicine," *Annals of translational medicine*, vol. 7, no. 7, 2019.

[8] L. Liu, Y. Yu, Z. Fei, *et al.*, "An interpretable boosting model to predict side effects of analgesics for osteoarthritis," *BMC systems biology*, vol. 12, no. 6, pp. 29–38, 2018.

[9] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting," *Energy*, vol. 214, p. 118 874, 2021.

[10] Z. Pan, S. Fang, and H. Wang, "Lightgbm technique and differential evolution algorithm-based multi-objective optimization design of ds-apmm," *IEEE Transactions on Energy Conversion*, vol. 36, no. 1, pp. 441–455, 2020.

[11] P. Hunter, "The advent of ai and deep learning in diagnostics and imaging: Machine learning systems have potential to improve diagnostics in healthcare and imaging systems in research," *EMBO reports*, vol. 20, no. 7, e48559, 2019.

[12] E. Abadi, W. P. Segars, B. M. Tsui, *et al.*, "Virtual clinical trials in medical imaging: A review," *Journal of Medical Imaging*, vol. 7, no. 4, p. 042 805, 2020.

[13] S. Wang, B. Kang, J. Ma, *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *European radiology*, vol. 31, no. 8, pp. 6096–6104, 2021.

[14] M. Tougaccar, B. Ergen, and Z. Comert, "Detection of lung cancer on chest ct images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 23–39, 2020.

[15] K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with neuromorphic computing," *Nature*, vol. 575, no. 7784, pp. 607–617, 2019.

[16] K. Y. Lim and R. Mandava, "A multi-phase semi-automatic approach for multisequence brain tumor image segmentation," *Expert systems with applications*, vol. 112, pp. 288–300, 2018.

[17] S. Saladi and N. Amutha Prabha, "Analysis of denoising filters on mri brain images," *International Journal of Imaging Systems and Technology*, vol. 27, no. 3, pp. 201–208, 2017.

[18] J. Cui, K. Gong, N. Guo, *et al.*, "Pet image denoising using unsupervised deep learning," *European journal of nuclear medicine and molecular imaging*, vol. 46, no. 13, pp. 2780–2789, 2019.

[19] H. Amri, A. Khalfallah, M. Gargouri, N. Nebhani, J.-C. Lapayre, and M.-S. Bouhlel, "Medical image compression approach based on image resizing, digital watermarking and lossless compression," *Journal of Signal Processing Systems*, vol. 87, no. 2, pp. 203–214, 2017.

[20] M. J. Willemink, W. A. Koszek, C. Hardell, *et al.*, "Preparing medical imaging data for machine learning," *Radiology*, vol. 295, no. 1, p. 4, 2020.

[21] M. Diwakar, P. Kumar, and A. K. Singh, "Ct image denoising using nlm and its method noise thresholding," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 14 449–14 464, 2020.

[22] H. V. Bhujle and B. H. Vadavadagi, "Nlm based magnetic resonance image denoising–a review," *Biomedical Signal Processing and Control*, vol. 47, pp. 252–261, 2019.

[23] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.

[24] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.

[25] B. K. Shah, V. Kedia, R. Raut, S. Ansari, and A. Shroff, "Evaluation and comparative study of edge detection techniques," *IOSR Journal of Computer Engineering*, vol. 22, no. 5, pp. 6–15, 2020.

[26] X. Li, H. Jiao, and Y. Wang, "Edge detection algorithm of cancer image based on deep learning," *Bioengineered*, vol. 11, no. 1, pp. 693–707, 2020.

[27] M. Gholizadeh-Ansari, J. Alirezaie, and P. Babyn, "Deep learning for low-dose ct denoising using perceptual loss and edge detection layer," *Journal of digital imaging*, vol. 33, no. 2, pp. 504–515, 2020.

[28] Y. Yao, L. Su, and Z. Lu, "Deepgfl: Deep feature learning via graph for attack detection on flow-based network traffic," in *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, IEEE, 2018, pp. 579–584.

[29] R. A. Rossi, R. Zhou, and N. K. Ahmed, "Deep feature learning for graphs," *arXiv preprint arXiv:1704.08829*, 2017.

[30] B. O. Ayinde, T. Inanc, and J. M. Zurada, "Redundant feature pruning for accelerated inference in deep neural networks," *Neural Networks*, vol. 118, pp. 148–158, 2019.

[31] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *Journal of Signal and Information Processing*, vol. 4, 2020.

[32] F. Kherif and A. Latypova, "Principal component analysis," in *Machine Learning*, Elsevier, 2020, pp. 209–225.

[33] H.-P. Deutsch and M. W. Beinker, "Principal component analysis," in *Derivatives and internal models*, Springer, 2019, pp. 793–804.

[34] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia detection using cnn based feature extraction," in *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)*, IEEE, 2019, pp. 1–7.

[35] M. Zhang, W. Li, Q. Du, L. Gao, and B. Zhang, "Feature extraction for classification of hyperspectral and lidar data using patch-to-patch cnn," *IEEE transactions on cybernetics*, vol. 50, no. 1, pp. 100–111, 2018.

[36] A. K. Pradhan, S. Swain, J. K. Rout, and N. K. Ray, "Prediction of stroke disease using different types of gradient boosting classifiers," in *Advances in Data Computing, Communication and Security*, Springer, 2022, pp. 337–346.

[37] A. H. Khan, D. N. Awang Iskandar, J. F. Al-Asad, H. Mewada, and M. A. Sherazi, "Ensemble learning of deep learning and traditional machine learning approaches for skin lesion segmentation and classification," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 13, e6907, 2022.

[38] K. B. Abou Omar, "Xgboost and lgbm for porto seguro's kaggle challenge: A comparison," *Preprint Semester Project*, 2018.

[39] D. Segelke, L. Gehrke, and J. Wabbersen, "Imputation of genetic characteristics using deep learning methods," *Interbull Bulletin*, no. 55, pp. 103–106, 2019.

[40] B. S. Ahamed, S. Arya, *et al.*, "Lgbm classifier based technique for predicting type-2 diabetes," *European Journal of Molecular & Clinical Medicine*, vol. 8, no. 3, pp. 454–467, 2021.

[41] K.-L. Liu, T. Wu, P.-T. Chen, *et al.*, "Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: A retrospective study with cross-racial external validation," *The Lancet Digital Health*, vol. 2, no. 6, e303–e313, 2020.

[42] W. Bakasa and S. Viriri, "Pancreatic cancer survival prediction: A survey of the state-of-the-art," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.

[43] N. Bantilan, "Pandera: Statistical data validation of pandas dataframes," in *Proceedings of the Python in Science Conference (SciPy)*, 2020, pp. 116–124.

[44] B. H. Menze, B. M. Kelm, R. Masuch, *et al.*, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, pp. 1–16, 2009.

[45] A. I. Adler and A. Painsky, "Feature importance in gradient boosting trees with cross-validation feature selection," *Entropy*, vol. 24, no. 5, p. 687, 2022.

[46] Y. M. Banadaki, "Detecting malicious dns over https traffic in domain name system using machine learning classifiers," *Journal of Computer Sciences and Applications*, vol. 8, no. 2, pp. 46–55, 2020.

[47] K. Preuss, N. Thach, X. Liang, *et al.*, "Using quantitative imaging for personalized medicine in pancreatic cancer: A review of radiomics and deep learning applications," *Cancers*, vol. 14, no. 7, p. 1654, 2022.

[48] V. M. Campello, P. Gkontra, C. Izquierdo, *et al.*, "Multi-centre, multi-vendor and multi-disease cardiac segmentation: The m&ms challenge," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.

[49] N. F. Greenwald, G. Miller, E. Moen, *et al.*, "Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning," *Nature biotechnology*, vol. 40, no. 4, pp. 555–565, 2022.

[50] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE access*, vol. 7, pp. 81 542–81 554, 2019.

[51] M.-L. Huang and Y.-Z. Wu, "Semantic segmentation of pancreatic medical images by using convolutional neural network," *Biomedical Signal Processing and Control*, vol. 73, p. 103 458, 2022.

[52] P. M. Paithane and S. Kakarwal, "Automatic pancreas segmentation using a novel modified semantic deep learning bottom-up approach," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 1, pp. 98–104, 2022.

[53] K. Dmitriev, J. Marino, K. Baker, and A. E. Kaufman, "Visual analytics of a computer-aided diagnosis system for pancreatic lesions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2174–2185, 2019.

[54] M. Ahmad, S. F. Qadri, M. U. Ashraf, *et al.*, "Efficient liver segmentation from computed tomography images using deep learning," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[55] T. Siriapisith, W. Kusakunniran, and P. Haddawy, "Pyramid graph cut: Integrating intensity and gradient information for grayscale medical image segmentation," *Computers in Biology and Medicine*, vol. 126, p. 103 997, 2020.

[56] A. M. Awe, V. R. Rendell, M. G. Lubner, and E. R. Winslow, "Texture analysis: An emerging clinical tool for pancreatic lesions," *Pancreas*, vol. 49, no. 3, p. 301, 2020.

4. 3 Pancreas Instance Segmentation using Deep Learning Techniques

This paper proposes instance segmentation of CT pancreas images using a combined model of UNet and Watershed. In addition to the bounding box parameters and class labels that were previously described, semantic segmentation also yields other results. The result is a full, high-resolution, labelled PDAC image. Max Pooling enhances comprehension of "WHAT" in the visual by enlarging the receptive field. But it frequently forgets where things are. Semantic segmentation requires knowledge of both "WHAT" and "WHERE" in the image. Therefore, we require a method of up-sampling the image from low resolution to high resolution to recover the "WHERE" information. The preferred technique for up-sampling is transposed convolution, which uses backpropagation to learn parameters to transform low-resolution images into high-resolution images.

Pancreas Instance Segmentation Using Deep Learning Techniques

Wilson Bakasa and Serestina Viriri(✉)

School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban, South Africa
viriris@ukzn.ac.za

Abstract. The segmentation of pancreatic ductal adenocarcinoma (PDAC) images facilitates computer vision applications. This paper proposes using UNet, a fully convolutional neural network based on a residual neural network (ResNet34), as an encoder and turning point for boundary class separation, designed for biomedical applications. Image segmentation separates one image into a number of smaller pieces. When doing PDAC image segmentation tasks, these resulting fragments or numerous segments are helpful. Another important requirement for image segmentation tasks is masks and edge classes. TensorFlow and Keras are deep learning frameworks used to build U-Net architectures. UNet aims to capture the characteristics of both context and localisation. The main idea of the implementation is to use a continuous shrink layer and immediately use the upsampling operator to get a higher resolution output on the input image. A total of 800 images were used, 0.2 for testing and 0.8 for training. When using the original UNet model, the accuracy was 0.86. With the help of masking and boundary classes, binary images consisting of null or non-zero values, we got an accuracy of 0.89 from the segmentation tasks.

Keywords: Segmentation · UNet · ResNet-34 · Watershed · Tensorflow · Border classes · PDAC medical images

1 Introduction

Computer vision [1] is an interdisciplinary branch of science that studies how to make computers grasp complex information from digital photos or movies. It aims to construct things that the human visual system can perform by automating them. In recent years, the discipline of computer vision [2] has advanced thanks to deep learning. By utilising boundary classes [3–5] of pancreatic ductal adenocarcinoma (PDAC) images, the authors will discuss a computer vision problem called instance segmentation via Semantic segmentation. Although, there are several approaches that academics have developed to address this issue. We will focus on one specific architecture called U-Net [6] that uses ResNet-34 [7] as its foundation and a Fully Convolutional Network (FCN) Model to perform segmentation.

T. M. Ngatched Nkouatchah et al. (Eds.): PAAISS 2022, LNICST 459, pp. 205–223, 2023.
https://doi.org/10.1007/978-3-031-25271-6_13

This paper looks at instance segmentation of the pancreas using deep learning techniques. In Sect. 2, the original UNet model first analyses how the encoder and decoder are used in deep learning. The proposed model is discussed in Sect. 3, implementing ResNet34 as an encoder to UNet, plus Watershed to form a hybrid model. The max pooling and convolution operations are discussed in Sect. 4 under methods and techniques used in this paper. In Sect. 5, we looked at the experiments done to generate the border classes and the implementation of UNetet plus Watershed, including the training. Results are discussed in Sect. 6, showing the batch normalisation and max pooling from implementing the original against the proposed model.

1.1 Semantic Segmentation

In the computer vision task of semantic segmentation [8,9], each pixel of an image frame is categorised or labelled according to the class to which it belongs. ResNet and UNet are two common Convolutional Neural Networks (CNNs) used for segmentation tasks. This study uses the UNet architecture to achieve multi-class segmentation on the PDAC images dataset. Semantic image segmentation aims to assign a class of what is represented to each image pixel.

1.2 Instance Segmentation

The time needed to perform diagnostic tests on PDAC patients can be drastically decreased using machines to supplement radiologists' analysis. Semantic image segmentation aims to assign a class of what is represented to each pixel of PDAC images [10]. This process is known as a dense prediction since we predict every pixel in the image. The result is a high-resolution image with each pixel assigned to a certain class. Thus, it is PDAC image classification at the pixel level.

2 UNet

In the original work, [11] created the UNet, an architecture for semantic segmentation for Biomedical Image Segmentation. Two paths are present in the architecture. The encoder used to record the context of the image is called the contraction path, which is the first path. The encoder is essentially a standard stack of max-pooling and convolutional layers. The second method is the symmetric expanding path, or decoder, which uses transposed convolutions to provide exact localisation. Therefore, it is an end-to-end, FCN [12–14]; it has no dense layers and solely has convolutional layers. Any size photograph can be accepted.

The feature map must be up-sampled before each step in the expansion route, which is then followed by a 2×2 up-convolution that reduces the number of feature channels in half. Cropping is necessary since each convolution loses boundary pixels. There are a total of 23 convolutional layers in the network [11].

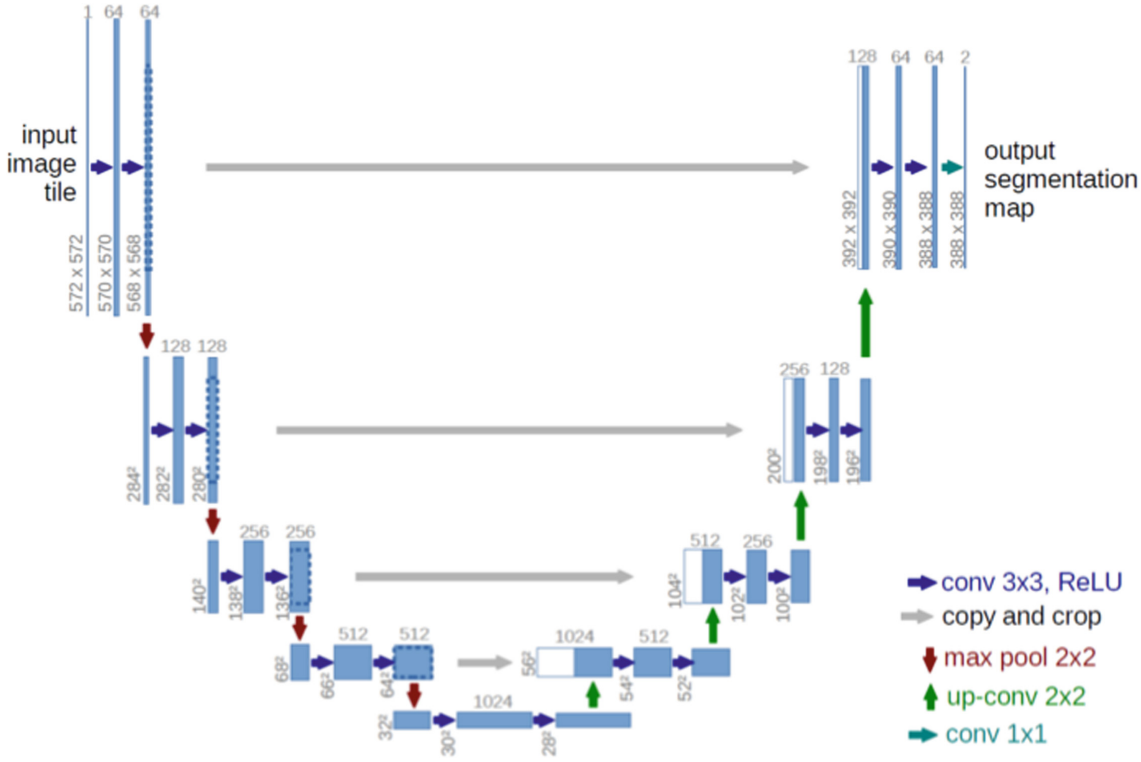


Fig. 1. 32×32 pixel original U-Net architecture.

According to the design Fig. 1, a model is applied to an input image before a few convolutional layers with the ReLU activation function are applied. The image size decreases from 572×572 to 570×570, then to 568×568. This reduction is because they used unpadded convolutions, which reduced the total dimensionality. We also see an encoder block on the left side, a decoder block on the right side, and the Convolution blocks [15,16].

The encoder block using the max-pooling layers of strides two continuously decreases the size of the image. The encoder architecture also includes recurrent convolutional layers with many filters. When we get to the decoder part, we see that the convolutional layers' number of filters starts to go down and that the subsequent layers gradually upsample the top layer [17].

2.1 Encoder

Because the encoder in this scenario has repetitive blocks as seen in Fig. 2, it is best to add functions to render the code modular. In these encoder blocks, ReLU will initiate the two 2D Convolution Layer (Conv2D) layers, which will be followed by the MaxPooling and Dropout layers. Each stage will include more filters due to the pooling layer, and the dimensionality of the features will decrease [18,19].

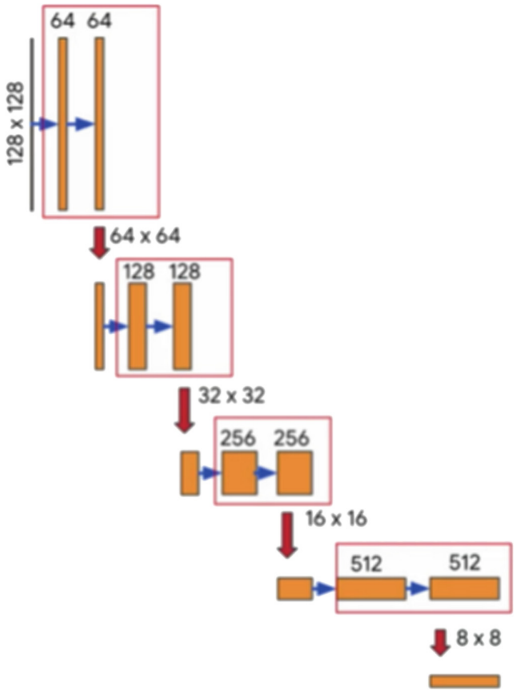


Fig. 2. Encoder left side

2.2 Decoder

The decoder Fig. 3 [20,21], which restores the features' original image size by upsampling, is the last component. We will concatenate the output of the relevant encoder block at each level of upsampling before passing it on to the following decoder block.

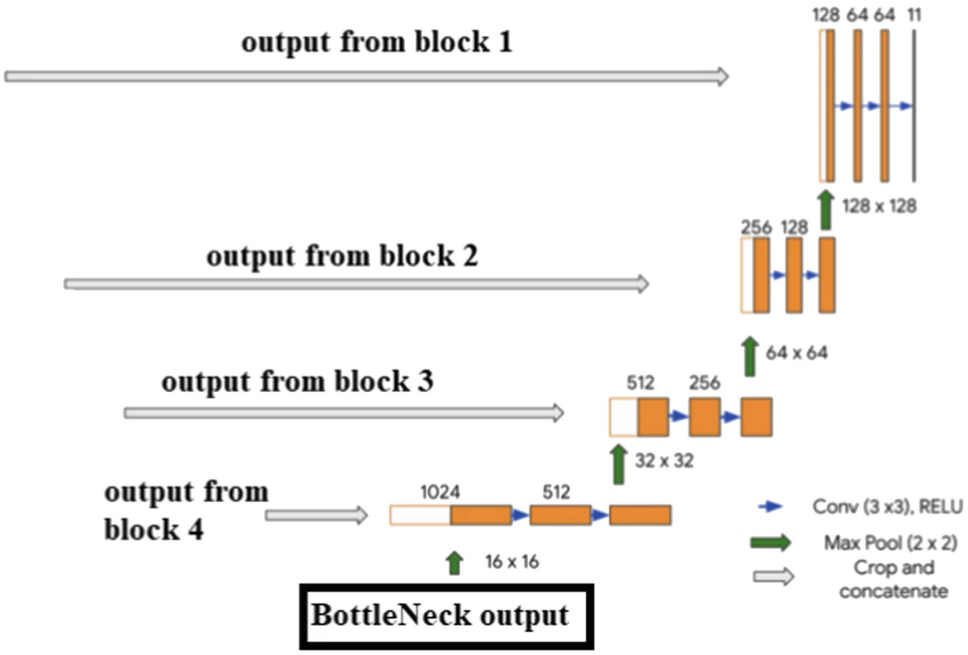


Fig. 3. Decoder right side

3 Proposed Model

This research suggests UNet [11] segmentation of PDAC images followed by object separation based on watersheds as in Fig. 4. In order to create boundary classes [5], the object properties will also be computed. The UNet backbone will be made of ResNet-34.

$$[Semanticsegmentation(UNet)withResNet34] + watershed$$
$$-- > Instancesegmentation$$

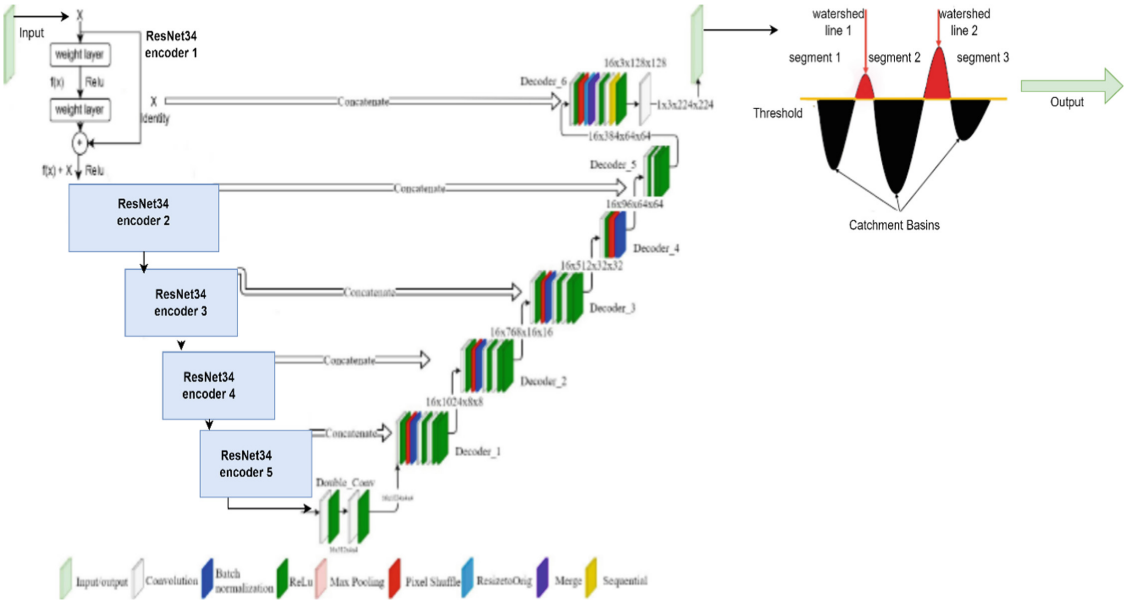


Fig. 4. The proposed model: ResNet-34 blocks as encoders 1–5. The decoder is the usual original UNet decoder. The output from the decoder is then fed onto Watershed for object separation using border classes.

3.1 Residual Networks (ResNet)

ResNet is a CNN architecture composed of numerous residual blocks (ResBlocks), which are described below in Fig. 5. Skip connections are what set ResNets apart from other CNNs. ResNet is best used as the encoder because it addresses the vanishing gradient problem. Accuracy does not increase or degrade when additional layers have been added, but training grows slower as a result. Because of the network's skip connections, this is conceivable [22–24].

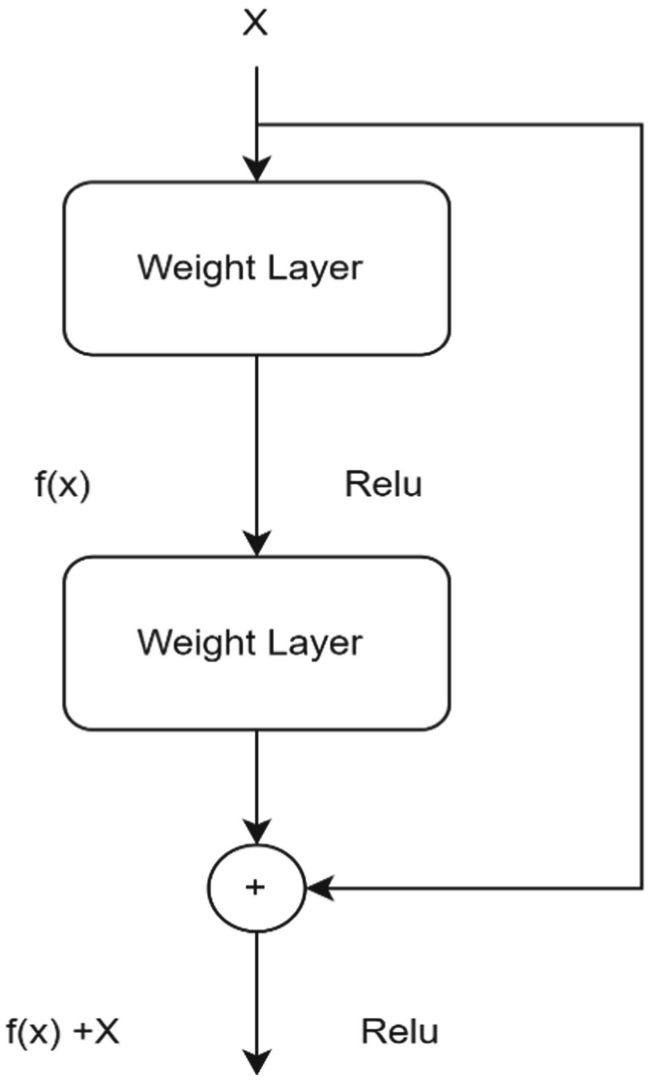


Fig. 5. ResNet-34 building block

Two connections from the input are present in each ResBlock, one of which passes across a series of linear functions, batch normalization, and convolutions. These convolutions and functions are bypassed by the identity when it crosses or skips connections [7].

The combined tensor outputs of the two connections are used. Adding cross-connections [25] between network layers enables vast areas to be bypassed if necessary, resulting in the lost surface. Making it much simpler to train the model with the right weights to minimise the loss.

3.2 Watershed

The image measure module, the "region props" function, extracts attributes of observed classes and produces relevant parameters for each item. Remove any backgrounds or areas that could be considered objects [26–28].

4 Methods and Techniques

Utilising the Keras API [29–31], the paper implemented UNet in Python. Convolution, max-pooling, down- and upsampling, and transposed convolution were all used in the implementation. The study covers UNet image segmentation followed by watershed-based object separation. Additionally, object properties will be computed. Uses a model trained on the common UNet framework based on ResNet-34 for semantic segmentation, implementing Watershed in border classes separation. Segmenting instances is the outcome of all this.

4.1 Max Pooling Operation

We have less parameters in the network because of the pooling function, Fig. 6, which shrinks the size of the feature map [32].

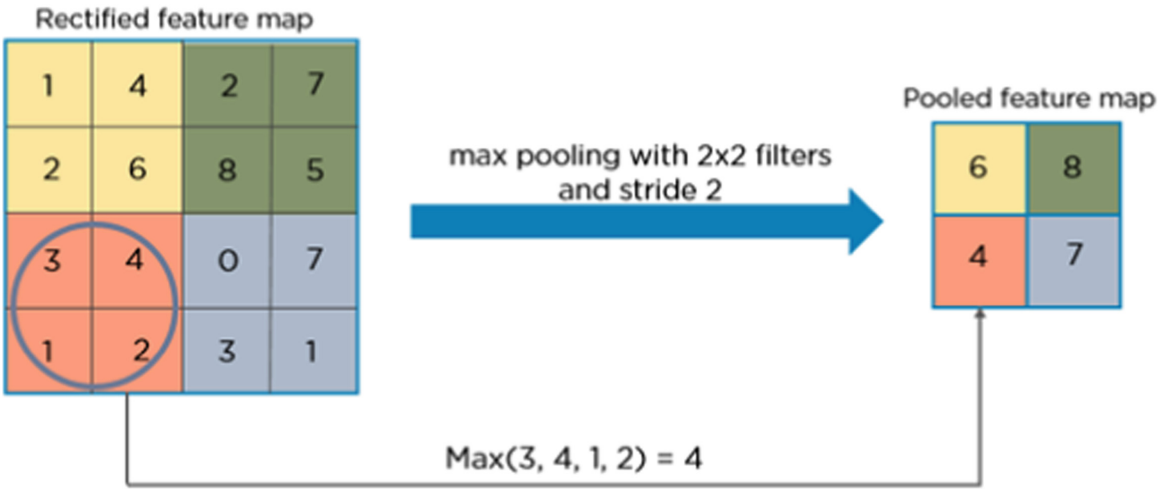


Fig. 6. Max pooling operations

In essence, we choose the highest pixel value from each 2×2 block of the input feature map to create a pooled feature map. Strides and filter size are two crucial [33] processes. The goal is to keep the pixels with the highest values from each region and discard the data that best fits the image's context. The fact that the convolution decreases the image size and, in particular, the pooling operation is crucial to keep in mind.

The term "downsampling" [34,35] refers to this method. In the illustration above, the image size was 4×4 *before and* 2×2 after pooling.

4.2 ResNet-34 Encoder

The encoder/downsampling portion of the UNet can utilise a ResNet-34 [22] (the left half of the U). ResNet-34, a 34-layer ResNet architecture, was employed in our models because it is more efficient, trains quicker than ResNet-50 [36], and

requires less memory. Using a pre-trained model greatly reduces training time for an image generation/prediction model, which is necessary for the model to understand how to carry out its prediction. The features that need to be found and enhanced are then known to the model as a starting point. Model and weights already trained on PDAC images are used as input, which turns the ResNet-34 encoder into a UNet with cross-connections.

5 Experiments

The U-Net blocks are more akin to DenseBlocks than ResBlocks [19] since their outputs are concatenated. However, stride two convolutions shrink the grid size once more, preventing excessive memory consumption.

5.1 Generate Border Classes Steps

With binary masks, generate border pixels. When training a multi-class UNet semantic segmenter, these border pixels can be added as a new class. The benefit is that we can do watershed and segment "instances" using border pixels.

1. `Call function to define the border.`
2. `Erode some pixels into objects and dilate to outside the objects.`
3. `This region would be the border. Replace border pixel value with something other than 255.`
4. `Start by eroding edge pixels.`
5. `Define the kernel size for dilation based on the desired border size, adding 1 to keep it odd.`
6. `Replace 255 values with 127 for all pixels. Eventually, we will only define border pixels with this value.`
7. `The above-dilated image converts the eroded object parts to pixel value 255.`
8. `What is remaining with a value of 127 would be the boundary pixels.`

5.2 Implementing UNet Plus Watershed

Instance segmentation was done via semantic segmentation (U-Net) plus the Watershed. The steps are as follows:

1. `Load the U-Net(ResNet-34 backbone) trained model and corresponding weights.`
2. `The test image, which needs to be segmented, should be loaded and processed.`
3. `Predict and threshold for probabilities that are greater than 0.5.`
4. `Binaries the U-Net result to create an image.`

5. Utilizing OTSU, threshold image to binary. The value 255 will be applied to each thresholded pixel.
6. Morphological operations to eliminate small noise - opening. To get rid of holes, we can use closing.
7. Finding certain foreground areas using distance transform and thresholding.

Verify the total number of regions discovered before and after using the "Opening" function. We now understand that the areas in the middle of cells are unquestionably necessary. The distant region serves as the backdrop. We can utilise "erode" to extract specific regions that we need to. However, certain PDAC pictures feature regions that contact, making erosion alone ineffective.

The optimal strategy would be to threshold after distance transform to distinguish contacting items. Let us begin by determining a specific backdrop area and increasing the cell boundary to the background by distorting pixels a few times. In this manner, everything that is left will be history. Our ambiguous area is located between the foreground and the background. The watershed [26] should discover this location.

i The foreground points'intensities are adjusted to correspond to the distances between them and the nearest 0 value, which is the boundary.
ii Let's threshold the distance transform by beginning at its maximum value, $\frac{1}{2}$.
iii We now make a marker and give the inside sections labels. For some places, both the foreground and the backdrop will be given positive numbers. Regions that are unknown will be labeled 0. Let's use Connected Components as markers.
iv One issue is that all backdrop pixels have a value of 0, which indicates to the Watershed that this area is unknowable. For a specific backdrop to be 10 rather than 0, let us add 10 to all labels.
v Put a 0 marker in the area that is currently unknown.
vi Now we are ready for watershed filling.
vii Now that regions have been recognized, it is time to extract their properties; each object's valuable parameters are calculated using the "region props" function in the skimage measure module.

5.3 TensorFlow Implementation of UNet

We will implement TensorFlow to the UNet architecture for the computation of the model. We shall stick to the TensorFlow library for this model. We will import all necessary libraries to build our UNet architecture before starting from scratch. However, we will make a few important adjustments that will enhance the model's overall performance while making it a little less complicated [37,38].

Importing the Required Libraries. We will construct the UNet architecture using the TensorFlow deep learning framework. As a result, we will import the Keras framework and the TensorFlow library, which are now both part of the TensorFlow model structures. According to what we know, one of the primary imports for the UNet architecture is the activation function ReLU for the basic modelling structure, along with the convolutional layer, the max-pooling layer, an input layer, and the layer with the activation function. Then, we will conduct an upsampling for the decoder blocks we want to use by adding additional layers, such as the Conv2DTranspose layer. Additionally, we will stabilise the training process with batch normalisation layers and combine the required skip connections with concatenating layers [37,39].

Building the Convolution Block. After importing the required libraries, we can move on to creating the UNet architecture. We may do this in a single class by correctly specifying each argument's value and executing the process until the end or a few iterative blocks. The latter strategy will be employed since, with the help of a few building pieces, most people can more easily understand the model architecture of U-Net. The three iterative blocks we will employ are the convolution operation block, the ResNet-34 (encoder block), and the decoder block, as shown in the architecture representation. By using these three construction blocks, we can create the U-Net architecture. Now that we have gone through and comprehended each of these function code sections [15,16].

Constructing the Encoder and Decoder Blocks. Consecutive inputs from the top layer to the base layer will be used in the ResNet-34 encoder architecture. The convolutional block has two convolutional layers and corresponding batch normalisation and ReLU layers. The block will make up the ResNet-34 encoder function as stated [34,35].

Construct the UNet Architecture. The U-Net architecture contains a high number of processing blocks. Thus the entire structure might be huge if we tried to create it from scratch in a single layer. By dividing our functions into the three independent code blocks of convolutional operation, ResNet-34 (encoder) structure, and decoder structure, we can easily develop the UNet architecture in a few lines of code. The input layer, which contains the various forms from our input image, will be used [11].

5.4 Training

Since there are two classes, the model is built using the Adam optimiser, and we utilise the binary cross-entropy loss function (salt and no salt). Keras callbacks are employed to implement:

i monitor if the validation loss remains unchanged for five consecutive epochs, the learning rate decays.

ii Early termination, if the validation loss does not decrease during ten consecutive epochs.
iii Only keep the weights if the validation loss improves. A batch size of 32 is employed.

6 Results

The diagram below contrasts the segmented PDAC image and its markers with the original PDAC image from the original dataset, shown in Fig. 8. The optimisation approach known as the local gradient is frequently used to train neural networks and machine learning models. Using training data, these models learn new things over time. In gradient descent, the cost function acts as a barometer by evaluating the precision of each iteration of parameter changes. Until the function is close to or equal to zero, the model will keep altering its parameters to provide the least amount of error. Once their accuracy has been improved, machine learning models can be useful tools for computer science and artificial intelligence (AI) applications (Fig. 7).

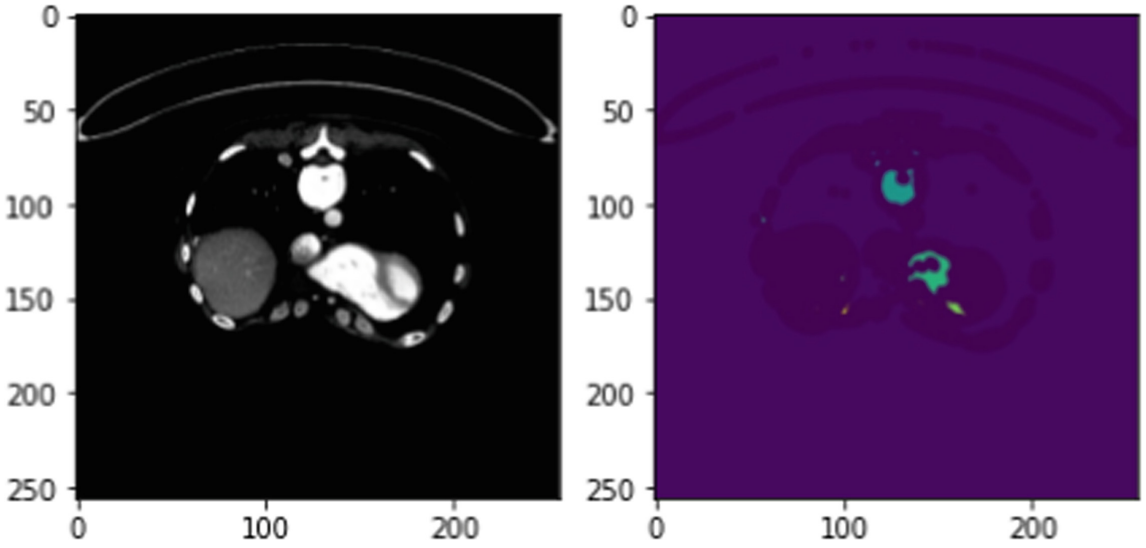


Fig. 7. Showing the original image and the testing label image.

The training of deep neural networks using batch normalisation [40], which normalises the contributions to a layer for each mini-batch, is illustrated in Fig. 9. As a result, the learning process is stabilised, and a significant reduction in the number of training epochs required to build deep neural networks. Between one hidden layer and the subsequent hidden layer, there is just another network layer called "Batch Norm". Before sending them on as the input of the next hidden layer, it has the responsibility of normalising the outputs from the previously hidden layer.

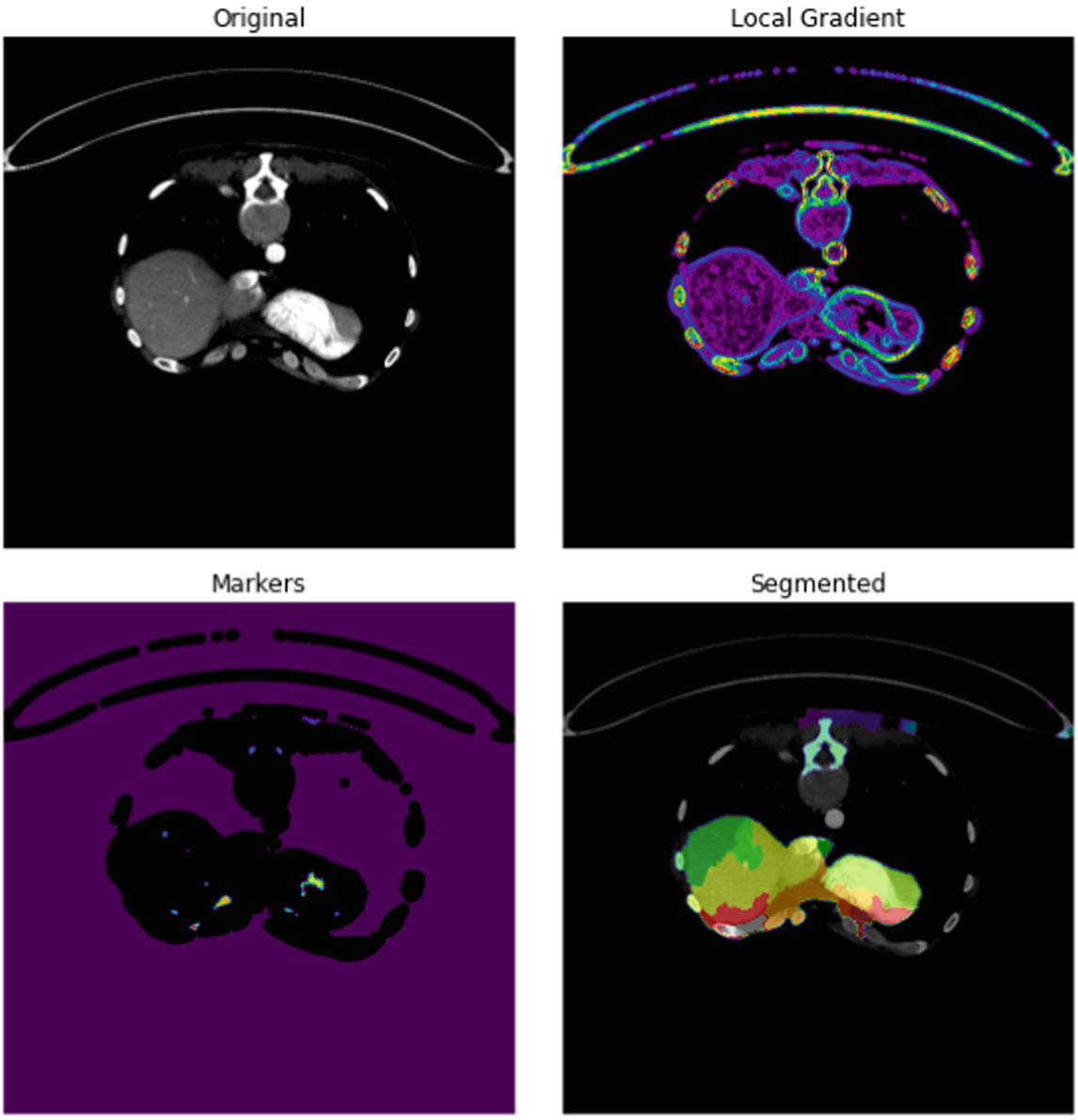


Fig. 8. Diagram for the segmented image

Additionally displayed are the total, trainable, and non-trainable model parameters. In the context of deep learning, they are the weight and bias characteristics of the training data that will be learned during the learning process. A parameter is frequently used to gauge a model's performance.

```
100%|          | 800/800 [00:15<00:00, 50.02it/s]
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 256, 256, 3)]     0

 conv2d (Conv2D)             (None, 256, 256, 64)      1792

 batch_normalization (BatchN  (None, 256, 256, 64)     256
 ormalization)

 activation (Activation)     (None, 256, 256, 64)      0

 conv2d_1 (Conv2D)           (None, 256, 256, 64)      36928

 batch_normalization_1 (Batc  (None, 256, 256, 64)     256
 hNormalization)

 activation_1 (Activation)   (None, 256, 256, 64)      0

 max_pooling2d (MaxPooling2D) (None, 128, 128, 64)     0

 conv2d_2 (Conv2D)           (None, 128, 128, 128)     73856

 batch_normalization_2 (Batc  (None, 128, 128, 128)    512
 hNormalization)

 activation_2 (Activation)   (None, 128, 128, 128)     0

 conv2d_3 (Conv2D)           (None, 128, 128, 128)     147584

 batch_normalization_3 (Batc  (None, 128, 128, 128)    512
 hNormalization)

 activation_3 (Activation)   (None, 128, 128, 128)     0

 max_pooling2d_1 (MaxPooling  (None, 64, 64, 128)      0
 2D)
```

Fig. 9. Batch normalisation and max pooling from implementing UNet model.

Epochs are represented in Fig. 10 and refer to training the neural network using all training data for a single cycle.

Figure 11 depicts the reconstructed image. Deep learning-based picture reconstruction is an effective use of artificial intelligence and cutting-edge machine learning. Deep learning has been widely used in computer vision and image analysis, which deal with existing images, enhance them and derive features from them (Figs. 12 and 13).

We may plot the train and validation loss to see how the training went as in Fig. 14, which should show typically decreasing values per epoch. The training

```
=================================================================
Total params: 27,923,523
Trainable params: 27,911,747
Non-trainable params: 11,776
_________________________________________________________________
None
Epoch 1/5
1/1 [==============================] - 110s 110s/step - loss: 0.2671 -
accuracy: 0.6961
Epoch 2/5
1/1 [==============================] - 103s 103s/step - loss: 0.1819 -
accuracy: 0.8558
Epoch 3/5
1/1 [==============================] - 101s 101s/step - loss: 0.1370 -
accuracy: 0.7828
Epoch 4/5
1/1 [==============================] - 101s 101s/step - loss: 0.1078 -
accuracy: 0.8763
Epoch 5/5
1/1 [==============================] - 101s 101s/step - loss: 0.0897 -
accuracy: 0.8623
```

Fig. 10. Loss and accuracy after 5 epochs obtained from implementing UNet model

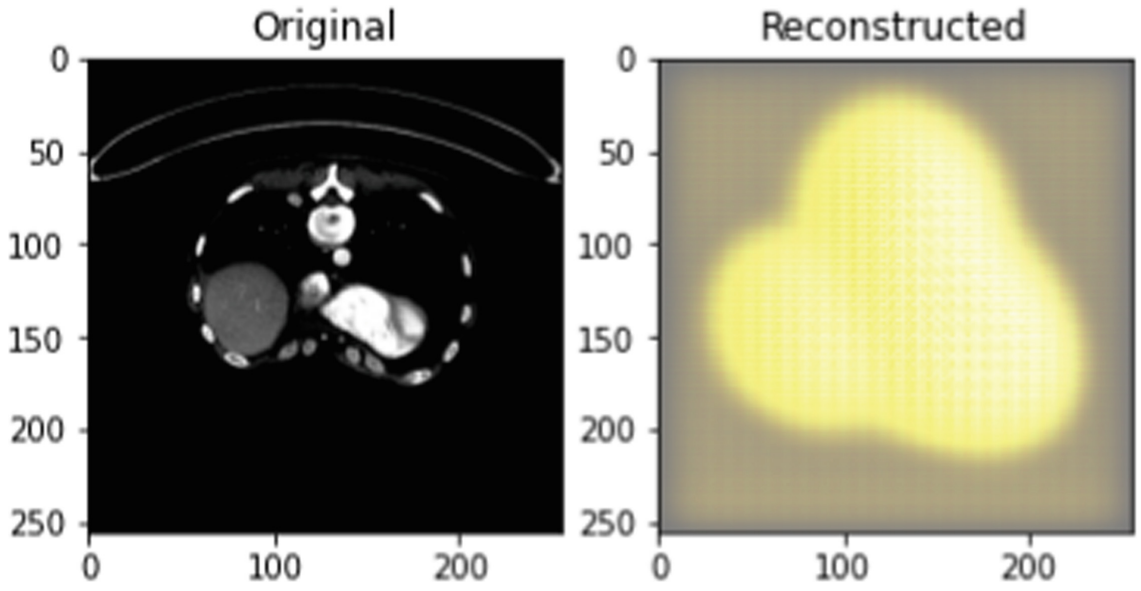


Fig. 11. The reconstruction of images

loss is the degree to which a deep learning model matches the training set of data. It assesses the model's variance on the training set, in other words. Keep in mind that the dataset utilised train the model was split up into the training set. The training loss is computed using the total mistakes for each sample in the training set.

Remember that each batch also concludes with a measurement of the training loss. A training loss curve is typically displayed to demonstrate this. On the other hand, a statistic known as validation loss is used to assess a deep learning

```
 batch_normalization_30 (BatchN  (None, 256, 256, 12  512
['conv2d_31[0][0]']
 ormalization)                  8)

 activation_30 (Activation)     (None, 256, 256, 12  0
['batch_normalization_30[0][0]']
                                8)

 conv2d_32 (Conv2D)             (None, 256, 256, 12  147584
['activation_30[0][0]']
                                8)

 batch_normalization_31 (BatchN  (None, 256, 256, 12  512
['conv2d_32[0][0]']
 ormalization)                  8)

 activation_31 (Activation)     (None, 256, 256, 12  0
['batch_normalization_31[0][0]']
                                8)

 max_pooling2d_9 (MaxPooling2D)  (None, 128, 128, 12  0
['activation_31[0][0]']

:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

batch_normalization_45 (BatchN  (None, 512, 512, 64  256
['conv2d_46[0][0]']
 ormalization)                  )

 activation_45 (Activation)     (None, 512, 512, 64  0
['batch_normalization_45[0][0]']
                                )

 conv2d_47 (Conv2D)             (None, 512, 512, 1) 65
['activation_45[0][0]']
==================================================================
Total params: 31,055,297
Trainable params: 31,043,521
Non-trainable params: 11,776
```

Fig. 12. Batch Normalisation and Max Pooling from implementing the proposed model of UNet (ResNet-34) plus Watershed.

model's performance on the validation set. The validation set is a portion of the dataset used to evaluate the model's performance. Similar to how we calculated the training loss, we calculated the validation loss by summing the mistakes for each sample in the validation set.

```
Epoch 1/5
25/25 [==============================] - 140s 6s/step - loss: 0.0211 -
accuracy: 0.4535
Epoch 2/5
25/25 [==============================] - 140s 6s/step - loss: 0.0092 -
accuracy: 0.5791
Epoch 3/5
25/25 [==============================] - 140s 6s/step - loss: 0.0071 -
accuracy: 0.8354
Epoch 4/5
25/25 [==============================] - 140s 6s/step - loss: 0.0060 -
accuracy: 0.8735
Epoch 5/5
25/25 [==============================] - 139s 6s/step - loss: 0.0052 -
accuracy: 0.8932
```

Fig. 13. Loss and accuracy after five epochs obtained from implementing the proposed model of UNet (ResNet-34) plus Watershed

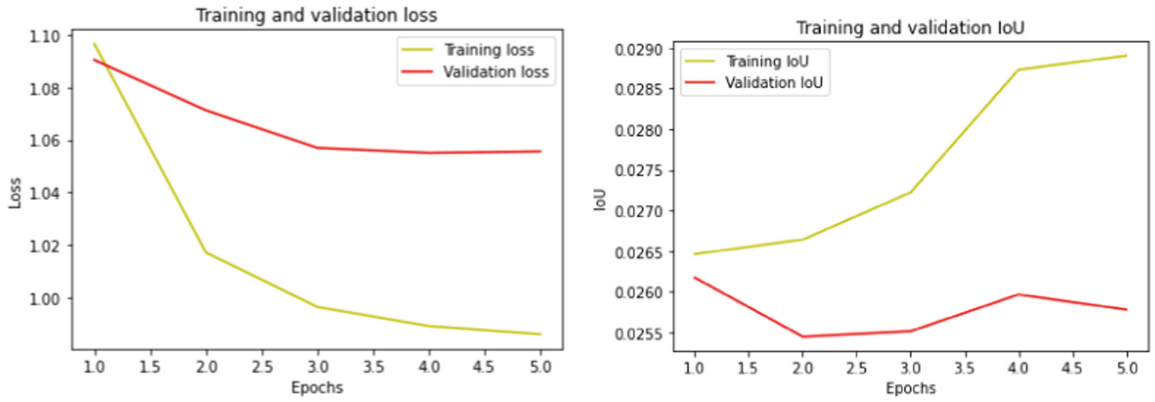


Fig. 14. Training against the Validation loss and IoU graph

7 Conclusion

Apart from a class label and a set of bounding box parameters, semantic segmentation generates further information, as was previously mentioned. The final product is a comprehensive, high-resolution PDAC image with labelled pixels. By widening the receptive field, Max Pooling facilitates understanding of "WHAT" in the visual. However, it frequently forgets the location of the objects. Semantic segmentation requires an understanding of "WHAT" in the image and "WHERE". Therefore, we must find a means to up-sample the image from low resolution to high resolution to recover the "WHERE" information. Backpropagation is used by transposed convolution to learn parameters. The preferred method for upsampling is to transform a low-resolution image into a high-resolution one.

Stride two convolutions gradually shrink the grid size, preventing excessive memory consumption. A modern GPU can quickly compute the segmentation of images with a size of 512×512 using this UNet architecture. However, the UNet (ResNet-34) plus watershed model's modern efficiency, speed, and simplicity will be enhanced PDAC image segmentation tasks. Using the value of convolution

as "same" is one of the changes made to this structure because numerous subsequent research studies have demonstrated that this alteration had no detrimental effects on the construction of the building. Additionally, the original architecture did not use this feature because batch normalisation was only introduced in 2016. Nevertheless, batch normalisation will be a part of our model implementation because it typically produces the best outcomes. In subsequent publications, we will investigate the functionality of the PyTorch-based U-Net architecture and a few alternative model structures.

References

1. Wiley, V., Lucas, T.: Computer vision and image processing: a paper review. Int. J. Artif. Intell. Res. **2**(1), 29–36 (2018)
2. Thevenot, J., López, M.B., Hadid, A.: A survey on computer vision for assistive medical diagnosis from faces. IEEE J. Biomed. Health Inform. **22**(5), 1497–1511 (2017)
3. Wang, R., Chen, S., Ji, C., Fan, J., Li, Y.: Boundary-aware context neural network for medical image segmentation. Med. Image Anal. **78**, 102395 (2022)
4. Chen, X., Williams, B.M., Vallabhaneni, S.R., Czanner, G., Williams, R., Zheng, Y.: Learning active contour models for medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11632–11640 (2019)
5. Asaturyan, H., Thomas, E.L., Fitzpatrick, J., Bell, J.D., Villarini, B.: Advancing pancreas segmentation in multi-protocol MRI volumes using Hausdorff-Sine loss function. In: Suk, H.-I., Liu, M., Yan, P., Lian, C. (eds.) MLMI 2019. LNCS, vol. 11861, pp. 27–35. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32692-0_4
6. Li, X., Chen, H., Qi, X., Dou, Q., Chi-Wing, F., Heng, P.-A.: H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans. Med. Imaging **37**(12), 2663–2674 (2018)
7. Koonce, B.: ResNet 34. In: Koonce, B. (ed.) Convolutional Neural Networks with Swift for Tensorflow, pp. 51–61. Springer, Heidelberg (2021). https://doi.org/10.1007/978-1-4842-6168-2_5
8. Guo, Y., Liu, Y., Georgiou, T., Lew, M.S.: A review of semantic segmentation using deep neural networks. Int. J. Multimed. Inf. Retrieval **7**(2), 87–93 (2018)
9. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J.: A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857 (2017)
10. Bakasa, W., Viriri, S.: Pancreatic cancer survival prediction: a survey of the state-of-the-art. Comput. Math. Methods Med. **2021**, 1–17 (2021)
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
12. Roth, H.R., et al.: An application of cascaded 3D fully convolutional networks for medical image segmentation. Computer. Med. Imaging Graph. **66**, 90–99 (2018)
13. Dolz, J., Desrosiers, C., Ayed, I.B.: 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. NeuroImage **170**, 456–470 (2018)

14. Dai, J., He, K., Li, Y., Ren, S., Sun, J.: Instance-sensitive fully convolutional networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 534–549. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_32
15. Teng, L., Li, H., Karim, S. DMCNN: a deep multiscale convolutional neural network model for medical image segmentation. J. Healthc. Eng. **2019**, 1–11 (2019)
16. Khanna, A., Londhe, N.D., Gupta, S., Semwal, A.: A deep residual U-net convolutional neural network for automated lung segmentation in computed tomography images. Biocybern. Biomed. Eng. **40**(3), 1314–1327 (2020)
17. Zongwei Zhou, Md., Siddiquee, M.R., Tajbakhsh, N., Liang, J.: UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging **39**(6), 1856–1867 (2019)
18. Sathananthavathi, V., Indumathi, G.: Encoder enhanced atrous (EEA) UNet architecture for retinal blood vessel segmentation. Cogn. Syst. Res. **67**, 84–95 (2021)
19. Kiran, I., Raza, B., Ijaz, A., Khan, M.A.: DenseRes-Unet: segmentation of overlapped/clustered nuclei from multi organ histopathology images. Comput. Biol. Med. **143**, 105267 (2022)
20. Choi, S.: Utilizing Unet for the future traffic map prediction task traffic4cast challenge 2020. arXiv preprint arXiv:2012.00125 (2020)
21. Arrastia, J.L., et al.: Deeply supervised UNet for semantic segmentation to assist dermatopathological assessment of basal cell carcinoma. J. imaging **7**(4), 71 (2021)
22. Al-Moosawi, N.M.A.-M.M., Khudeyer, R.S., et al. Resnet-34/dr: A residual convolutional neural network for the diagnosis of diabetic retinopathy. Informatica **45**(7), 115–124 (2021)
23. Xue, Yu., Wang, Y., Liang, J., Slowik, A.: A self-adaptive mutation neural architecture search algorithm based on blocks. IEEE Comput. Intell. Mag. **16**(3), 67–78 (2021)
24. Zhou, S., Nie, D., Adeli, E., Yin, J., Lian, J., Shen, D.: High-resolution encoder-decoder networks for low-contrast medical image segmentation. IEEE Trans. Image Process. **29**, 461–475 (2019)
25. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Trans. Med. Imaging **39**(7), 2494–2505 (2020)
26. Zanaty, E., Afifi, A.: A watershed approach for improving medical image segmentation. Comput. Methods Biomech. Biomed. Eng. **16**(12), 1262–1272 (2013)
27. Ramesh, K.K.D., Kiran Kumar, G., Swapna, K., Datta, D., Rajest, S.S.: A review of medical image segmentation algorithms. EAI Endors. Trans. Pervasive Health Technol. **7**(27), e6 (2021)
28. Shen, T., Wang, Y.: Medical image segmentation based on improved watershed algorithm. In: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 1695–1698. IEEE (2018)
29. Arnold, T.B.: kerasr: R interface to the keras deep learning library. J. Open Source Softw. **2**(14), 296 (2017)
30. Manaswi, N.K.: Understanding and working with Keras. In: Manaswi, N.V. (ed.) Deep Learning with Applications Using Python, pp. 31–43. Springer, Heidelberg (2018). https://doi.org/10.1007/978-1-4842-3516-4_2
31. Gulli, A., Kapoor, A., Pal, S.: Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API. Packt Publishing Ltd. (2019)
32. Murray, N., Perronnin, F.: Generalized max pooling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2473–2480 (2014)

33. Loussaief, S., Abdelkrim, A.: Convolutional neural network hyper-parameters optimization based on genetic algorithms. Int. J. Adv. Comput. Sci. Appl. **9**(10) (2018)
34. Chen, Q., Xu, J., Koltun, V.: Fast image processing with fully-convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2497–2506 (2017)
35. Zhang, Y., Zhao, D., Zhang, J., Xiong, R., Gao, W.: Interpolation-dependent image downsampling. IEEE Trans. Image Process. **20**(11), 3291–3296 (2011)
36. Wen, L., Li, X., Gao, L.: A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Comput. Appl. **32**(10), 6111–6124 (2020)
37. Qamar, S., Jin, H., Zheng, R., Ahmad, P., Usama, M.: A variant form of 3D-UNet for infant brain segmentation. Futur. Gener. Comput. Syst. **108**, 613–623 (2020)
38. Cai, S., Tian, Y., Lui, H., Zeng, H., Yi, W., Chen, G.: Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quant. Imaging Med. Surg. **10**(6), 1275 (2020)
39. Tuan, T.A., Tuan, T.A., Bao, P.T.: Brain tumor segmentation using bit-plane and UNET. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11384, pp. 466–475. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_41
40. Daimary, D., Bora, M.B., Amitab, K., Kandar, D.: Brain tumor segmentation from MRI images using hybrid convolutional neural networks. Proc. Comput. Sci. **167**, 2419–2428 (2020)

4. 4 Intelligent Automated Pancreas Segmentation using U-Net Model Variants

The best model for the job cannot be determined solely by performance. The model must satisfy additional requirements, including memory needs, training and prediction times, interpretability, and data format. The MA can test them on a validation dataset when it is challenging to select the optimal model from a few alternatives for new data. The MA might compare these metrics to evaluate each model and reach a judgement. Analysing the requirements and desired outcomes is crucial before deciding whether to build a deep learning model. Increased accuracy and performance are the benefits, even though it might take more time and work.

IA will assist in reducing model training and prediction time because pre-trained models are used. Again, accuracy is enhanced by providing a baseline for model comparison. This paper offers an intelligent automation model to help readers choose the optimal model for a particular segmentation assignment.

Intelligent Automated Pancreas Segmentation using U-Net Model Variants

Wilson Bakasa[0000−0003−0927−1301] and Serestina Viriri[0000−0002−2850−8645]

School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal, Durban, South Africa
219098448@stu.ukzn.ac.za; viriris@ukzn.ac.za

Abstract. Various algorithms and models can be generated for any given machine-learning problem. Many options are beneficial, but deciding which model to implement in segmentation is critical. Even though we have a variety of performance metrics to evaluate a model, it is not prudent to implement every algorithm for every problem. This will take significant time and effort to decide the best model for the task. As a result, knowing how to choose the best algorithm for a given task is critical. Three models will be trained for the pancreas CT image segmentation task: U-Net, Attention U-Net, and Residual Attention U-Net. The authors examine the factors that can aid in selecting a model that best suits any segmentation task and with specific requirements in this article. We'll do this by considering several factors that can help us narrow our choices. A Model Analyser (MA) and an Intelligent Automation Model implement these factors in the Inference Engine. Understanding these factors will assist the MA in understanding the task that each model can perform, depending on the complexity of a problem, memory requirements, training time, number of features and parameters, training data size, and accuracy or interpretability. Many factors influence which algorithm is used, from the type of problem at hand to the expected output. This guide provides several considerations when determining the best U-Net model for a new dataset. Accuracy and mean IoU are the performance metrics used to evaluate segmentation models.

Keywords: Pancreas · Segmentation · U-Net · Attention U-Net · Residual Attention U-Net · Intelligent Automation Model.

1 Introduction

Image segmentation [1] is a computer vision task that divides an image into multiple areas by labelling each pixel. It provides significantly more information about an image than object detection, which draws a bounding box around the detected object, or image classification, which labels the object. Segmentation is useful in real-world applications such as medical imaging segmentation. Image segmentation is classified into two types [2]:

1. Semantic segmentation: classify each pixel with a label.
2. Instance segmentation: classify each pixel and differentiate each object instance.

U-Net [3] is a semantic segmentation technique developed initially for medical imaging segmentation. It was one of the first deep-learning segmentation models. The U-Net architecture was later modelled into other variants, such as Attention U-Net [4] and Residual Attention U-Net [5].

2 Related Work

Statistical shape models [6], or multi-atlas techniques [7], [8] were used in early work on pancreas segmentation from abdominal CT. Atlas approaches, in particular, benefit from implicit shape constraints enforced by manual annotation propagation. Dice similarity coefficients (DSC) for atlas-based frameworks range from 69.6% to 73.9% in public benchmarks such as the TCIA dataset [9]. In [10], a classification-based framework is proposed to alleviate the atlas' reliance on image registration.

Importantly, previous applications of Attention U-Net have only been in medical settings, such as brain tumour segmentation [11], liver CT scan segmentation [12], and gland segmentation [13]. Previously, U-Net was used to successfully segment deforestation in the Amazon Rainforest using Sentinel-2 satellite imagery [14], and this paper aims to investigate the incorporation of an attention mechanism into U-Net to improve upon U-Net.

Several methods for segmenting surgical instruments have recently been proposed. Ozturk et al. [15] presented a network based on Fully Convolutional Networks (FCN) and optic flow to solve surgical instrument occlusion and deformation problems. RASNet [4] used an attention module to draw attention to the target region and improve feature representation. Olivier et al. [16] proposed a novel U-shape network for instrument segmentation and pose estimation simultaneously. Arunkumar et al. [17] used a method that combined the recurrent and convolutional networks to improve segmentation accuracy.

3 Variants of U-Net Model

The variants of the U-Net model that were experimented on in this paper are U-Net, Attention U-Net and Residual Attention U-Net.

3.1 U-Net Model for image segmentation:

Although some interesting non-learning algorithms, such as the Otsu method or the Watershed algorithm, have been used for image segmentation, most of today's real-world segmentation problems are solved by training encoder-decoder convolutional neural networks [CNNs]. U-Net is one of the most well-known network architectures. U-net is made up of two parts [18]:

1. encoder - converts an image into an abstract representation of image features by employing a series of convolutional blocks that gradually decrease the height and width of the representation while increasing the number of channels corresponding to image features.
2. decoder decodes an image representation into a binary mask by applying a series of up-convolutions that gradually increase the representation's height and width to the original image's size and decrease the number of channels to the number of classes we are segmenting.

In addition, U-Net implements skip connections that connect corresponding encoder and decoder levels. They allow the model to avoid "losing" features extracted by earlier encoder blocks, which improves segmentation performance [19]. U-Net extends the standard FCN architecture by introducing skip connections, which allow blocks of layers in the contraction phase to send their output directly to blocks in the expansion phase, greatly improving the ability to extract high-level features from images.

3.2 Attention U-Net Model for image segmentation:

Attention The U-Net architecture [13] is the foundation for U-Net, a fully convolutional network (FCN), a family of neural networks distinguished by an encoder-decoder. These are intended for semantic segmentation, also referred to as pixel-wise classification. The goal of an attention mechanism is to mimic the human ability to direct focus or concentrate on specific stimuli. This is done in the domain of neural networks by learning which parts of the input to focus attention on.

Attention U-Net is created by adding an attention gate to the U-Net skip connection. Rather than concatenating each up-scaled layer in the expansion phase with the corresponding contraction-phase layer, the up-scaled layer is concatenated with the attention mechanism output, which is a function of the pre-up-scaled layer and the aforementioned contraction-phase layer [20].

3.3 Residual Attention U-Net Model for image segmentation:

The Residual Attention U-Net (RAUNet) employs an encoder-decoder architecture to obtain high-resolution masks. To extract semantic features, the encoder is ResNet34 [5] pre-trained on ImageNet. It contributes to model reduction and faster inference. A new augmented attention module (AAM) is designed in the decoder to fuse multi-level features and capture global context. Furthermore, transposed convolution is used to perform up-sampling to obtain refined edges.

4 Methods and Techniques

To do experiments, the TCIA dataset was used. It was randomly divided into Dataset A and Dataset B. Dataset A was used to train the models, and then Dataset B was used as new data to test for model selection.

4.1 Model Selection Factors

To come up with the decision to select the right model to use for the segmentation of every new data, the MA will look at the following factors:

Number of Features and Parameters: Longer training usually results in better, more accurate model performance. If there is time to let the model train longer, more features and parameters can be specified to interpret [21].

Memory requirements: Many models can be applied if the entire dataset can be loaded into the computer's RAM. If this is not possible, incremental learning models can be used. Incremental learning is a machine learning method in which input data is continuously used to extend the knowledge of an existing model, designed to adapt to new data without forgetting previous knowledge, eliminating the need to retrain the model [22].

Training time: The time it takes an algorithm to learn and create a model is called training time. It is, therefore, critical to consider the time required to train the model. It is chosen if a pre-trained model can reduce training time. The long time it takes to train a neural network model is well-known. Some models require different training times depending on the number of CPU cores used [23].

Prediction time: Prediction time is the time it takes the model to make predictions. Even if an algorithm produces good results, it is useless if it is slow at making predictions. However, It is worth noting that accuracy is more important than prediction time for medical imaging. This is true in situations like cancerous cell segmentation [24].

The number of data points and features: Algorithms such as neural networks perform well with large amounts of data and many features. However, some models only support a limited number of features. The size of the data and the number of features should be considered when choosing a model [17].

Data format: Data is frequently derived from a combination of open-source and custom data resources, and as a result, it can be in various formats. Categorical and numerical data formats are the most common. Any given dataset may contain only categorical, numerical, or both data. As a result, a model should be chosen because it was trained using the same data input format [25].

Size of the Training Data: Collecting a large amount of data is usually recommended to obtain reliable results. However, data availability is frequently a constraint. So, if the training data is smaller or the dataset has fewer observations but a greater number of features, a model capable of segmenting such pancreas images must be chosen. If the training data is large enough and the number of observations exceeds the number of features, a matching model must be chosen for the task [17].

Accuracy and/or Interpretability of the Output: A model's accuracy predicts a response value for a given observation close to its true response value. A highly interpretable model means that any individual predictor can be easily associated with the response, whereas flexible models provide higher accuracy at the expense of low interpretability [26].

4.2 Proposed Model

The authors propose an Intelligent Automated segmentation model, Fig: 1, for the pancreas using the variants of the U-Net model, that is, U-Net, Attention U-Net, and the Residual Attention U-Net models. Each model is trained on the same data to determine the best results. The training information is sent to the Model Analyser (MA) [27] for storage to be used as a reference for model selection when a new dataset is made available. The process of decision-making for the model to use when a segmentation task is presented is done in the Inference Engine, which consists of two parts, the MA and the Intelligent Automation Model (IAM).

Inference Engine: An inference engine interprets and evaluates the facts in the MA to provide the best model for the new dataset. Deep learning model experts will be interviewed and observed to gather facts for an MA. This knowledge is then represented as "if-then" rules known as production rules: "If a condition is met, the following inference can be made, or a model selection action can be taken, or data is sent for training a model [28]." In this paper, the inference engine is divided into two parts, the MA, which represents the knowledge base and the IAM, for decision-making and data interpretation.

Model Analyser (MA): A unified interactive framework for investigating segmentation models and applications. It analyses and evaluates a model's behaviour and limitations and its robustness to changes in the dataset. It seeks to improve segmentation accuracy by matching model behaviour to dataset properties and investigating the factors considered when manually selecting a model when we have new pancreas images to segment. When working with segmentation models in deep learning, exploring and analysing alternative possibilities to a model's decision is necessary. During this exploration and analysis, data scientists will ask questions like the ones below [27], [29]:

- Which model is best for the dataset given different requirements?;
- How will the model behave if the features of the selected data point are changed and data size changes?;
- Did the model performs well, producing accurate results by considering metrics like IOU and dice coefficient?;

The above questions are asked to determine the robustness of a segmentation model. The MA enables data scientists to answer those questions by automatically selecting models based on the new data segmentation constraints and the capabilities of each model. Each model's performance benchmark on different data constraints is saved for future reference.

Intelligent Automation Model (IAM): Automation is now a well-known technical concept in segmentation, with many organisations automating manual, repeatable tasks to improve efficiency. The intelligent aspect is introduced by emerging technologies such as artificial intelligence and deep learning, which can carry out human actions and act independently by making decisions or interpreting data without direct human intervention. So, Intelligent Automation is the combination of smart technologies and automated processes to improve segmentation process automation **khaparde2022differential**.

The IAM is presented with new pancreas images to segment, and it has to decide which model, from the trained U-Net variants models, to use to segment the images and update the MA. This decision-making process is done using information from the MA. The MA provides the knowledge base with information on each model and the segmentation constraints used to train the models. If no model is suitable to segment the new pancreas images, then new training is done using the new dataset, and the MA is updated.

A model selected to do the segmentation task is executed. For it to be successful, it must satisfy the condition that its IoU and accuracy are greater than or equal to that of the previous task the model was trained on. If it is less, the IAM is informed to update the MA. If the IAM decides that no model is suitable for the segmentation task, the New dataset is sent to be used for training the models, and all components must be updated on the results.

Proposed Segmentation Intelligent Automated System (IAS): We proposed an IAS for medical imaging focusing on CT pancreas images for the following reasons:

1. **Increased process efficiency** – speed up segmentation processes while ensuring they are completed to a high standard;
2. **Technology Integration** – rather than executing complete individual segmentation tasks in isolation, integrate them with other technologies to achieve intelligent automation;
3. **Release boredom with repetitive tasks** – To be confident that routine segmentation tasks are being carried out fully, effectively, and error-free with intelligent technologies helping to make informed decisions;

4. **Interpret Big Data** – intelligent automation can handle and interpret large amounts of data that would take humans a long time to manage;
5. **Reduced costs** – Tasks are segmented using best-selected models rather than doing trial and error, wasting resources.

Image pre-processing of the pancreas CT images is done, and the U-Net variants models are trained. Validation and testing are done to have the IoU and accuracy. If below a threshold image, pre-processing and retraining are done again in case the images are not well prepared, as it also affects the results. Retraining is necessary in case the model has not learned enough. Training is also done if no model can do the segmentation task given the new dataset. If a model is selected to segment the new dataset, the inference engine must be updated on the evaluation matrices and model configurations.

Model Selection Models are trained, and the MA is updated on the performance of each model given different data constraints provided in the data description or according to the model's analysis. For the Inference Engine to select the model to use on new data, it will follow the following steps:

1. Train and test data is input for image pre-processing;
2. All three models are trained using the same train and test pre-processed images;
3. The results from the training are stored in the MA, which makes a summary of all the parameters;
4. When new data is available, the IAM analyse the data description and retrieves the performance metrics summary from the MA and analyse if the new data has the same requirements;
   - if parameters match to a threshold of 75% and above, the the model which produced the best segmentation results is selected to work on the new data
   - If parameters match a threshold of less than 75%, then the data goes for pre-processing and training as completely new data.

5 Results and Discussion

This section will discuss some performance metrics the MA will analyse to decide which model best fits the new Data. The models were tested on two different samples set of train and test data [(A(train and test set)] and [B(train and test set)]. Fig: 2 shows a sample of some images from the dataset that were segmented using the three variants of the U-Net model. Attention U-Net segmented the images better than the other variants, followed by Residual Attention U-Net. This could also be seen in Tables: 1, 2 and 3, where the accuracy, IOU, and dice coefficient of Attention U-Net is always higher than that of the other variants. Results from these tables are the factors that will be stored in the MA and used to determine the best model to segment new data based on these factors. From the results of Table: 1, we notice that the U-Net model requires less memory than other variants and has less training time. Compared to other variants, Attention U-Net Table: 2 is good when the data size is large and works well with high data parameters. Residual Attention U-Net, Table: 3, proves to work well with more features as indicated by increased accuracy, IOU, and dice coefficient as the number of features increases.

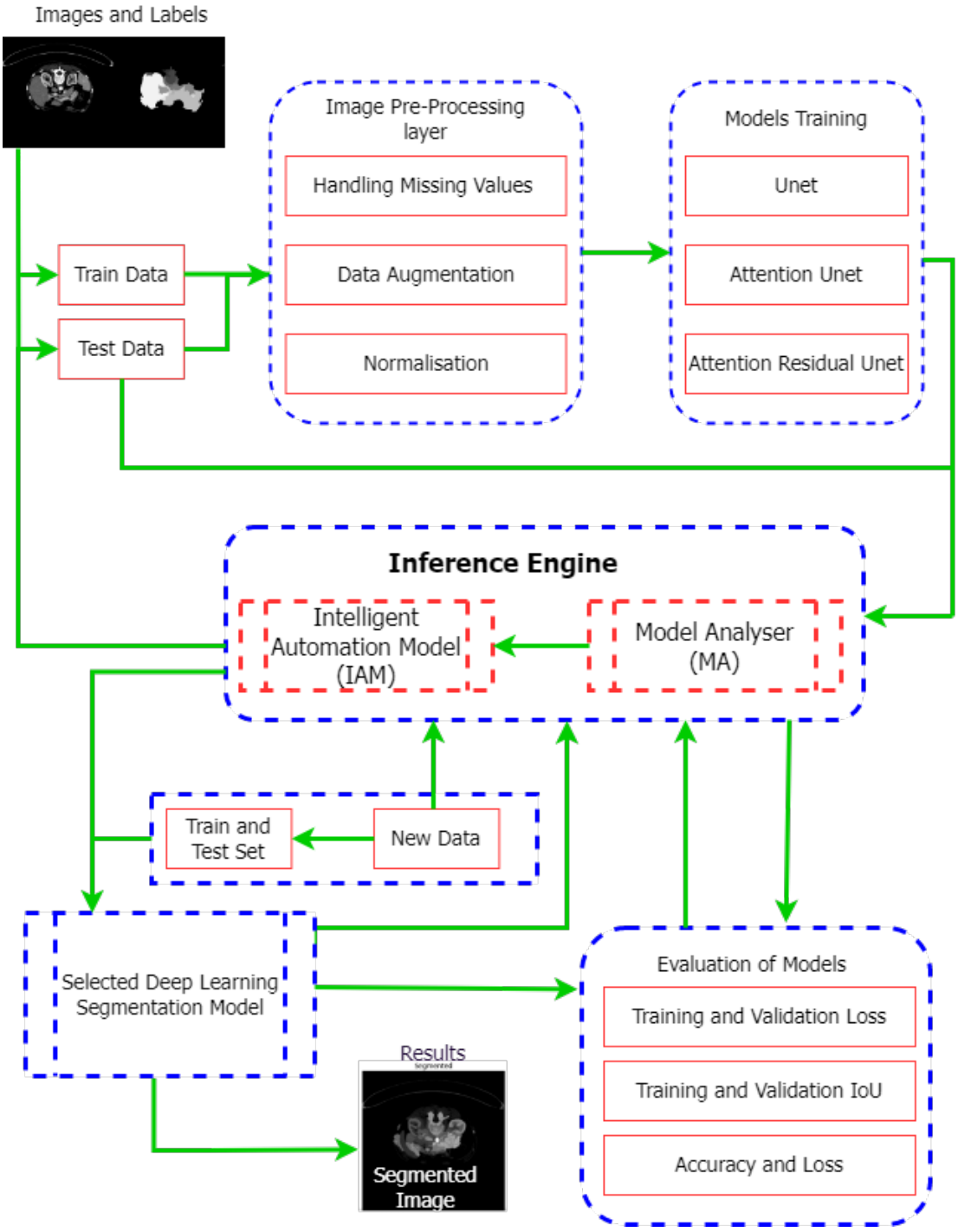


Fig. 1. Proposed Segmentation Intelligent Automated System Architecture

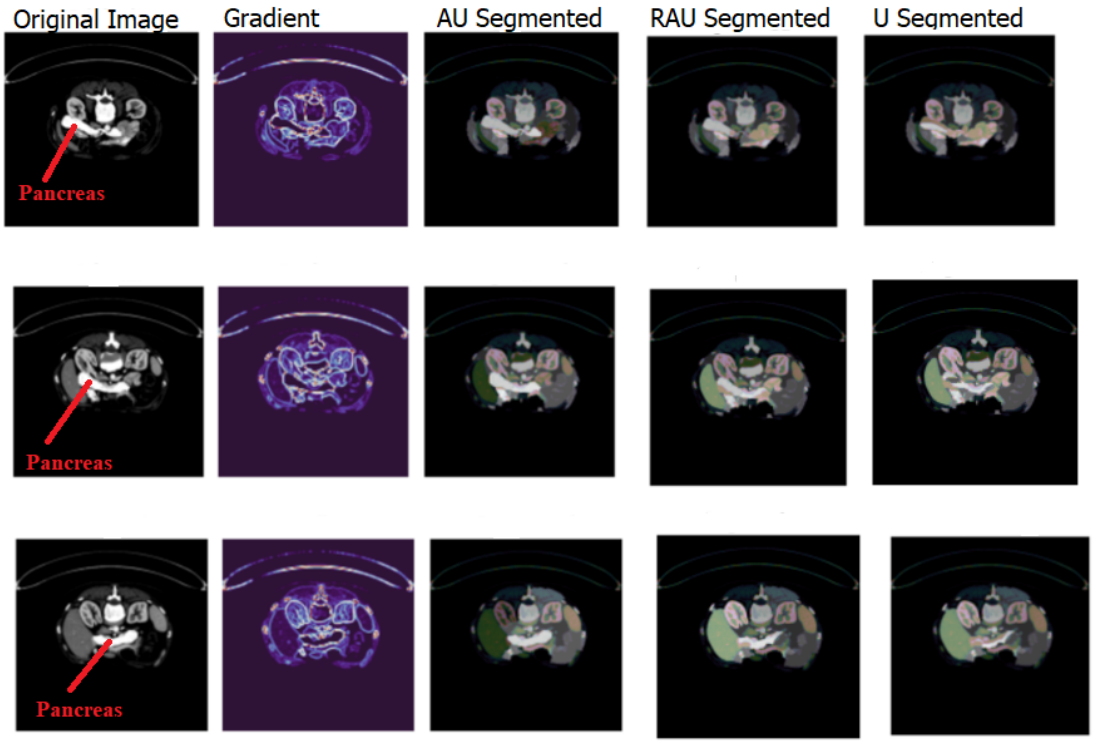


Fig. 2. an illustration of a sample of original images randomly selected from the dataset and segmented using Attention U-Net, Residual Attention U-Net, and the U-Net. The Gradient images show colour intensity changes.

Table 1. U-Net results for different factors tested, which are considered for selecting the best model by the MA. UV(U-net Variants), F(number of Features), DF(Data Format), STD(Size of Training Data - images)), MR(Memory Requirements - Gigs), TT(Training Time - seconds), ACC(Accuracy), P(Parameters - millions), IOU(Intersection Over Union), DC(Dice Coefficient, and TPE(Time Per Epoch - seconds). Requires less memory for training and takes less time as compared to other variants.

UV	F	DF	STD	MR	TT	ACC	P	IOU	DC	TPF
U-Net	8									
		dicom	800	1.2	33	0.89	31.4	0.8421	0.88	26
		nrrd	800	1.0	29	0.97	31.1	0.8033	0.83	26
		nifti	800	0.75	19	0.89	32.0	0.8350	0.83	18
		dicom	1200	2.1	45	0.91	38.3	0.8555	0.89	49
		nrrd	1200	1.3	42	0.87	37.0	0.8461	0.84	37
		nifti	1200	0.87	26	0.87	38.2	0.8489	0.84	25
	12									
		dicom	800	1.24	33	0.90	33.3	0.8622	0.90	33
		nrrd	800	1.2	35	0.88	33	0.8231	0.85	34
		nifti	800	0.75	24	0.89	34.5	0.8350	0.86	23
		dicom	1200	2.2	50	0.90	38.3	0.8232	0.90	50
		nrrd	1200	1.25	42	0.89	37.8	0.8461	0.87	42
		nifti	1200	0.89	24	0.90	36.6	0.8589	0.87	32
	16									
		dicom	800	1.4	34	0.84	40.3	0.7321	0.84	37
		nrrd	800	1.4	33	0.84	37.1	0.7432	0.81	38
		nifti	800	0.80	29	0.82	33.9	0.7733	0.84	26
		dicom	1200	2.3	52	0.87	44.3	0.7813	0.83	50
		nrrd	1200	2.0	40	0.89	48.2	0.7347	0.82	46
		nifti	1200	1.1	26	0.80	38.4	0.7588	0.88	38

Table 2. Attention U-Net results for different factors tested, which are considered for selecting the best model by the MA. UV(U-net Variants), F(number of Features), DF(Data Format), STD(Size of Training Data - images)), MR(Memory Requirements - Gigs), TT(Training Time - seconds), ACC(Accuracy), P(Parameters - millions), IOU(Intersection Over Union), DC(Dice Coefficient, and TPE(Time Per Epoch - seconds). Best when the data size is large, and there are more parameters.

UV	F	DF	STD	MR	TT	ACC	P	IOU	DC	TPF
AU	8									
		dicom	800	2.1	42.3	0.96	31.6	0.8532	0.91	34
		nrrd	800	1.7	37	0.94	31.2	0.8499	0.90	32
		nifti	800	1.7	23	0.94	32.2	0.8343	0.91	23
		dicom	1200	3.2	61	0.98	38.6	0.8675	0.93	56
		nrrd	1200	2.3	48	0.97	37.3	0.8500	0.91	56
		nifti	1200	2.1	38	0.97	38.5	0.8489	0.91	31
	12									
		dicom	800	2.3	48	0.96	31.7	0.8536	0.91	36
		nrrd	800	1.7	37	0.97	31.5	0.8500	0.90	32
		nifti	800	1.9	22	0.94	32.8	0.8366	0.91	21
		dicom	1200	3.1	63	0.98	38.2	0.8657	0.94	53
		nrrd	1200	2.1	43	0.97	37.6	0.8633	0.92	53
		nifti	1200	2.4	37	0.98	38.8	0.8532	0.92	34
	16									
		dicom	800	1.4	35	0.84	32.4	0.8421	0.88	31
		nrrd	800	1.3	31	0.83	33.1	0.8033	0.83	29
		nifti	800	0.79	20	0.83	30.0	0.8350	0.83	24
		dicom	1200	2.2	49	0.90	39.3	0.8555	0.89	48
		nrrd	1200	1.4	46	0.90	37.0	0.8461	0.84	33
		nifti	1200	0.87	28	0.91	36.2	0.8489	0.85	29

Table 3. Residual Attention U-Net results for different factors tested, which are considered for selecting the best model by the MA. UV(U-net Variants), F(number of Features), DF(Data Format), STD(Size of Training Data - images)), MR(Memory Requirements - Gigs), TT(Training Time - seconds), ACC(Accuracy), P(Parameters - millions), IOU(Intersection Over Union), DC(Dice Coefficient, and TPE(Time Per Epoch - seconds). Proves to be efficient with many features and can work well with all data formats.

UV	F	DF	STD	MR	TT	ACC	P	IOU	DC	TPF
RAU	8									
		dicom	800	2.1	42.3	0.93	31.6	0.8421	0.87	28
		nrrd	800	1.6	36	0.92	31.2	0.8033	0.83	28
		nifti	800	1.7	23	0.91	32.2	0.8350	0.80	20
		dicom	1200	3.2	57	0.92	38.6	0.8555	0.88	48
		nrrd	1200	1.9	48	0.91	37.3	0.8461	0.88	41
		nifti	1200	2.1	36	0.90	38.5	0.8489	0.89	27
	12									
		dicom	800	2.1	42.3	0.95	31.6	0.8531	0.87	31
		nrrd	800	1.6	36	0.93	31.2	0.8124	0.83	33
		nifti	800	1.7	23	0.91	32.2	0.8650	0.80	30
		dicom	1200	3.2	57	0.93	38.6	0.8601	0.88	37
		nrrd	1200	1.9	48	0.94	37.3	0.8441	0.88	37
		nifti	1200	2.1	36	0.92	38.5	0.8420	0.89	26
	16									
		dicom	800	3.1	43	0.96	35.6	0.8531	0.89	34
		nrrd	800	2.0	38	0.95	31.9	0.8424	0.85	40
		nifti	800	1.9	22	0.91	36.2	0.8750	0.85	31
		dicom	1200	3.1	54	0.94	37.9	0.8501	0.89	39
		nrrd	1200	2.1	50	0.94	37.3	0.8641	0.87	36
		nifti	1200	2.3	37	0.95	38.4	0.8620	0.91	29

6 Conclusion

When choosing a model for a deep learning task, performance may appear to be the most obvious metric. However, performance alone will not help you choose the best model for the job. Additional criteria include memory requirements, training and prediction time, interpretability, and data format, which the model must meet. Makes a more confident decision by incorporating a broader range of factors. When choosing the best model for new data from a few options is difficult, the MA can test them on a validation dataset. This provides metrics for the MA to compare each model and make a final decision.

References

[1] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[2] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: State of the art," *International journal of multimedia information retrieval*, vol. 9, no. 3, pp. 171–189, 2020.

[3] J. Le'Clerc Arrastia, N. Heilenkötter, D. Otero Baguer, *et al.*, "Deeply supervised unet for semantic segmentation to assist dermatopathological assessment of basal cell carcinoma," *Journal of imaging*, vol. 7, no. 4, p. 71, 2021.

[4] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 36–46.

[5] Z. Li, H. Zhang, Z. Li, and Z. Ren, "Residual-attention unet++: A nested residual-attention u-net for medical image segmentation," *Applied Sciences*, vol. 12, no. 14, p. 7149, 2022.

[6] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative," *Medical image analysis*, vol. 52, pp. 109–118, 2019.

[7] L. Sun, L. Zhang, and D. Zhang, "Multi-atlas based methods in brain mr image segmentation," *Chinese Medical Sciences Journal*, vol. 34, no. 2, pp. 110–119, 2019.

[8] J. Wu and X. Tang, "Brain segmentation based on multi-atlas and diffeomorphism guided 3d fully convolutional network ensembles," *Pattern Recognition*, vol. 115, p. 107 904, 2021.

[9] G. Suman, A. Patra, P. Korfiatis, *et al.*, "Quality gaps in public pancreas imaging datasets: Implications & challenges for ai applications," *Pancreatology*, vol. 21, no. 5, pp. 1001–1008, 2021.

[10] S. F. Qadri, L. Shen, M. Ahmad, S. Qadri, S. S. Zareen, and S. Khan, "Op-convnet: A patch classification-based framework for ct vertebrae segmentation," *IEEE Access*, vol. 9, pp. 158 227–158 240, 2021.

[11] M. K. Islam, M. S. Ali, M. S. Miah, M. M. Rahman, M. S. Alam, and M. A. Hossain, "Brain tumor detection in mr image using superpixels, principal component analysis and template based k-means clustering algorithm," *Machine Learning with Applications*, vol. 5, p. 100 044, 2021.

[12] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, and M. A.-M. Salem, "Liver tumor segmentation in ct scans using modified segnet," *Sensors*, vol. 20, no. 5, p. 1516, 2020.

[13] P. Zhao, J. Zhang, W. Fang, and S. Deng, "Scau-net: Spatial-channel attention u-net for gland segmentation," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 670, 2020.

[14] L. Bragagnolo, R. V. da Silva, and J. M. V. Grzybowski, "Amazon forest cover change mapping based on semantic segmentation by u-nets," *Ecological Informatics*, vol. 62, p. 101 279, 2021.

[15] O. Ozturk, B. Saritürk, and D. Z. Seker, "Comparison of fully convolutional networks (fcn) and u-net for road segmentation from high resolution imageries," *International journal of environment and geoinformatics*, vol. 7, no. 3, pp. 272–279, 2020.

[16] A. Olivier, O. Moal, B. Moal, *et al.*, "Active learning strategy and hybrid training for infarct segmentation on diffusion mri with a u-shaped network," *Journal of Medical Imaging*, vol. 6, no. 4, p. 044 001, 2019.

[17] N. Arunkumar, M. A. Mohammed, S. A. Mostafa, D. A. Ibrahim, J. J. Rodrigues, and V. H. C. de Albuquerque, "Fully automatic model-based segmentation and classification approach for mri brain tumor using artificial neural networks," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, e4962, 2020.

[18] M. Tripathi, "Facial image denoising using autoencoder and unet," *Heritage and Sustainable Development*, vol. 3, no. 2, pp. 89–96, 2021.

[19] B. Wang, S. Qiu, and H. He, "Dual encoding u-net for retinal vessel segmentation," in *International conference on medical image computing and computer-assisted intervention*, Springer, 2019, pp. 84–92.

[20] T. Falk, D. Mai, R. Bensch, *et al.*, "U-net: Deep learning for cell counting, detection, and morphometry," *Nature methods*, vol. 16, no. 1, pp. 67–70, 2019.

[21] E. Thrane and C. Talbot, "An introduction to bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models," *Publications of the Astronomical Society of Australia*, vol. 36, 2019.

[22] F. Isensee, P. F. Jäger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," *arXiv preprint arXiv:1904.08128*, 2019.

[23] R. Hemelings, B. Elen, I. Stalmans, K. Van Keer, P. De Boever, and M. B. Blaschko, "Artery–vein segmentation in fundus images using a fully convolutional network," *Computerized Medical Imaging and Graphics*, vol. 76, p. 101 636, 2019.

[24] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE access*, vol. 7, pp. 60 134–60 149, 2019.

[25] J. C. Caicedo, A. Goodman, K. W. Karhohs, *et al.*, "Nucleus segmentation across imaging experiments: The 2018 data science bowl," *Nature methods*, vol. 16, no. 12, pp. 1247–1253, 2019.

[26] S. Nam, Y. Chong, C. K. Jung, *et al.*, "Introduction to digital pathology and computer-aided pathology," *Journal of pathology and translational medicine*, vol. 54, no. 2, pp. 125–134, 2020.

[27] I. Zakarija, F. Škopljanac-Mačina, and B. Blašković, "Automated simulation and verification of process models discovered by process mining," *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, vol. 61, no. 2, pp. 312–324, 2020.

[28] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. Abo-Elsoud, "Detecting covid-19 patients based on fuzzy inference engine and deep neural network," *Applied soft computing*, vol. 99, p. 106 906, 2021.

[29] S. L. Oh, J. Vicnesh, E. J. Ciaccio, R. Yuvaraj, and U. R. Acharya, "Deep convolutional neural network model for automated diagnosis of schizophrenia using eeg signals," *Applied Sciences*, vol. 9, no. 14, p. 2870, 2019.

4. 5 Conclusion

In conclusion, applying deep learning techniques to pancreas instance segmentation represents a significant advancement in medical image analysis. The ability to accurately and efficiently identify and delineate individual instances of the pancreas within complex medical images holds immense promise for improving diagnostic accuracy, treatment planning, and disease monitoring. The studies and approaches discussed in this context underscore the potential of convolutional neural networks and other deep learning architectures to handle the intricate structures and variations in pancreas images. As this technology continues to evolve, it is foreseeable that the synergy between deep learning algorithms and medical imaging will pave the way for more precise and personalised patient care, ultimately contributing to better outcomes in pancreatic disease management. However, it is crucial to acknowledge the ongoing challenges related to data availability, model generalisation, and clinical integration that must be addressed to translate these techniques into real-world medical practice seamlessly.

Chapter 5

Classification of CT Pancreas Images

5. 1 Introduction

The fusion of technology and medicine has paved the way for remarkable advancements in the ever-evolving healthcare field. Computed Tomography (CT) scans have emerged as a pivotal tool in diagnosing and understanding pancreatic disorders, offering intricate visualisations of this vital organ. Within this realm, the art of classifying CT pancreas images takes centre stage as deep learning and artificial intelligence techniques synergise with medical expertise to unlock new dimensions of accuracy and efficiency in disease identification. In this discussion, we delve into a crucial aspect of radiology – the classification of CT pancreas images. Combing through intricate cross-sectional scans, radiologists and machine learning algorithms are tasked with deciphering the intricate details of pancreas morphology to detect abnormalities and facilitate accurate diagnoses. This study examines the classification process, its significance in early disease detection, and the symbiotic relationship between human expertise and cutting-edge technology in pancreatic health assessment.

5. 2 VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction.

This paper proposes a combined model of XGBoost and VGG16 for classifying CT pancreas images. The learning process benefits the structures, and deep learning approaches are applicable in real-world settings. They have been utilised in medical imaging, and this application may develop quickly in the next few years. Significant therapeutic implications arise from deep learning in medical imaging.

Patient treatment will be improved because of this branch of study. For deep learning approaches to be employed most successfully, it is essential to show their benefits. In the context of medical image analysis, deep learning algorithms can classify, count, and categorise disease patterns. They also make it possible to broaden analytical goals and create models for patients' therapy prognoses. Researchers working in medical imaging are considering these challenges, and deep learning is thriving in this area.

Journal of *Imaging*

MDPI

Article

VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction

Wilson Bakasa † **and Serestina Viriri** *,†

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban 4041, South Africa; 219098448@stu.ukzn.ac.za

* Correspondence: viriris@ukzn.ac.za

† These authors contributed equally to this work.

Abstract: The prognosis of patients with pancreatic ductal adenocarcinoma (PDAC) is greatly improved by an early and accurate diagnosis. Several studies have created automated methods to forecast PDAC development utilising various medical imaging modalities. These papers give a general overview of the classification, segmentation, or grading of many cancer types utilising conventional machine learning techniques and hand-engineered characteristics, including pancreatic cancer. This study uses cutting-edge deep learning techniques to identify PDAC utilising computerised tomography (CT) medical imaging modalities. This work suggests that the hybrid model VGG16–XGBoost (VGG16—backbone feature extractor and Extreme Gradient Boosting—classifier) for PDAC images. According to studies, the proposed hybrid model performs better, obtaining an accuracy of 0.97 and a weighted F1 score of 0.97 for the dataset under study. The experimental validation of the VGG16–XGBoost model uses the Cancer Imaging Archive (TCIA) public access dataset, which has pancreas CT images. The results of this study can be extremely helpful for PDAC diagnosis from computerised tomography (CT) pancreas images, categorising them into five different tumours (T), node (N), and metastases (M) (TNM) staging system class labels, which are T0, T1, T2, T3, and T4.

Keywords: feature extraction; classification; computerised tomography; VGG16; XGBoost

Citation: Bakasa, W.; Viriri, S. VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction. *J. Imaging* **2023**, *9*, 138. https://doi.org/10.3390/jimaging9070138

Academic Editors: Pier Luigi Mazzeo and Yudong Zhang

Received: 15 May 2023
Revised: 19 June 2023
Accepted: 4 July 2023
Published: 7 July 2023

1. Introduction

Machine learning uses enormous datasets that require a lot of processing power to produce predictions and recommendations that are more accurate. The efficient feature extraction (FE) [1] technique allows for resource reduction without sacrificing important data. Machine learning models can be made more effective and accurate by extracting features. The amount of redundant data in the collection is decreased with feature extraction. In the end, the data reduction speeds up the learning and generalisation phases of the machine learning process while also enabling the model to be built with less machine effort [2]. Extracting the features can be helpful when one has a large dataset and needs to conserve resources without sacrificing any crucial or pertinent data. Finally, the data reduction speeds up the learning and generalisation phases of the machine learning process while requiring less machine work to develop the model.

Machine learning is more accurate and effective when features are extracted. By removing redundant and unwanted data, feature extraction cuts through the noise. Only the data necessary to train the model for its intended medical diagnosis use are necessary to build the most accurate machine learning models [3]. The model's accuracy suffers with the inclusion of ancillary data. The learning process is slowed by using training data that are not specifically relevant. A reduction in data means processing jobs that do not bring value do not occupy computational resources.

In image analysis activities, image classification poses a significant problem, particularly regarding the choice of methodologies and strategies for utilising the output of image

processing and pattern recognition, classification methods, and, ultimately, verifying the image classification result against medical expert knowledge [4]. In addition to achieving high accuracy, the primary goal of medical image classification is to pinpoint the specific areas of the pancreas that are infected.

However, extracting the right image features to capture the essential information from a dataset accurately is still difficult. This study suggests using a hybrid VGG16–Extreme Gradient Boost (XGBoost) model to extract and classify pancreatic ductal adenocarcinoma (PDAC) features from computerised tomography (CT) images. There are stages 0 (zero) and I to IV (1 to 4). For clinicians to collaborate and develop the most effective treatment plans, the use of stages offers a common manner of classifying cancer. The "T" plus a letter or number (0 to 4) in the tumour (T), node (N), and metastases (M) (TNM) staging system [5,6] is used to indicate the size and location of the tumour. Centimetres (cm) are used to assess tumour size. A normal pen or pencil's width is about equivalent to one centimetre. The tumour stage aids the physician in creating the most effective treatment strategy for each patient. Details on each tumour stage are listed below:

- T0 (time plus 0): There are no signs of cancer in the pancreas;
- T1: The tumour is solely in the pancreas and is no more than 2 cm in size (according to item T1). T1a, T1b, and T1c are other stages that can be identified based on the size of the tumour;
- T2: The tumour, which is only in the pancreas, is more than 2 cm but less than 4 cm in size;
- T3: The tumour is larger than 4 cm and extends past the pancreas. There is no involvement of the major arteries or veins near the pancreas;
- T4: The tumour spread affects the pancreas and nearby major arteries and veins. A T4 tumour cannot be completely removed during surgery.

Each PDAC tumour must be separately diagnosed because it may affect the prognosis and course of treatment. Accurate identification of every PDAC multi-classification category is necessary for proper diagnosis.

Deep learning [7–9] allows computers to understand images and learn from facts. This intelligence allows computers to extract patterns which are particular to a dataset and use those patterns to support autonomous reasoning. Based on some expectations, digital image processing considerably influences decision-making processes. It provides more accurate feature extraction.

1. XGBoost classifier [10], pre-trained VGG16 model [11,12], and deep learning were combined to help with the early diagnosis of PDAC patients;
2. Nine different combinations, including VGG16–XGBoost, VGG16–RF, and VGG16–SVM, Inception V3–XGBoost, Inception V3–RF, and Inception V3–SVM, as well as LGBM–XGBoost, LGBM–RF, and LGBM–SVM, were tested to identify PDAC in pancreas CT images. In these combinations, XGBoost, RF [13,14], and SVM [15] were employed as classifiers, while Inception V3, VGG16, and LGBM [16] were used as deep feature extractors;
3. To evaluate the effectiveness of the proposed framework in terms of accuracy, precision, recall, F1-score, and confusion matrix, a detailed experimental investigation was performed.

Several deep learning algorithm phases are applied in medical imaging, as illustrated in Figure 1.

Developing algorithms and models to assist radiologists in detecting and diagnosing diseases and applying machine learning techniques to improve image processing and diagnosis are among the difficulties confronting computer-aided diagnosis [17]. Another problem is improving image visibility or diagnostic quality so that discrete structures or regions of interest within an image can be detected and separated. Image registration, or aligning images received in different modalities or at different times to make them more comparable, should also be considered.

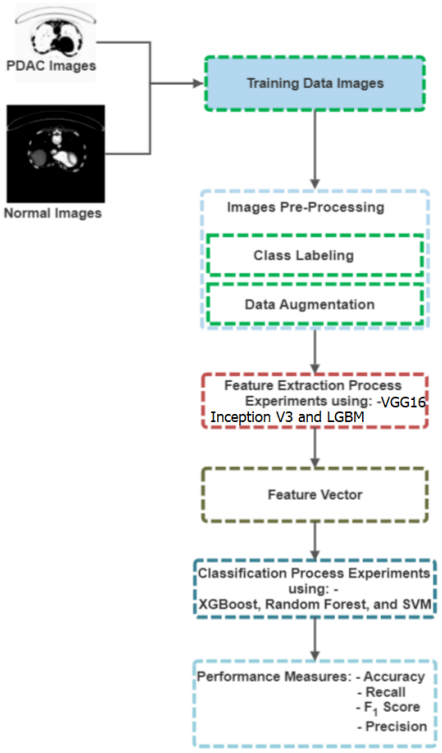


Figure 1. Different steps to follow when using deep learning algorithms in medical image processing.

One of the most significant issues in image recognition is classifying medical images into several groups to aid in diagnosing diseases or further scientific investigation [18]. Two steps can be used to classify medical images in general. Finding useful elements in the image is the first step. In the following step, classification models for the dataset of images are created utilising the features. The classification of medical images into several classes has traditionally been a challenging, tedious, and time-consuming task performed by doctors using their professional experience. This strategy is likely to provide unstable or unpredictable results. Medical image classification application study has had significant worth compared to prior research [19]. Numerous papers in this field have been published due to the scholars' work. However, as of right now, we are still unable to complete our objective efficiently. Doctors could diagnose ailments with additional research if we complete the categorisation task. Consequently, figuring out how to complete this activity effectively is crucial.

The main objectives of this study are to automate the diagnosis and to categorise phases of PDAC from CT images with the same accuracy as that of an expert in medical science. This paper contributes to building a hybrid classification model by taking advantage of different algorithms used in deep learning to address the above problems. VGG16 is used to extract features from images and classify them using XGBoost. To do this, we propose the VGG16–XGBoost model.

The automatic diagnosis of diseases on CT images is primarily improved by a good classification model [20]. This is necessary so that the diagnostic algorithms can adjust to the image groupings produced due to the categorisation. Firstly, this study aims to automatically recognise and classify medical images using deep learning (DL) algorithms—VGG16, Inception V3, and LGBM—to aid in the diagnosis, treatment planning, and monitoring of illness. Secondly, this study experiments and compares three model classifiers: XGBoost, Support Vector Machines (SVM) and Random Forest (RF). Lastly, after experimenting on individual models, the study aims to use a hybrid model to extract and classify PDAC features from CT images [21].

Our literature analysis shows that, before developing deep architectures, many researchers have classified medical images using shallow models, which primarily rely on

shape, colour, texture elements, and their combinations [22]. The biggest issue with these models is that the extracted features are sometimes referred to as low-level features; these features have weak generalisation abilities and cannot represent high-level concepts in the problem domain. On the other hand, the non-medical imaging area has seen a lot of success with deep architecture. Deep learning-based methods, the most amazing branch of machine learning, are an efficient technique to build an end-to-end model that can calculate final classification labels using the raw pixels of medical images.

2. Related Work

Ref. [23] suggested machine learning algorithms that aid in diabetes prediction and found that LGBM performed better than all other models, with an accuracy of 89.85% and an area under the ROC curve (AUC) of 0.95. As a result, LGBM is a more effective algorithm for differentiating between diabetes and non-diabetic individuals.

Using a few selected features, [24] classified colon cancer tissues using SVM. They used 40 colon cancer tumours and 22 healthy colon tissues and obtained an accuracy of 94.51%. The top 50–200 genes were then utilised for training SVMs, which distinguished between non–tumour and tumour specimens as well as or better than the entire repertoire of 1988 genes.

On the patient dataset for COVID-19, [25] offered a model to use the RF approach, which had an F1 Score of 0.866 and was enhanced by the AdaBoost algorithm. On unbalanced datasets, the Boosted Random Forest algorithm also provided detailed forecasts and an accuracy of 94%. The information analysed in this investigation showed that Wuhan natives had higher death rates.

To classify brain tumours from MRI scans, [26] used both VGG16 and AlexNet, with each model capturing 1000 characteristics. The collected features were further assessed using the recursive feature elimination (RFE) selection technique to determine the most effective features. With 200 features selected, the SVM classifier achieved 96.77% accuracy. An SVM for pneumonia identification using X-rays was presented by Eid and Elawady, based on ResNet. The created model employed an SVM classifier to identify pneumonia based on the relevant features after using a boosting technique to choose the features from chest X-rays that were important to the problem [27]. Following training with 5863 X-rays, the model's accuracy was 98.13%.

3. Materials and Methods

The feature extraction process converts unprocessed data into numerical features that machine learning algorithms can use. Data scientists can develop new features appropriate for machine learning applications by extracting the geometry of an object or the redness value from images. A crucial aspect of image processing is feature extraction [28]. This method is used with other devices to identify elements in digital images, such as edges, forms, or movements. Once these have been located, the data can be processed to carry out various functions relating to image analysis.

Utilising feature descriptors such as Scale-Invariant Feature Transform (SIFT) and speeded-up robust features (SURF) as a method for object detection, using well-established Computer Vision (CV) algorithms is the conventional method [29]. For applications such as image classification before the development of DL, a procedure called feature extraction was used. Features are discrete areas of an image that are "interesting", "descriptive", or "informative". This stage might incorporate CV methods, including threshold segmentation, corner detection, and edge detection. A bag-of-words, or definition, of each object class is created using the most features that can be practically derived from an image [30]. During the deployment phase, these definitions are looked up in other images. An image is labelled as containing a particular object if it contains a sizeable number of features from one bag of words in another image.

End-to-end learning is a concept introduced by deep learning (DL) [31]. In this approach, the computer is given a dataset of images tagged with the classes of objects present

in each image. As a result, a DL model is "trained" on the input data, where neural networks identify the underlying patterns in class images and automatically determine which features are the most salient and descriptive for each object in each class [32]. Although there are trade-offs in computing resources and training time, it is widely documented that DNNs outperform conventional techniques. The CV engineer's workflow has significantly changed due to all state-of-the-art approaches in CV using this methodology. Instead of extracting hand-crafted features, which was once their area of expertise, they now iterate through deep learning architectures, as shown in Figure 2.

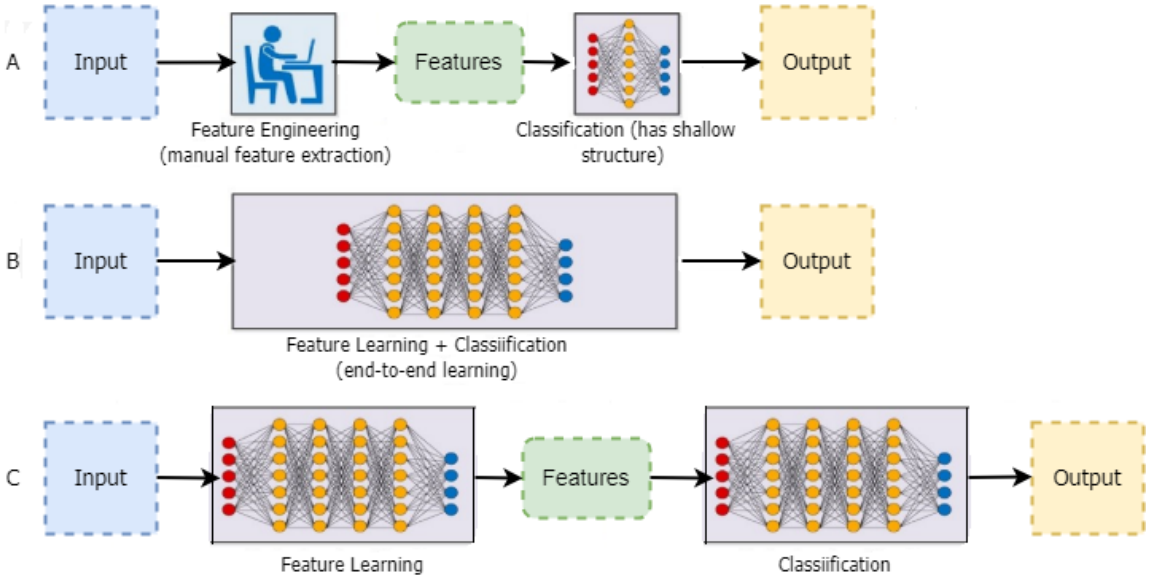


Figure 2. (**A**) The classification process using traditional techniques, (**B**) the classification process using DL techniques, and (**C**) the classification process using the proposed model.

Three algorithms, VGG16, LGBM, and Inception V3, were experimented on to find which one would extract highly accurate features and consider different metrics. Three classifiers, SVM, RF, and XGBoost, were also experimented on to find the one which gave high accuracy. The best feature extraction, VGG16, was combined with the best classifier, XGBoost, to solve the problem and obtain better outcomes. First, experiments compared three deep learning (DL) feature descriptors: VGG16, Inception V3 and LGBM. Secondly, experiments compared three model classifiers: XGBoost, Support Vector Machines (SVM), and Random Forest (RF). Lastly, after experimenting on individual models, the study used a hybrid model to extract and classify PDAC features from CT images provided by the TCIA dataset. The classifier receives the features extracted from the feature-extracting models as input and then passes them to the classifier model.

3.1. CT Images Dataset

The TCIA Public Access dataset was used. In this dataset, 82 abdominal contrast-enhanced 3D CT images were obtained in portal-venous from 53 male and 27 female patients 70 s after intravenous contrast injection. The subjects' ages ranged from 18 to 76, with a mean age of 46.8 $\pm$ 16.7. The CT scans, collected using Philips and Siemens MDCT scanners at 120 kVp tube voltage, have resolutions of 512 $\times$ 512 pixels, with varied pixel sizes and slice thicknesses ranging from 1.5 to 2.5 mm [33].

3.2. Feature Extractors (FE)

To find common features, descriptors are compared between the PDAC CT images. For two images, we might obtain a collection of pairings with the notation $(Xi, Yi) \longleftrightarrow (Xi', Yi')$, where (Xi, Yi) denotes a feature in one image and (Xi', Yi') denotes its corresponding feature in the other. There are two key techniques for FE: human

feature extraction and automatic feature extraction. Machine learning models, such as SVM and decision trees, are employed for manual feature extraction [34]. However, these methods are time-consuming and inefficient for classifying PDAC CT images. This feature extraction may be carried out automatically using the deep learning technique, which is why deep learning has become so popular recently. For a given outcome, different trained and untrained models are available. In this study, three DL models, VGG16, LGBM and Inception V3, were experimented on to find the one with high accuracy when implemented as FE for the given data. Traditional FE techniques include the following:

- Statistic Pixel-Level Features (SPLF) [35]: The pixels within a segmented region can be quantitatively described by these properties. The SPL features are as follows: mean, variance, a histogram of the grey values of the pixels in the region as well as the region's area, details on the contrast of the pixels inside the region, and edge gradients of the pixels defining the region's boundaries;
- Feature Shape Circularity: compactness, moments, chain codes, and the Hough transform are among the characteristics that reveal details about the shape of the region boundary [36]. To describe shapes, morphological processing techniques have also been employed;
- Texture characteristics [37]: determined using the second-order statistical histogram or co-occurrence matrices, these provide information on the local texture within the region or related area of the image. Additionally, wavelet processing describes local texture information in spatial frequency analysis;
- Relational characteristics [37]: these reveal the relational and hierarchical organisation of the regions connected to a single object or a collection of objects.

This conventional method presents a challenge in that it necessitates selecting which elements in each image are crucial. Feature extraction becomes more difficult as there are more classes to categorise [38]. The CV engineer must determine which features best define various types of objects through judgement and a protracted process of trial and error. Additionally, the CV engineer must fine-tune many parameters for each feature definition.

Traditional approaches primarily rely on shape, colour, and/or texture features and their combinations; most are problem-specific and have been demonstrated to be complementary in medical images [39]. This results in a system that cannot represent high-level problem domain concepts and has poor model generalisation ability. A comprehensive model that can generate final classification labels from the raw pixels of medical images can be built using recent DL techniques.

Deep learning rejects the conventional programming paradigm, substituting issue analysis for a training framework in which the system is fed many training patterns, which are collections of inputs for which the intended outputs are known, which it learns and utilises to compute new patterns. DL is utilised in digital image processing to handle challenging issues (such as image colourisation, classification, segmentation, and detection). With massive data and plenty of processing power, DL techniques such as Convolutional Neural Networks (CNNs) have pushed the envelope of what is feasible, primarily increasing prediction performance. Superhuman accuracy is used to tackle issues previously thought to be unsolvable. This is best illustrated via image classification. In recent years, the CV field has seen significant change, largely due to the invention of CNNs, which has significantly improved object recognition [40]. A rise in computer power and the amount of data accessible for training neural networks have contributed to this recent surge in productivity.

The performance and cost-effectiveness of vision-based applications have been enhanced, further accelerating their adoption. This is due to the rapid advancements in DL and device capabilities, such as CPU power, memory capacity, power consumption, and optics [27]. Thanks to DL approaches, CV engineers may now more accurately complete tasks such as object detection, Simultaneous Localisation and Mapping (SLAM), semantic segmentation, image classification, and more. Since DL neural networks are trained rather than coded, applications utilising this method frequently require less specialised analysis

and fine-tuning and take advantage of the vast amounts of video data already available in systems [27]. Furthermore, DL is more adaptable than CV algorithms, which are typically more domain-specific, because CNN models and frameworks can be retrained using a customised dataset for any use case.

To identify characteristics (such as edges) throughout an image, CNNs use kernels, commonly referred to as filters. A kernel is a matrix of weights trained to identify particular features. The primary principle of CNNs, Figure 3, is to spatially convolve the kernel on an input image and check for the presence of the feature it is supposed to detect, as the name suggests [41]. By computing the dot product of the kernel and the input area where the kernel is overlapped (the region of the original image the kernel is looking at is known as the receptive field), a convolution operation is performed to provide a value representing how confident it is that a specific feature is present.

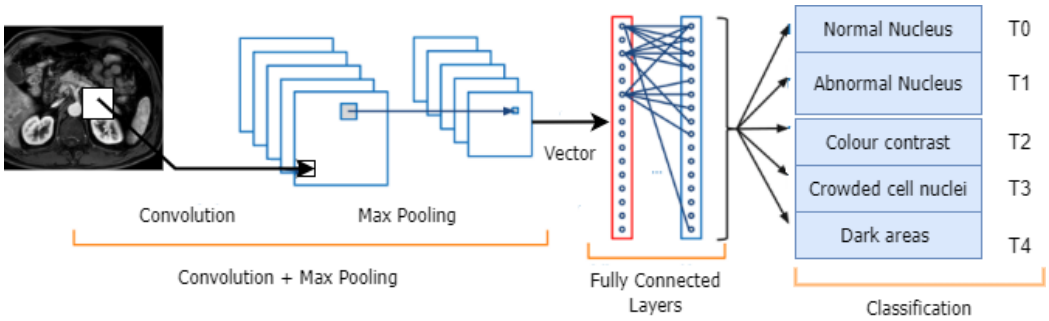


Figure 3. Convolutional Neural Networks' architecture blocks.

The convolutional layer is frequently followed by a pooling layer to eliminate redundancy in the input feature, which speeds up the training process and lowers the amount of memory the network needs [42]. For instance, max pooling outputs the maximum value from a window moved over the input, decreasing the number of pixels output to just the crucial ones in an image. Figure 3 illustrates how deep CNNs may contain many pairs of convolutional and pooling layers. The output layer then computes the scores (confidence or probability) for the output classes/features using a dense network after a Fully Connected layer flattens the volume of the previous layer into a feature vector. After that, this output is fed into a regression function, e.g., Softmax, which converts everything into a vector whose elements add up to one.

The output of the convolution layer is added to a bias term and sent to a nonlinear activation function to speed up the learning of kernel weights [43]. Most of the time, activation functions are non-linear, including Sigmoid, TanH, and ReLU (Rectified Linear Unit). These activation functions are chosen following the characteristics of data and classification tasks. Since neurons in the brain either fire or do not, ReLUs are known to have a greater biological representation [44]. As a result, they provide sparser, more effective representations and produce positive outcomes for image identification tasks, since they are less sensitive to the vanishing gradient problem.

3.2.1. LGBM Feature Extractor

LGBM is a gradient-boosting framework built on decision trees to enhance model performance while utilising less memory. The Gradient Boosting Decision Tree (GBDT) [45] framework overcomes the limitations of the histogram-based approach, which is widely used in all GBDT frameworks, by using Gradient-based One Side Sampling and Exclusive Feature Bundling (EFB) [46,47].

LGBM [48] divides the tree leaf-wise instead of growing trees level-by-level as other boosting algorithms do. The leaf chosen for growth is the one with the highest delta loss. Since the leaf is fixed, the leaf-wise algorithm has a lower loss than the level-wise algorithm. The leaf-wise tree growth could cause the model's complexity to increase, which

could potentially lead to overfitting in small samples. The diagrammatic representation of leaf-wise tree growth is shown below in Figure 4.

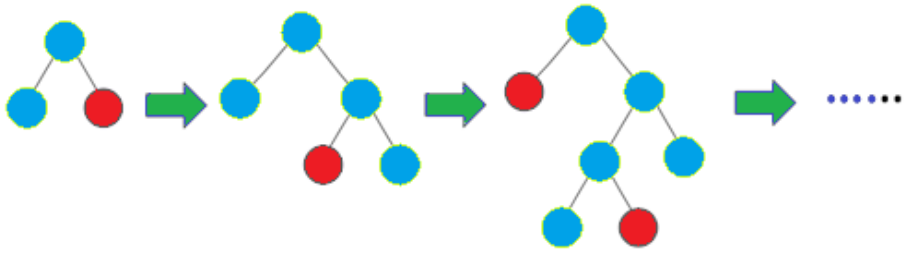


Figure 4. Leaf-wise tree growth of LGBM.

3.2.2. VGG16 Feature Extractor

VGG16 (Figure 5) is a highly deep CNN for large-scale image identification [49]. The model attains 92.7% top-5 test accuracy.

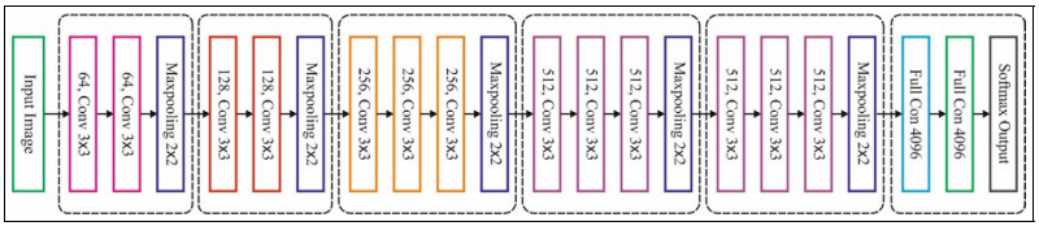


Figure 5. VGG16 architecture.

Convolution Layer

It contains a filter that convolves over the input volume's width and height. The output of the convolution layer is produced by performing a dot product operation between each area of the image and the filter weight content [50].

Pooling Layer

When drawing a pooling layer, to decrease dimensionality and complexity, speed up processing, and avoid overfitting, it uses non-linear down-sampling on the activation map [51].

Max Pooling

Similar to the convolutional layer, this layer has the same fundamental properties, i.e., using the maximum neighbouring value from each point in the input image is crucial. It is carried out on every individual input channel [50].

Softmax Layer

Using the Softmax layer, the output layer of CNN is utilised to represent the categorical distribution over labels and delivers the probability of the inputs to the labels [52].

3.2.3. Inception V3

Google suggested the GoogLeNet network, which is a CNN, in 2014 [53]. Inception v1 (2014), Inception v2 (2015), Inception v3 (2015) in Figure 6, Inception v4 (2016), and Inception-ResNet (2016) are the five core versions of GoogLeNet. The fundamental component of Inception v3 is a Keras-created network structure that has already been trained on Image Net. Unlike the Inception v1 and v2 network structures, the Inception v3 network structure uses a convolution kernel splitting technique [54] to separate large volume integrals into smaller convolutions.

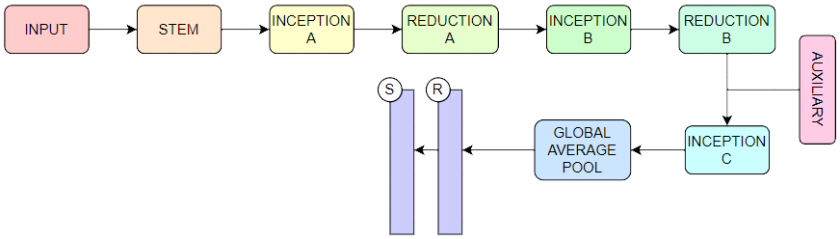


Figure 6. Inception V3 architecture.

3.3. Classifiers

A classifier [55] is an algorithm used in machine learning that automatically sorts data into one or more categories. In the case of this study, the classifier scans the pancreas CT images and filters them according to the class label.

3.3.1. SVM Classifier

Vladimir Vapnik and his colleagues initially proposed SVM. This method of classifying aims to divide the categories along a surface that maximise the margin. The data points near the decision surface are called "support vectors". The chosen boundary is referred to as a "hyperplane". Using training data with a known class label, SVM [56] creates the model. Support vectors, which are essential training data elements, are created during the SVM learning stage. The model and support vectors eventually categorise the test data. SVM-based categorisation for linear data is modelled mathematically using optimisation problems [57]. The non-linear SVM idea is implemented using kernels. For the classification of non-linear data, SVM transforms the basic training data into a higher dimension via non-linear mapping. In the upper dimension, it searches for the linear separating hyperplane.

3.3.2. RF Classifier

To construct a random forest, which joins a group of "weak learners" to form a "strong learner", the Bagging approach and random variable selection were used in this study. In this technique, each decision tree in the group is constructed using a sample with replacement from the training data. Each decision tree in the group acts as a base estimator to estimate the class label of an unlabeled instance. A majority vote completes the process. For each tree in RF, Bootstrap sampling is applied, and data are separated according to the presented problem type. The Gini index is used to divide the data for classification, and each tree model can be trained to minimise the sum of squares of mean deviations for regression. The utilisation of ensembles of trees without pruning is one of the advantages of employing RFs [58].

Additionally, RF is highly resistant to overfitting and does not require standardisation or normalisation because it is unaffected by the value range. The number of trees and the features randomly sampled at each split are two factors that should be changed for the RF model. Various decision tree architectures comprise the RF [59] classifier. The individual samples from the trees connected to this classifier are chosen randomly.

3.3.3. XGBoost Classifier

Training is sped up and overfitting is decreased using XGBoost [60] randomisation techniques such as random subsamples and column subsampling. XGBoost decreases the cost of computing the ideal split by storing the data in a compressed, pre-sorted column-based system. This column-based storage structure enables a parallel search for the optimal division of each taken-into-consideration attribute. Additionally, rather than scanning all potential candidate splits, the XGBoost uses a data percentiles-based strategy to assess a smaller subset of candidate splits and calculate their gain using combined statistics. Therefore, the node-level data subsampling has come close to realising this

concept. XGBoost uses a sparsity-aware approach to exclude null values from the split candidate's loss gain calculation. The XGBoost architecture is demonstrated in Figure 7.

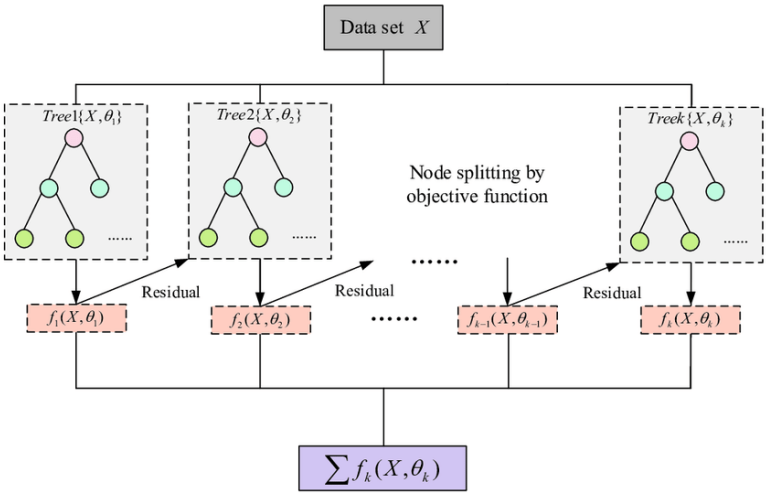


Figure 7. XGBoost architecture.

3.4. Proposed Model

PDAC arises from macroscopic cystic or microscopic precursor lesions [61]. Through the accumulation of genetic and epigenetic alterations, the normal ductal epithelium can progress to a series of precursor lesions, invasive carcinoma, and, eventually, metastatic carcinoma, as shown in Figure 8.

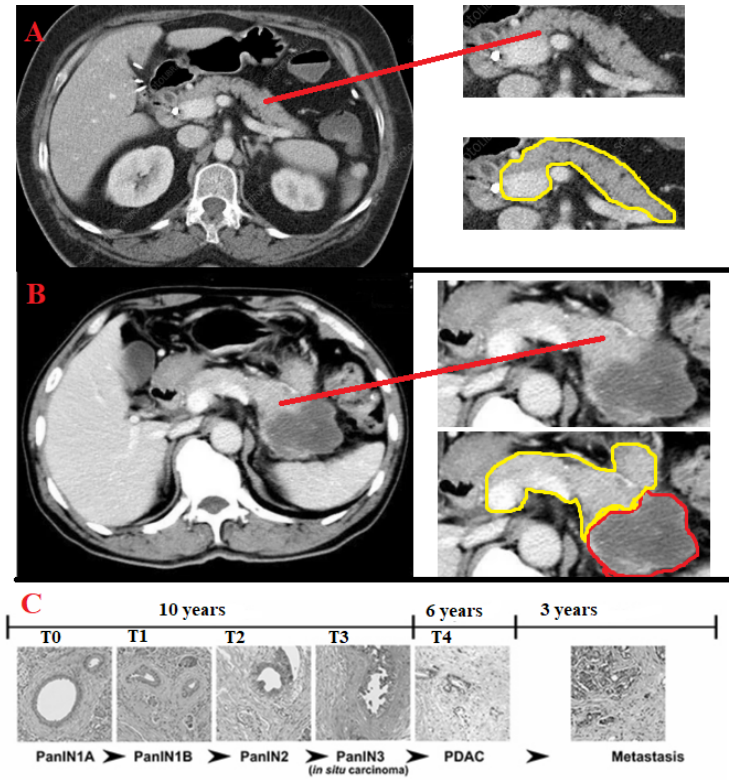


Figure 8. (**A**) A normal pancreas CT image; (**B**) PDAC at an advanced stage, with spreading to other organs; (**C**) Stages of PDAC development, which are represented as T0 to T4. The other stage, metastasis,

is not represented because cancer at that point has spread to other organs surrounding the pancreas. PanIN (pancreatic intraepithelial neoplasia) with low-grade pancreatic intraepithelial neoplasia encompasses three older terms—PanIN-1A, PanIN-1B, and PanIN-2. These include Pancreatic Intraepithelial Neoplasia 1-A, Pancreatic Intraepithelial Neoplasia 1-B, and Pancreatic Intraepithelial Neoplasia 2. Pancreatic Intraepithelial Neoplasia 3 is high-grade pancreatic intraepithelial neoplasia, which then advances to PDAC.

This study suggests a hybrid VGG16–XGBoost model (Figure 9) to sort pancreas CT images into T0, T1, T2, T3, and T4 classes; for this task, we use an XGBoost classifier. A technique for machine learning that performs classification and regression modelling is called XGBoost. This system consists of a set of gradient-boosted decision trees. New models are created using the gradient boosting technique to anticipate the errors or residuals of prior models, which are then integrated to obtain the final prediction. To reduce errors, boosting algorithms employ gradient descent techniques.

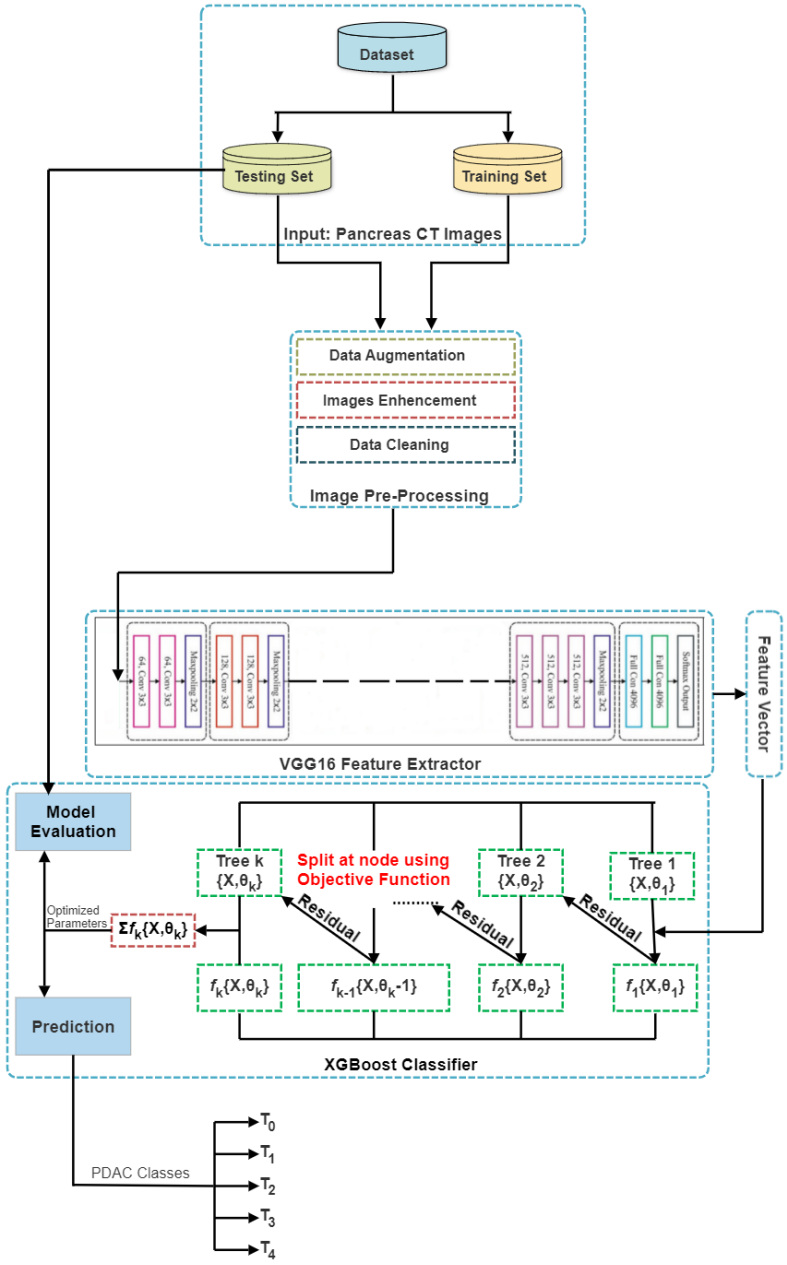


Figure 9. Proposed model, illustrating VGG16 as the feature extractor and XGBoost as the classifier.

4. Experiments

Data were split, and the training set was composed of 80% of the data, while the test set was 20%, i.e., 229 images used for training and 49 images used for testing. The classification job was trained for RF, SVM, and XGBoost. The proposed multi-feature combination model exhibited the greatest performance when the following performance metrics were utilised to assess the proposed system: accuracy (ACC), precision (Pre), recall (Rec), and F1-score (F1). The XGBoost classifier outperformed the RF and SVM classification algorithms regarding performance (Algorithms 1 and 2).

We also adjusted XGBoost hyperparameters to test which settings produced better results. The following parameters were then used: model_colsample_bytree = 0.9, model_gamma = 0.0, model_learning_rate = 0.01, model_max_depth = 5, model_min_child_weight = 1, model_n_estimators = 250, model_reg_lambda = 1, and model_subsample = 0.5

Keras, an open-source Python interface for Convolutional Neural Networks, was used as the library. Keras is a TensorFlow library interface. One Hot Encoding was performed to encode categorical features for machine learning models. This type of encoding creates a new binary feature for each viable category and assigns a value of 1 to each sample's feature that corresponds to its original category. The Jupyter Notebook is a tremendously effective tool for generating and presenting deep learning projects interactively.

Ductal adenocarcinomas typically present as poorly defined masses surrounded by the broad desmoplastic response. They enhance weakly compared to normal pancreatic tissue and seem hypoattenuating on arterial phase scans in 75–90% cases. However, they may become iso-attenuating on delayed scans, necessitating multiple-phase scanning when pancreatic cancer is suspected. The double duct indicator is visible. Calcifications in adenocarcinoma are exceedingly rare and are more likely to result from a pre-existing condition, such as chronic pancreatitis [62].

Algorithm 1: Model algorithm: Implementing LGBM to Classifiers

Input: Input: train_images.
Input: Input: test_labels.
/* Read training and testing data. */
Read input images and assign labels based on folder names.
Resize images.
Capture images and labels into arrays. /* Start by creating empty lists. Do the same for test/validation images. */
Encode labels. /* from the text (folder names) to integers. */
Separate the data into train and test datasets. /* Already split but assigned to a meaningful convention. Split the dataset here if there is just one. */
Normalise pixel values to between 0 and 1.
Feature extractor function. /* Input shape is (n, x, y, c) - number of images, x, y, and channels */
Create a temporary data frame. /* To capture information for each loop. Reset the data frame to blank after each loop. */
Start adding data to the data frame.
Add features from the current image to the item in the dataset.
Extract features from training images.
Reshape to a vector. /* For Random Forest/SVM training */
Define the classifier.
Model Train.
/* Using the training data. */
Send test data using the identical feature extractor technique.
Make prediction. /* Use the learned RF model. */
Extract features. /* Also reshape them to the right dimensions */
Predict.
return *Classification_Report*

Regarding predicting unresectability, CT correlates strongly with surgical results, with a positive predictive value of 89–100%. The most significant aspect of evaluating locally is the tumour's connection to nearby arteries, including the superior mesenteric artery and the celiac axis. T4 illness is unresectable if a tumour covers a vessel by more than 180 degrees [63].

Algorithm 2: Model algorithm: Implementing VGG16 to Classifiers

Input: Input: train_images.
Input: Input: test_labels.
/* Read training and testing data. */
Read input images and assign labels. /* Based on folder names */
Resize images.
Capture images and labels into arrays. /* Start by creating empty lists. Do the same for test/validation images. */
Encode labels. /* From the text (folder names) to integers. */
Separate the data into train and test datasets. /* Already split but assigned to a meaningful convention. Split the dataset here if there is just one. */
Normalise pixel values to between 0 and 1.
One hot encode y values. /* For the neural network. */
Load model. /* Without classifier/fully connected layers */
Make loaded layers non-trainable. /* It is important to work with pre-trained weights. */
/* Now, let us use features from the convolutional network */
Define the classifier.
Train the model. /* Using the training data. */
Send test data. /* Using the identical feature extractor technique. */
Make a prediction. /* Use the learned RF model. */
Extract features and reshape them to the right dimensions.
Predict.
return *Classification_Report*

5. Experimental Results

The XGBoost classifier outperformed the RF and SVM classification algorithms regarding performance. As a result, the ACC's correct recognition rate when XGBoost was utilised as a classifier was 97%. These findings suggest that the suggested strategy might aid in a more precise diagnosis of instances that are challenging to categorise in pancreas CT imaging.

Performance Assessment Transfer learning was used to assess the effectiveness of the pre-trained VGG-16 model using the precision, recall, and F1 score metrics. In the equations, Equations (1)–(5), the words True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP) are employed. A confusion matrix is a table-based illustration of the effectiveness of the prediction model. Each entry in a confusion matrix indicates the number of predictions the model made when correctly or incorrectly categorising the classes (Figure 10).

Utilising a confusion matrix as the deep learning model's evaluation criterion is usually preferable. It provides very straightforward yet effective performance metrics for the model. The confusion matrix's most widely used performance metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Misclassification\ Rate = \frac{FP + FN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\text{-}score = \frac{2TP}{2TP + FP + FN} \tag{5}$$

In an ideal case, we would like a model with a precision of 1 and a recall of 1, which amounts to an F1 score of 1 or 100% accuracy, which is typically not the case for a machine learning model. We must attempt to increase precision while maintaining a greater recall value. A confusion matrix can assess the model's recall, precision, and accuracy performance.

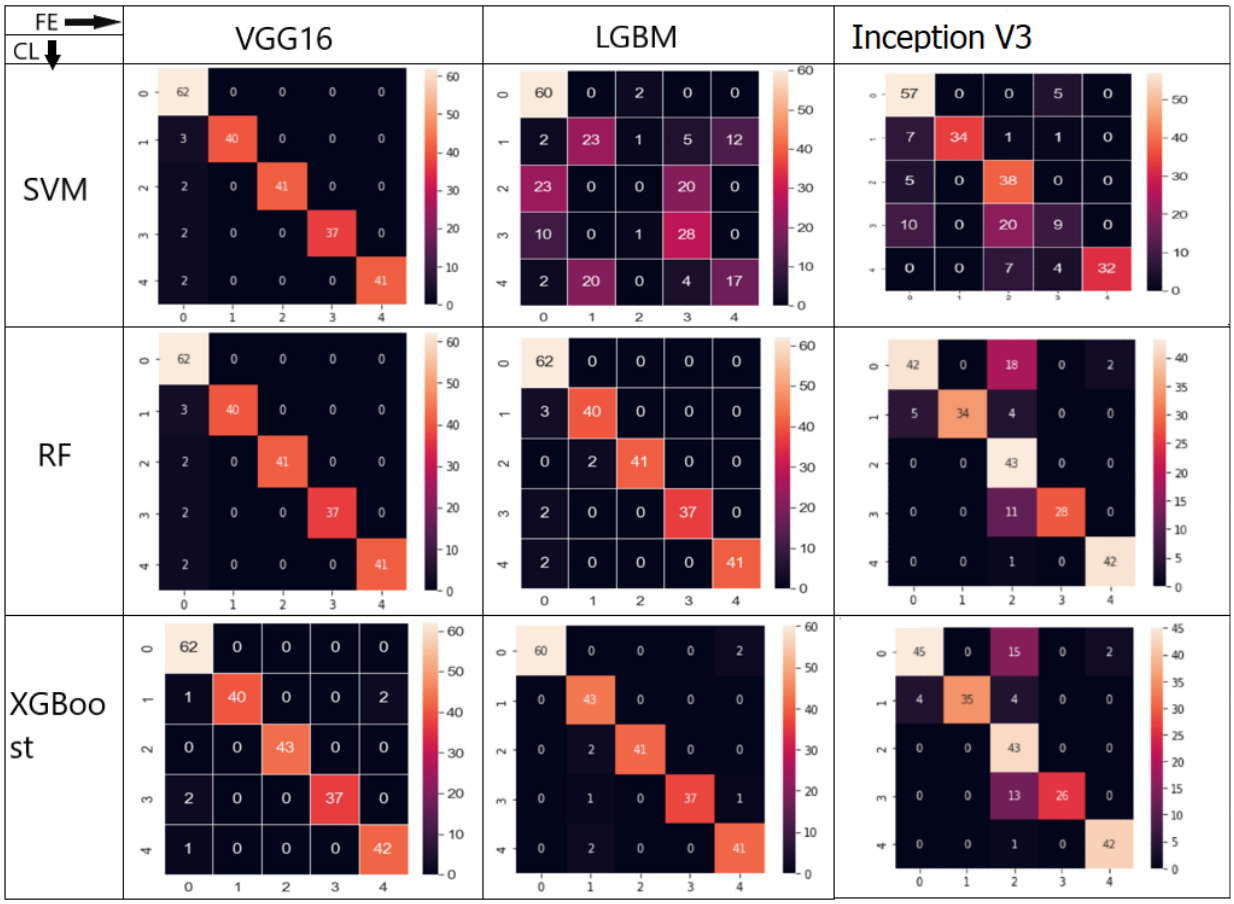


Figure 10. Confusion Matrix for using VGG16–XGBoost, LGBM–XGBoost, Inception V3–XGBoost, VGG16–RF, LGBM–RF, Inception V3–RF, VGG16–SVM, LGBM–SVM, and Inception V3–SVM combinations. The different classification (CL) models with the used feature extraction (FE) models are shown.

Tables 1–3 show the performance measures for using LGBM, VGG16, and Inception V3 as the backbone implemented for feature extraction to XGBoost, RF, and SVM as the classifiers. XGBoost using VGG16 had the highest ACC (Accuracy), WP (Weighted Precision), and WFS (Weighted F1-Score) as performance measures.

Table 1. Evaluation measures results for the model using LGBM (Light Gradient Boosting Machine) as a feature extractor and experiments carried out on three classifiers (CF), that is, SVM (Support Vector Machines), RF (Random Forest), and XGBoost (Extreme Gradient Boosting). The metrics used are AV (Averages), that is, MAV (Micro Average) and WAV (Weighted Average), ACC (Accuracy), WP (Weighted Precision), WFS (Weighted F1 Score), PR (Precision), RE (Recall), and FS (F10Score).

CF	ACC	WP	WFS	CL	AV	PR	RE	FS
SVM	0.56	0.46	0.49					
					MAV	0.45	0.52	0.47
					WAV	0.46	0.56	0.49
				T0		0.62	0.97	0.75
				T1		0.53	0.53	0.53
				T2		0.57	0.60	0.57
				T3		0.49	0.72	0.58
				T4		0.59	0.40	0.47
RF	0.92	0.91	0.93					
					MAV	0.97	0.96	0.96
					WAV	0.96	0.96	0.96
				T0		0.90	1.00	0.95
				T1		0.95	0.93	0.94
				T2		0.97	0.95	0.98
				T3		0.97	0.95	0.97
				T4		0.97	0.95	0.98
XGBoost	0.94	0.93	0.93					
					MAV	0.97	0.96	0.96
					WAV	0.96	0.96	0.96
				T0		0.90	1.00	0.95
				T1		0.95	0.93	0.94
				T2		1.0	0.95	0.98
				T3		0.95	0.95	0.97
				T4		0.94	0.95	0.97

Table 2. Evaluation measures results for the model using VGG16 (a Convolutional Neural Network of 16 convolutional layers by Visual Geometry Group) as a feature extractor and experiments carried out on three classifiers (CF), that is, SVM (Support Vector Machines), RF (Random Forest), and XGBoost (Extreme Gradient Boosting). The metrics used are AV (Averages), that is, MAV (Micro Average) and WAV (Weighted Average), ACC (Accuracy), WP (Weighted Precision), WFS (Weighted F1 Score), PR (Precision), RE (Recall), and FS (F1-Score).

CF	ACC	WP	WFS	CL	AV	PR	RE	FS
SVM	0.95	0.96	0.94					
					MAV	0.97	0.96	0.96
					WAV	0.97	0.96	0.96
				T0		0.87	1.00	0.93
				T1		1.00	0.93	0.96
				T2		1.00	0.95	0.98
				T3		1.00	0.95	0.97
				T4		1.00	0.95	0.96
RF	0.96	0.95	0.95					
					MAV	0.97	0.96	0.96
					WAV	0.97	0.96	0.96
				T0		0.87	1.00	0.93
				T1		1.00	0.93	0.96
				T2		1.00	0.95	0.98
				T3		1.00	0.95	0.97
				T4		1.00	0.95	0.98

Table 2. *Cont.*

CF	ACC	WP	WFS	CL	AV	PR	RE	FS
XGBoost	0.98	0.98	0.97					
					MAV	0.98	0.97	0.97
					WAV	0.98	0.97	0.97
				T0		0.94	1.00	0.97
				T1		1.00	0.94	0.96
				T2		1.0	1.00	1.00
				T3		1.00	0.95	0.98
				T4		0.96	0.98	0.99

Table 3. Evaluation measures results for the model using Inception V3 as a feature extractor and experiments carried out on three classifiers (CF), that is, SVM (Support Vector Machines), RF (Random Forest), and XGBoost (Extreme Gradient Boosting). The metrics used are AV (Averages), that is, MAV (Micro Average) and WAV (Weighted Average), ACC (Accuracy), WP (Weighted Precision), WFS (Weighted F1 Score), PR (Precision), RE (Recall), and FS (F1-Score).

CF	ACC	WP	WFS	CL	AV	PR	RE	FS
SVM	0.63	0.60	0.61					
					MAV	0.57	0.62	0.58
					WAV	0.55	0.60	0.58
				T0		0.68	0.94	0.62
				T1		0.58	0.62	0.60
				T2		0.59	0.59	0.67
				T3		0.60	0.69	0.61
				T4		0.68	0.68	0.51
RF	0.86	0.87	0.89					
					MAV	0.89	0.90	0.92
					WAV	0.87	0.91	0.90
				T0		0.90	0.90	0.92
				T1		0.91	0.89	0.89
				T2		0.89	0.89	0.89
				T3		0.89	0.88	0.91
				T4		0.87	0.91	0.91
XGBoost	0.95	0.95	0.95					
					MAV	0.92	0.93	0.94
					WAV	0.91	0.91	0.93
				T0		0.90	0.96	0.93
				T1		0.91	0.95	0.95
				T2		0.92	0.95	0.95
				T3		0.93	0.95	0.94
				T4		0.93	0.94	0.95

5.1. ROC

A graph depicting the effectiveness of a classification model at all classification thresholds is called a ROC curve (receiver operating characteristic curve) (Figure 11). The True Positive and False Positive rates are plotted on this curve. The chart demonstrates the connection between the True and False Positive rates. It was decided to compare each class to every other using the One-vs.-Rest methodology. The interpretation of the AUC score os as follows:

1. $\leq$0.5 = There is no prejudice;
2. 0.5–0.7 = Inadequate discrimination;
3. 0.7–0.8 = Acceptable discrimination;
4. 0.8–0.9 = Good discrimination;
5. 0.9 = Excellent discrimination.

Although all the models produce acceptable AUC scores, XGBoost had the highest score, followed by Random Forest, and then SVM. The higher the AUC, the better the model predicts 0 classes as 0 and 1 classes as 1. By analogy, the higher the AUC, the better the model distinguishes between images with and without PDAC.

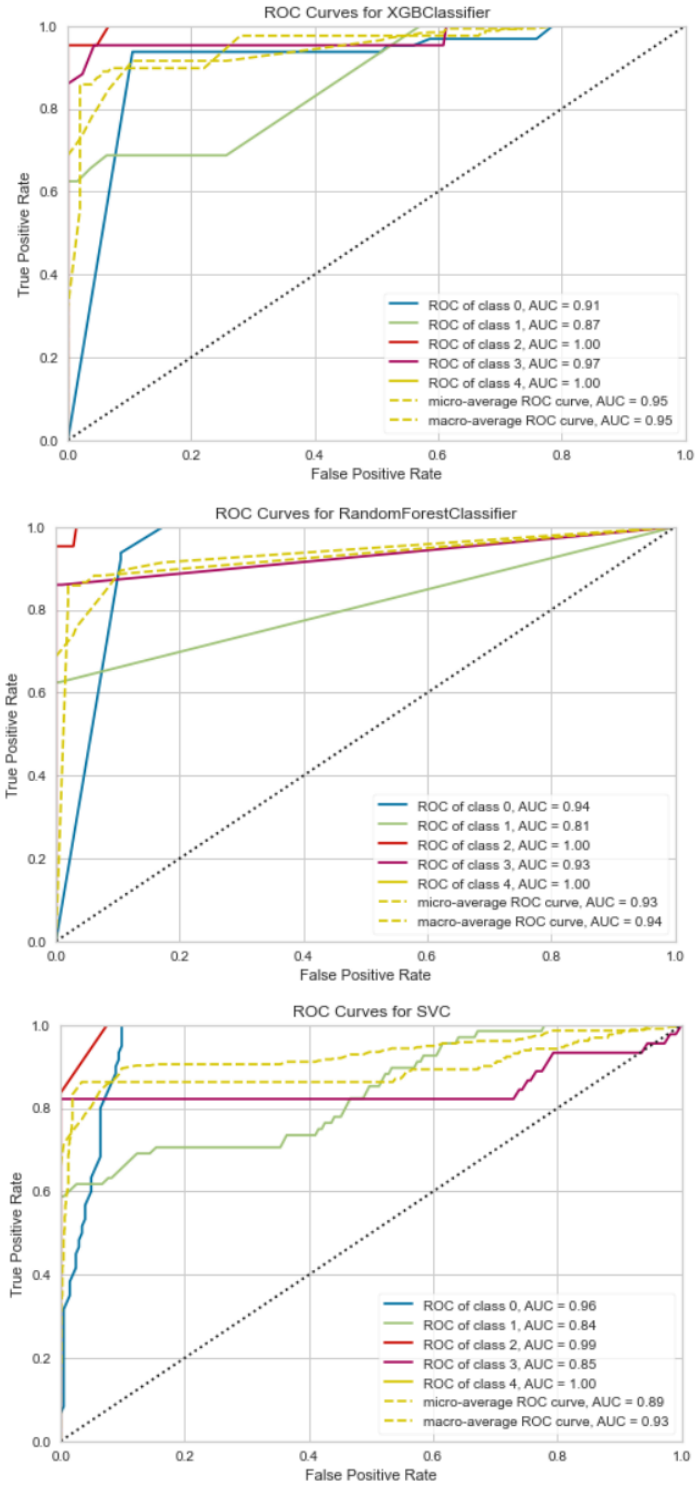


Figure 11. ROC curves for SVM, RF, and XGBoost, showing True Positive against False Positive rates.

The model helps to diagnose the stages of progression for PDAC tumour cells. Surgery for stages T0 and T1 offers a potential cure but comes with a significant morbidity of 20–30% and a fatality rate of 5%. A Whipple procedure is used to remove pancreatic head

tumours [63]. Even when resection is possible, most patients die from recurrence, barely doubling survival in operated patients after five years, from 5% to 10%. Almost a fifth of individuals will have died 12 months after their diagnosis. This makes it necessary to have models of high accuracy to diagnose PDAC, and the VGG16–XGBoost hybrid model with a classification accuracy of 0.97 will be a good tool in medical imaging [64].

6. Discussion

As shown in Figure 10, combining VGG16 and XGBoost results in a lower misclassification rate than other models, indicating that the model can accurately identify place-infected regions. Our approach can accurately identify those positive locations, even if the input patch only has a minor PDAC tumour presence. These findings suggest that our model recognised PDAC mostly through nuclear characteristics. Our system identified tumour cell nuclei as large, irregular, crowded, and dark regions. For tumour identification, these characteristics are essential. However, there are also False Positive and False Negative locations, including those overpopulated with nuclei.

Additionally, we found that the classification model tended to identify packed and asymmetrical groupings of aberrant cell nuclei, wherein the model may be making judgements based on colour contrast. The distribution of cell nuclei is asymmetrical in a few False Positive locations.

Low accuracy is obtained when using LGBM as the feature extractor. The SVM classifier had 0.56 accuracy, while RF and XGBoost had 0.92 and 0.94, respectively. Using VGG16 as the feature extractor resulted in all the classifiers producing the highest accuracy, compared to Inception V3 and SVM. For Inception V3, the classifiers had an accuracy of 0.63 (SVM), 0.86 (RF), and 0.95 (XGBoost), while for VGG16, the classifiers' accuracy was 0.95 (SVM), 0.96 (RF), and 0.98 (XGBoost). Combining VGG16 and XGBoost models resulted in better accuracy than single end-to-end models, as shown in Table 4.

Table 4. Comparison table for accuracy obtained when applying different models in diagnosis.

Paper	Application	Model	Accuracy
[23]	Diabetes Prediction	K-Nearest Neighbour, SVM, Random Forest, LGBM, AdaBoost, and Decision Tree	89.85%
[24]	Identify Colon Cancer	SVM	94.51%
[25]	Identify COVID-19	Random Forest + AdaBoost	94%
[65]	Pancreatic Cystic Lesion	CNN of VGG16	93.11%
[66]	Pancreatic Cancer	Twin SVM	98%
[67]	Skin Cancer		
		CNN	83.2%
		ResNet50	83.7%
		Inception V3	85.8%
		Inception ResNet	84%

Shape characteristics, colour histogram features, colour moment features, and texture features are the most often utilised features in the classification of medical images, according to [67]. Most of the earlier studies use global characteristics to categorise medical images. The most often utilised attributes for recognising targets are texture and colour moment, which are both global features. In [68], an SVM classifier produced SE = 93 and SP = 92 after they retrieved colour, texture, and form variables. To perform well on a dataset of 217 melanomas and 588 photos, [69,70] used colour moment characteristics and texture features.

The Twin SVM model had good results of 0.98, contributing to implementing combined models to form a hybrid. To address the CV issues mentioned above, classic machine learning classification algorithms such as Support Vector Machines and K-Nearest Neighbours are typically supplemented with feature descriptors such as Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) [71]. These days, classical techniques are utilised when the problem can be simplified so that it can be implemented on low-cost

microcontrollers or to restrict the problem for deep learning techniques by emphasising specific elements in the data, enhancing the data, or helping with dataset annotation.

To classify breast cancer into carcinoma and noncarcinoma using histopathology images, [72] presented an ensemble DL technique. In this instance, the framework was created using the VGG16 and VGG19 models. Three further convolutional layers are added to the same fundamental design of VGG16 to create VGG19. The models were adjusted during training, except for the first block, to improve performance. An overall accuracy 95.29% was obtained after the combined tuned VGG16 and VGG19 models.

Algorithms for machine learning have their limits, and it might not be easy to create a model with a high degree of accuracy. The accuracy of 0.98, which is a significant improvement over past research, shows that, if we create and combine many models, we have the potential to increase the overall accuracy. Thus, greater resilience is a fundamental benefit of using the VGG16–XGBoost model. For a specific dataset, one method might not be able to predict the outcome perfectly. By dividing the FE and classification tasks to be completed by separate models, the model reduces the burden on one model once more. As a result, the task division does not put demand on one algorithm's resources.

7. Conclusions

Deep learning techniques are viable in real-world situations, and the structures benefit from the learning process. They have been used in relation to medical imaging and this application might develop quickly in the upcoming years. Deep learning in medical imaging has significant implications for the treatment. Significantly, this field of study guarantees that patients will receive better care. The advantages of deep learning approaches are crucial to demonstrating that they are used most effectively. Deep learning algorithms can categorise, classify, and count illness patterns from image processing in medical image analysis. Additionally, they enable the expansion of analytical objectives and produce therapy prediction models for patients. These difficulties are being taken into account by researchers in medical imaging, and deep learning is blossoming in this field.

Similar to deep learning in many other applications other than health care, it is advancing quickly. A fast-developing area of clinical image analysis called "radiomics" has great potential to help with cancer diagnosis and treatment decision-making. To implement modern computer vision and machine learning techniques effectively, this goal of more effective decision-making requires significant individual and combined expertise from domain specialists in medicine, biology, and computer science. If the raw data are made available, deep learning and machine learning can considerably progress the field of radiomics in the upcoming years, regardless of the patient population or type of tumour.

One of the limitations of this research is that, due to limited datasets for PDAC, we were limited to testing the model on one dataset. This does not limit the implementation of the model to another dataset. For future work, we will experiment on other datasets and try to use the ensemble method for feature extraction so that some models may detect hidden features.

To catch up to other imaging fields, deep models' applications in the medical image analysis domain need a lot of work to be put in because their huge dataset requirements for excellent feature extraction make them inefficient for small datasets. However, medical datasets are frequently tiny because medical images are frequently challenging. As a result, if we employ a deep model to address a limited dataset directly, the method will likely result in model overfitting. Aside from these issues, the model's interpretability is fairly poor, and training a deep model often involves significant computing.

Author Contributions: Conceptualization, W.B. and S.V.; methodology, W.B.; software, W.B.; validation, S.V; formal analysis, W.B. and S.V.; investigation, W.B.; resources, S.V.; data curation, W.B.; writing—original draft preparation, W.B.; writing—review and editing, S.V.; visualization, W.B.; supervision, S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: A publicly available benchmarked dataset was used in this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ghojogh, B.; Samad, M.N.; Mashhadi, S.A.; Kapoor, T.; Ali, W.; Karray, F.; Crowley, M. Feature selection and feature extraction in pattern analysis: A literature review. *arXiv* **2019**, arXiv:1905.02845.
2. Waring, J.; Lindvall, C.; Umeton, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* **2020**, *104*, 101822. [CrossRef] [PubMed]
3. Maghdid, H.S.; Asaad, A.T.; Ghafoor, K.Z.; Sadiq, A.S.; Mirjalili, S.; Khan, M.K. Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. In *Multimodal Image Exploitation and Learning 2021*; SPIE: Bellingham, WA, USA, 2021; Volume 11734, pp. 99–110
4. Gupta, A.; Gupta, K.; Gupta, K.; Gupta, K. A survey on human activity recognition and classification. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 28–30 July 2020; pp. 915–919.
5. Rosen, R.D.; Sapra, A. TNM classification. In *StatPearls [Internet]*; StatPearls Publishing: Tapa, FL, USA, 2021.
6. Rosen, R.D.; Sapra, A. TNM classification. In *StatPearls [Internet]*; StatPearls Publishing: Tapa, FL, USA, 2022.
7. Kelleher, J.D. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2019.
8. Anaya-Isaza, A.; Mera-Jiménez, L.; Zequera-Diaz, M. An overview of deep learning in medical imaging. *Inform. Med. Unlocked* **2021**, *26*, 100723. [CrossRef]
9. Chan, H.P.; Samala, R.K.; Hadjiiski, L.M.; Zhou, C. Deep learning in medical image analysis. *Adv. Exp. Med. Biol.* **2020**, *1213*, 3–21.
10. Paleczek, A.; Grochala, D.; Rydosz, A. Artificial breath classification using XGBoost algorithm for diabetes detection. *Sensors* **2021**, *21*, 4187. [CrossRef]
11. Singh, D.; Taspinar, Y.S.; Kursun, R.; Cinar, I.; Koklu, M.; Ozkan, I.A.; Lee, H.N. Classification and analysis of pistachio species with pre-trained deep learning models. *Electronics* **2022**, *11*, 981. [CrossRef]
12. Albashish, D.; Al-Sayyed, R.; Abdullah, A.; Ryalat, M.H.; Almansour, N.A. Deep CNN model based on VGG16 for breast cancer classification. In Proceedings of the 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 14–15 July 2021; pp. 805–810.
13. Chowdhury, A.R.; Chatterjee, T.; Banerjee, S. A Random Forest classifier-based approach in the detection of abnormalities in the retina. *Med. Biol. Eng. Comput.* **2019**, *57*, 193–203. [CrossRef]
14. Guo, Z.; Li, X.; Huang, H.; Guo, N.; Li, Q. Deep learning-based image segmentation on multimodal medical imaging. *IEEE Trans. Radiat. Plasma Med. Sci.* **2019**, *3*, 162–169. [CrossRef]
15. Yadav, S.S.; Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **2019**, *6*, 113. [CrossRef]
16. Ono, Y.; Mitani, Y. Evaluation of feature extraction methods with ensemble learning for breast cancer classification. In Proceedings of the 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech), Osaka, Japan, 7–9 March 2022; pp. 194–195.
17. Battineni, G.; Sagaro, G.G.; Chinatalapudi, N.; Amenta, F. Applications of machine learning predictive models in the chronic disease diagnosis. *J. Pers. Med.* **2020**, *10*, 21. [CrossRef]
18. Arabahmadi, M.; Farahbakhsh, R.; Rezazadeh, J. Deep learning for smart Healthcare—A survey on brain tumor detection from medical imaging. *Sensors* **2022**, *22*, 1960. [CrossRef]
19. Liu, X.; Song, L.; Liu, S.; Zhang, Y. A review of deep-learning-based medical image segmentation methods. *Sustainability* **2021**, *13*, 1224. [CrossRef]
20. Ahsan, M.; Based, M.A.; Haider, J.; Kowalski, M. COVID-19 detection from chest X-ray images using feature fusion and deep learning. *Sensors* **2021**, *21*, 1480.
21. Masud, M.; Sikder, N.; Nahid, A.A.; Bairagi, A.K.; AlZain, M.A. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. *Sensors* **2021**, *21*, 748. [CrossRef]
22. Nguyen, D.T.; Lee, M.B.; Pham, T.D.; Batchuluun, G.; Arsalan, M.; Park, K.R. Enhanced image-based endoscopic pathological site classification using an ensemble of deep learning models. *Sensors* **2020**, *20*, 5982. [CrossRef] [PubMed]
23. Malini, M.; Gopalakrishnan, B.; Dhivya, K.; Naveena, S. Diabetic Patient Prediction using Machine Learning Algorithm. In Proceedings of the 2021 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 9–10 October 2021; pp. 1–5.
24. Huang, S.; Cai, N.; Pacheco, P.P.; Narrandes, S.; Wang, Y.; Xu, W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genom. Proteom.* **2018**, *15*, 41–51.
25. Iwendi, C.; Bashir, A.K.; Peshkar, A.; Sujatha, R.; Chatterjee, J.M.; Pasupuleti, S.; Mishra, R.; Pillai, S.; Jo, O. COVID-19 patient health prediction using boosted random forest algorithm. *Front. Public Health* **2020**, *8*, 357. [CrossRef]

26. Toğaçar, M.; Ergen, B.; Cömert, Z. BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Med. Hypotheses* **2020**, *134*, 109531. [CrossRef]
27. Ghimire, D.; Kil, D.; Kim, S.-h. A survey on efficient convolutional neural networks and hardware acceleration. *Electronics* **2022**, *11*, 945. [CrossRef]
28. Florkowski, M. Anomaly Detection, Trend Evolution, and Feature Extraction in Partial Discharge Patterns. *Energies* **2021**, *14*, 3886. [CrossRef]
29. Amerikanos, P.; Maglogiannis, I. Image Analysis in Digital Pathology Utilizing Machine Learning and Deep Neural Networks. *J. Pers. Med.* **2022**, *12*, 1444. [CrossRef] [PubMed]
30. Polap, D.; Wlodarczyk-Sielicka, M. Classification of non-conventional ships using a neural bag-of-words mechanism. *Sensors* **2020**, *20*, 1608. [CrossRef] [PubMed]
31. El-Rashidy, N.; El-Sappagh, S.; Islam, S.R.; El-Bakry, H.M.; Abdelrazek, S. End-to-end deep learning framework for coronavirus (COVID-19) detection and monitoring. *Electronics* **2020**, *9*, 1439. [CrossRef]
32. Naranjo-Torres, J.; Mora, M.; Hernández-García, R.; Barrientos, R.J.; Fredes, C.; Valenzuela, A. A review of convolutional neural network applied to fruit image processing. *Appl. Sci.* **2020**, *10*, 3443. [CrossRef]
33. Roth, F.; Turkbey, L.; Liu, S. The Cancer Imaging Archive. Data From Pancreas-CT (Version 2) [Data Set]. 2016 Available online: https://www.cancerimagingarchive.net/ (accessed on 20 May 2023).
34. Chang, C.C.; Li, Y.Z.; Wu, H.C.; Tseng, M.H. Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques. *Diagnostics* **2022**, *12*, 1747. [CrossRef] [PubMed]
35. Augustauskas, R.; Lipnickas, A. Improved pixel-level pavement-defect segmentation using a deep autoencoder. *Sensors* **2020**, *20*, 2557. [CrossRef]
36. Damian, F.A.; Moldovanu, S.; Dey, N.; Ashour, A.S.; Moraru, L. Feature selection of non-dermoscopic skin lesion images for nevus and melanoma classification. *Computation* **2020**, *8*, 41. [CrossRef]
37. Wang, Y.; Qi, Q.; Shen, X. Image segmentation of brain mri based on ltridp and superpixels of improved slic. *Brain Sci.* **2020**, *10*, 116. [CrossRef]
38. Alzubaidi, L.; Fadhel, M.A.; Al-Shamma, O.; Zhang, J.; Duan, Y. Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis. *Electronics* **2020**, *9*, 427. [CrossRef]
39. Tamang, L.D.; Kim, B.W. Deep learning approaches to colorectal cancer diagnosis: A review. *Appl. Sci.* **2021**, *11*, 10982. [CrossRef]
40. Bhatt, D.; Patel, C.; Talsania, H.; Patel, J.; Vaghela, R.; Pandya, S.; Modi, K.; Ghayvat, H. CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics* **2021**, *10*, 2470. [CrossRef]
41. Ali, M.; Ali, R. Multi-input dual-stream capsule network for improved lung and colon cancer classification. *Diagnostics* **2021**, *11*, 1485. [CrossRef] [PubMed]
42. Capra, M.; Bussolino, B.; Marchisio, A.; Shafique, M.; Masera, G.; Martina, M. An updated survey of efficient hardware architectures for accelerating deep convolutional neural networks. *Future Internet* **2020**, *12*, 113. [CrossRef]
43. Wang, Y.; Li, Y.; Song, Y.; Rong, X. The influence of the activation function in a convolution neural network model of facial expression recognition. *Appl. Sci.* **2020**, *10*, 1897. [CrossRef]
44. Solé, R.; Seoane, L.F. Evolution of brains and computers: The roads not taken. *Entropy* **2022**, *24*, 665. [CrossRef]
45. Rao, H.; Shi, X.; Rodrigue, A.K.; Feng, J.; Xia, Y.; Elhoseny, M.; Yuan, X.; Gu, L. Feature selection based on artificial bee colony and gradient boosting decision tree. *Appl. Soft Comput.* **2019**, *74*, 634–642. [CrossRef]
46. Zhang, D.; Gong, Y. The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. *IEEE Access* **2020**, *8*, 220990–221003. [CrossRef]
47. Fang, M.; Chen, Y.; Xue, R.; Wang, H.; Chakraborty, N.; Su, T.; Dai, Y. A hybrid machine learning approach for hypertension risk prediction. *Neural Comput. Appl.* **2021**, *35*, 14487–14497.
48. Chen, R.C.; Caraka, R.E.; Arnita, N.E.G.; Pomalingo, S.; Rachman, A.; Toharudin, T.; Tai, S.K.; Pardamean, B. An end to end of scalable tree boosting system. *Sylwan* **2020**, *165*, 1–11.
49. Xu, G.; Shen, X.; Chen, S.; Zong, Y.; Zhang, C.; Yue, H.; Liu, M.; Chen, F.; Che, W. A deep transfer convolutional neural network framework for EEG signal classification. *IEEE Access* **2019**, *7*, 112767–112776. [CrossRef]
50. Theckedath, D.; Sedamkar, R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **2020**, *1*, 1–7. [CrossRef]
51. Sajja, V.R.; Kalluri, H.K. Classification of Brain tumors using Fuzzy C-means and VGG16. *Turk. J. Comput. Math. Educ.* **2021**, *12*, 2103–2113.
52. Pardede, J.; Sitohang, B.; Akbar, S.; Khodra, M.L. Implementation of transfer learning using VGG16 on fruit ripeness detection. *Int. J. Intell. Syst. Appl.* **2021**, *13*, 52–61. [CrossRef]
53. Nahata, H.; Singh, S.P. Deep learning solutions for skin cancer detection and diagnosis. In *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*; Springer: Cham, Switzerland, 2020; pp. 159–182.
54. Wang, C.; Chen, D.; Hao, L.; Liu, X.; Zeng, Y.; Chen, J.; Zhang, G. Pulmonary image classification based on inception-v3 transfer learning model. *IEEE Access* **2019**, *7*, 146533–146541. [CrossRef]
55. Giger, M.L. Machine learning in medical imaging. *J. Am. Coll. Radiol.* **2018**, *15*, 512–520. [CrossRef]
56. Scholkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2018.

57. Hsu, C.W.; Chang, C.C.; Lin, C.J. A Practical Guide to Support Vector Classification. 2003. Available online: https://www.csie.ntu.edu.tw/~cjlin/ (accessed on 20 May 2023).
58. Asselman, A.; Nasseh, A.; Aammou, S. Revealing strengths, weaknesses and prospects of intelligent collaborative e-learning systems. *Adv. Sci. Technol. Eng. Syst. J.* **2018**, *3*, 67–79. [CrossRef]
59. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 157–175.
60. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA , 27–30 June 2016; pp. 770–778.
61. Ren, B.; Liu, X.; Suriawinata, A.A. Pancreatic ductal adenocarcinoma and its precursor lesions: Histopathology, cytopathology, and molecular pathology. *Am. J. Pathol.* **2019**, *189*, 9–21. [CrossRef] [PubMed]
62. Javadi, S.; Menias, C.O.; Korivi, B.R.; Shaaban, A.M.; Patnana, M.; Alhalabi, K.; Elsayes, K.M. Pancreatic calcifications and calcified pancreatic masses: Pattern recognition approach on CT. *Am. J. Roentgenol.* **2017**, *209*, 77–87. [CrossRef]
63. Lu, D.; Reber, H.A.; Krasny, R.M.; Kadell, B.M.; Sayre, J. Local staging of pancreatic cancer: Criteria for unresectability of major vessels as revealed by pancreatic-phase, thin-section helical CT. *AJR Am. J. Roentgenol.* **1997**, *168*, 1439–1443. [CrossRef] [PubMed]
64. DeVita, V.T., Jr.; Rosenberg, S.A.; Lawrence, T.S. *DeVita, Hellman, and Rosenberg's Cancer: Short Title*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2022.
65. Lee, T.C.; Angelina, C.L.; Kongkam, P.; Wang, H.P.; Rerknimitr, R.; Han, M.L.; Chang, H.T. Deep-Learning-Enabled Computer-Aided Diagnosis in the Classification of Pancreatic Cystic Lesions on Confocal Laser Endomicroscopy. *Diagnostics* **2023**, *13*, 1289. [CrossRef] [PubMed]
66. Sadewo, W.; Rustam, Z.; Hamidah, H.; Chusmarsyah, A.R. Pancreatic cancer early detection using twin support vector machine based on kernel. *Symmetry* **2020**, *12*, 667. [CrossRef]
67. Gouda, W.; Sama, N.U.; Al-Waakid, G.; Humayun, M.; Jhanjhi, N.Z. Detection of skin cancer based on skin lesion images using deep learning. *Healthcare* **2022**, *10*, 1183. [CrossRef] [PubMed]
68. Karim, A.M.; Kaya, H.; Güzel, M.S.; Tolun, M.R.; Çelebi, F.V.; Mishra, A. A novel framework using deep auto-encoders based linear model for data classification. *Sensors* **2020**, *20*, 6378. [CrossRef]
69. Caroprese, L.; Vocaturo, E.; Zumpano, E. Features for melanoma lesions: Extraction and classification. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence-Companion Volume, Thessaloniki, Greece, 14–17 October 2019; pp. 238–243.
70. Almeida, M.A.; Santos, I.A. Classification models for skin tumor detection using texture analysis in medical images. *J. Imaging* **2020**, *6*, 51. [CrossRef]
71. Yan, Y.; Yao, X.J.; Wang, S.H.; Zhang, Y.D. A survey of computer-aided tumor diagnosis based on convolutional neural network. *Biology* **2021**, *10*, 1084. [CrossRef] [PubMed]
72. Hameed, Z.; Zahia, S.; Garcia-Zapirain, B.; Javier Aguirre, J.; Maria Vanegas, A. Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors* **2020**, *20*, 4373. [CrossRef] [PubMed]

5. 3 Stacked Ensemble Deep Learning for Pancreas Cancer Classification using Extreme Gradient Boosting

Stacked ensemble deep learning is a technique that combines multiple individual models to improve the overall performance of a machine learning system. In the context of pancreatic cancer classification, it can be used to enhance the accuracy and robustness of the classification task. Extreme Gradient Boosting (XGBoost) is a popular gradient-boosting framework with excellent performance in various machine learning tasks. It is known for its efficiency, scalability, and ability to handle large datasets. It's worth noting that the success of stacked ensemble deep learning for pancreatic cancer classification depends on various factors such as data quality, feature selection, model architecture, and hyperparameter tuning.

Stacked Ensemble Deep Learning for Pancreas Cancer Classification using Extreme Gradient Boosting

Serestina Viriri[1*], Wilson Bakasa[1]

[1]University of KwaZulu-Natal, South Africa

Submitted to Journal:
Frontiers in Artificial Intelligence

Specialty Section:
Machine Learning and Artificial Intelligence

Article type:
Original Research Article

Manuscript ID:
1232640

Received on:
31 May 2023

Revised on:
03 Jul 2023

Journal website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

All authors contributed equally

Keywords

Stacking ensemble, deep learning, XGBoost, Hyperparameters, Classification, Pancreas segmentation

Abstract

Word count: 232

Ensemble learning aims to improve prediction performance by combining several models or forecasts. However, how much and which ensemble learning techniques are useful in deep learning-based pipelines for pancreas computed tomography (CT) image classification is a challenge. Ensemble approaches are the most advanced solution to many machine learning problems. These techniques entail training multiple models and combining their predictions to improve the predictive performance of a single model. This article introduces the idea of Stacked Ensemble Deep Learning (SEDL), a pipeline for classifying pancreas CT medical images. The weak learners are Inception V3, VGG16, and ResNet34, and we employed a stacking ensemble. By combining the first-level predictions, an input train set for XGBoost, the ensemble model at the second level of prediction, is created. The final classification will be made by Extreme Gradient Boosting (XGBoost), who is employed as a strong learner. Our findings showed that SEDL performed better, with a 98.8% ensemble accuracy, after some adjustments to the hyperparameters. The Cancer Imaging Archive (TCIA) public access dataset consists of 80 pancreas CT scans with a resolution of 512 * 512 pixels. We concluded that implementing the SEDL technique is an effective way to strengthen the robustness and increase the performance of the pipeline for classifying pancreas CT medical pictures. Interestingly, grouping like-minded or talented learners does not make a difference. Instead, combining several diverse, weak, heterogeneous machine learning methods enhances generalisation.

Contribution to the field

The ensemble models, which apply pooling functions on top of various deep convolutional neural network architectures, work by minimising bias and variance to improve the accuracy of models. Modern medical image classification pipelines frequently combine unique architectures or models that have undergone varied training to maximise performance. Utilising the prediction data from several methodologies leads to greater inference quality and bias or error reduction. Ensemble learning has proven effective and functional in many problem domains and with significant applications in medical imaging. Deep ensemble learning builds several classifiers or a set of basic learners and merges their output to lessen overall variance. Compared to using a single classifier or base learner, the accuracy is greatly increased when combined with a group of classifiers or base learners. A potent deep learning approach known as deep ensemble learning has demonstrated clear benefits in numerous applications. The generalisation capacity of an ensemble can be significantly higher than that of a single learner by utilising numerous learners.

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: No human studies are presented in this manuscript.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Stacked Ensemble Deep Learning for Pancreas Cancer Classification using Extreme Gradient Boosting

July 12, 2023

Abstract

Ensemble learning aims to improve prediction performance by combining several models or forecasts. However, how much and which ensemble learning techniques are useful in deep learning-based pipelines for pancreas computed tomography (CT) image classification is a challenge. Ensemble approaches are the most advanced solution to many machine learning problems. These techniques entail training multiple models and combining their predictions to improve the predictive performance of a single model. This article introduces the idea of Stacked Ensemble Deep Learning (SEDL), a pipeline for classifying pancreas CT medical images. The weak learners are Inception V3, VGG16, and ResNet34, and we employed a stacking ensemble. By combining the first-level predictions, an input train set for XGBoost, the ensemble model at the second level of prediction, is created. Extreme Gradient Boosting (XGBoost), employed as a strong learner, will make the final classification. Our findings showed that SEDL performed better, with a 98.8% ensemble accuracy, after some adjustments to the hyperparameters. The Cancer Imaging Archive (TCIA) public access dataset consists of 80 pancreas CT scans with a resolution of 512 * 512 pixels. We concluded that implementing the SEDL technique is an effective way to strengthen the robustness and increase the performance of the pipeline for classifying pancreas CT medical pictures. Interestingly, grouping like-minded or talented learners does not make a difference. Instead, combining several diverse, weak, heterogeneous machine learning methods enhances generalisation.

Keywords: Stacking Ensemble, Deep Learning, XGBoost, Hyperparameters, Classification, Pancreas Segmentation

1 Introduction

Recent years have seen a remarkable expansion in automated medical image analysis [1], [2]. Deep neural networks are one of the most well-liked and commonly used algorithms for computer vision problems [3]–[5]. Deep convolutional neural network architectures serve as the foundation for this trend. These designs performed well with clinicians and demonstrated strong prediction ability [6]–[8]. Deep learning-based automated medical image analysis is currently a prominent academic topic being integrated into clinical workflow.

Medical image classification categorises an entire image into specified classes according to a pancreatic cancer diagnosis or condition. To increase the accuracy of diagnoses or automate time-consuming procedures, these models are intended to be used as clinical decision support for physicians [9], [10]. Ensemble learning combines multiple machine learning models to produce superior

results. The fundamental idea is a deep learning model combination, such as VGG16 [11], Inception V3 [12], [13], and ResNet34 [14], [15], can produce more accurate results than any single machine learning model.

The world has moved towards a data-driven medical time, and artificial intelligence and machine learning are using this data to analyse diseases, forecast treatment results, and guide decision-making and drug development. The supervised machine learning classification approach is among the most popular and frequently used fields [16]. It aids in sorting data into several categories. It has many uses, including speech recognition, image classification, fraud detection, and spam identification in email and other diagnostic tests and medical procedures [17]. Identifying a set of target classes (feature to identify in images) is the first step in the supervised learning process for classifying images. A model is trained to identify the target classes using labelled sample images. Rough pixel data was used as the model's input in early computer vision algorithms.

Medical image classification is a work in medical image analysis that entails categorising medical pictures, such as X-rays, MRI scans, and CT scans, into different groups depending on the type of image or the presence of particular structures or disorders [18]. One of the most significant issues in the field of image recognition is medical image classification, which aims to categorise medical images to aid physicians in the diagnosis of disease or further study. In general, there are two parts to medical image classification. The first step is to take the image and extract its useful elements. The second phase entails using the features to create classification models for the dataset of images. A tough, tedious, and time-consuming operation, classifying medical images into multiple classes traditionally required clinicians to apply their professional skills to extract features [19]. This method is likely to produce unstable or unpredictable results. The application research for medical image categorization has had significant worth compared to previous studies.

Compared to individual models, ensemble approaches have higher predictive accuracy [20]. The approaches are particularly helpful when a dataset contains linear and non-linear data types since they combine several models to manage this data. Through ensemble approaches, bias and variance can be decreased, and the model is typically neither under nor overfitted [21]. Always less noisy and more stable is an ensemble of models. Utilising an ensemble of models rather than a single model was the main objective while developing our proposed SEDL model to boost robustness and performance. The spread or dispersion of the forecasts and model performance decreases, while an ensemble can produce superior predictions and performance than any contributing model.

Models are also known as learners. Each learner is regarded as "weak" on their own. On the other hand, a robust model can be created when several weak learners are linked in some way. In our case, this robust model—SEDL—will be called an ensemble model. The models should be as diverse as possible. Each model may perform well on different subsets of the data. When the individual models are combined, their flaws are balanced out. The premise of ensemble models is that there is greater strength in multiple models than in a single one. In the approach section, we outline our suggested SEDL model and go into ensemble learning techniques [22]. The experiments section describes our experiments and the procedures we used to develop and evaluate the model. The results section reviews the findings and some of the performance metrics we employed. Lastly, the concluding section provides a succinct summary.

2 Related Work

Recent research has shown that ensemble learning algorithms are a significant component of the most effective and precise medical picture categorisation pipelines [23]–[26]. Finding a model that

maximises prediction accuracy is the goal of machine learning. The method to combine numerous models into a better predictor closer to an optimal model was developed because it is challenging to determine the optimal model. The combining of models to provide improved prediction performance is what is referred to as ensemble learning. Deep ensemble learning [25], [27] incorporates ensemble learning techniques into a deep learning pipeline. Several recent research studies have effectively applied this method to increase a pipeline's performance and resilience for classifying medical images.

In the study [28], coronaviruses are categorised into two stages. Before this, the classification algorithm was applied to four subsets without first extracting the feature. Before SVM classification, the subsets underwent a vectorization step. The second stage involved the extraction of features using five distinct feature extraction methods, including the Discrete Wavelet Transform (DWT) (13), the Grey Level Size Zone Matrix (GLSZM) (12), Local Directional Patterns (LDP) (10), the Grey Level Run Length Matrix (GLRLM) (11), and the Grey Level Cooccurrence Matrix (GLCM) (7-9). SVM was used to classify the characteristics after that (14). The classification phase included cross-validation methods of 2x, 5x, and 10x. The mean classification results were found once the cross-validations were finished. The best classification accuracy was obtained as 99.68% with 10-fold cross-validation and GLSZM feature extraction method.

A convolutional Support Vector Machine is suggested in [29] and can categorise computed tomography pictures automatically. The CSVM model is trained from scratch instead of the pre-trained Convolutional Neural Networks trained with the transfer learning method. The dataset is separated into two parts: training (75%) and testing (25%) to assess the effectiveness of the CSVM approach. Three distinct numbers of SVM kernels are contained in each block of the CSVM model. With 94.03% ACC, 96.09% SEN, 92.01% SPE, 92.19% PRE, 94.10% F1-Score, 88.15% MCC, and 88.07% Kappa metric values, the performance of pre-trained CNN networks and CSVM models are compared. The CSVM (77, 33, 11) model performs the best.

According to empirical data, ensemble learning-based pipelines are often preferable because they combine the capabilities of various models to concentrate on distinct aspects while compensating for each model's specific limitations [30]. It is still unclear whether ensemble learning models are useful in deep learning-based medical image categorisation pipelines. Although generic ensemble learning is not new, the literature has yet to investigate how ensemble learning strategies affect deep learning-based classification. Few publications have begun to examine the deep ensemble learning field, unlike the authors who provide thorough reviews on general ensemble learning, such as according to [22]. While [31], [32] They provide descriptions or analyses of general deep ensemble learning methods, while [33] investigated deep learning-based ensemble learning methods in bioinformatics. In this study, we aim to create a SEDL pipeline to demonstrate the effect of ensemble learning techniques on deep convolution neural network performance for medical image categorisation. We wish to evaluate the performance of deep learning techniques as base learners for XGBoost [34] to uncover potential performance gains.

Firstly, the authors, [35], use typical image augmentation on minority classes, which involves rotating, scaling, and other modifications to images. The total number of photographs is now 260 more than before picture enhancement. Fourth, four manually produced features are selected from all of the images. Seventy-eight features are created for each image by combining these feature vectors. The 78 feature vectors from the 260 photos are then oversampled using the SMOTE technique. This methodology results in 495 feature vectors. Using these feature vectors, sAE and PCA are both trained. Out of a total of 78 features, 20 features were selected in this study using the sAE and PCA approaches. To classify data, SVM is trained using 495 vectors with 20

characteristics. The depth of the unbalanced structure in the dataset leads to the requirement for picture augmentation and data over-sampling. Only two photos will be created in many classes if image augmentation is the only technique used, and these two images will be nearly identical. Overfitting within the class takes place here. When solely employing a synthetic data oversampling method, performance data is obtained that is only a simulation of real-world performance, which may not improve. Two data replication methods are therefore combined for this purpose.

For detecting COVID-19 [36], where early diagnosis is crucial for human life, CT image features are retrieved using the convolutional neural network architecture, currently the most successful image processing technique. Representational power is improved by fusing data and four CNN architectures' output features. Finally, the features integrated with the feature ranking algorithm are sorted, and their length is decreased. The dimensional curse is averted in this way. A subset was produced from 16 16 (Subset-1) and 32 32 patches taken from 150 CT scans. Three thousand patch photos are designated as "COVID-19" or "No finding" within the parameters of the suggested approach for use during the training and testing phases. The data that has been processed was then categorised using the Support Vector Machine technique.

3 Methods and Techniques

Deep ensemble learning is commonly defined as assembling a group of many predictions derived from various deep convolutional neural network models [32]. Ensemble learning must now be defined in deep learning as combining data, most frequently predictions, for a single inference due to recent breakthrough methodologies. This data or these forecasts may come from a single model, several independent models, or none. Using numerous unique models in a stacking method, we investigated the performance impact of ensemble learning techniques in this investigation. The ensemble prediction is obtained by combining the predictions sent into the XGBoost [37] as a new train set.

On the TCIA dataset, the SEDL classification architectures we employed in this study had good results. VGG16, ResNet34, and Inception V3 are the architectures selected for first-level predictions. At the second level prediction, we used XGBoost. These networks have been specially designed for classifying pdac CT images and have undergone extensive training. Each of these networks has a different hyper-parameter value that is optimised. Here's a basic description of these architectures:

3.1 Deep Ensemble Learning

In a machine learning paradigm known as ensemble learning, multiple learners are trained to solve the same problem [22], [31]. In contrast to traditional deep learning approaches, which attempt to learn one hypothesis from training data, ensemble methods attempt to generate and combine several hypotheses. Several students who make up an ensemble are typically referred to as base students. An ensemble's capacity for generalisation is typically substantially greater than that of basic learners. Deep ensemble learning [26], [30] is intriguing because it can transform weak predictors that are only marginally better than random guesses into strong predictors that can make extremely precise predictions. In this sense, "base learners" and "weak learners" are synonyms. To create heterogeneous learners, we used a variety of learning algorithms. Since there is no one base learning algorithm in heterogeneous learning, some people prefer the terms "base learners" over "individual learners" or "component learners."

By considering the nature of machine learning as exploring a hypothesis space for the best accurate hypothesis, [38]–[40] provided explanations for why the generalisation ability of an ensemble is typically significantly stronger than that of a single learner. The training data may not contain enough details to identify only one best learner, which is the first justification. For instance, multiple students might perform equally well on the practice data set. Therefore, it could be preferable to combine these learners. The second concern is that the learning algorithms' search methods may be flawed. For instance, even if there is a single best hypothesis, it could be challenging since applying the algorithms yields less-than-ideal hypotheses.

3.2 Stacking

A technique for combining multiple regression or classification models, as illustrated in Figure 1. Bagging and boosting are the two most well-known ensemble modelling techniques. Bagging [41], [42] allows for averaging several comparable models with high volatility to reduce variance. Increasing [42], [43] generates many incremental models to reduce bias while minimizing variation.

Stacking is a different paradigm. Stacking's goal is to explore a space of several models for the same problem. A learning problem can be approached using several models to learn a portion of the problem but not the entire problem space. As a result, we can generate several learners and use them to generate intermediate predictions, one for each learned model. Then we add a new model that targets the same target as the intermediate forecasts.

The name of the last model comes from the claim that it is layered on top of the others. As a result, we may enhance our total performance, and frequently we produce a model superior to each intermediate model.

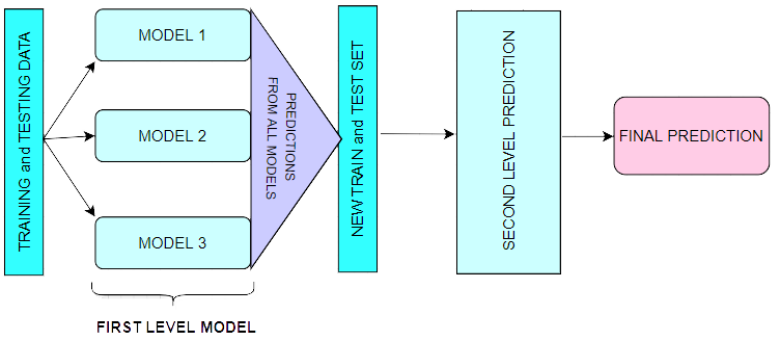


Figure 1: The demonstration of a stack for first-level prediction

3.3 Max Voting Implementation

Voting is effective because it gives most models' opinions more weight than the vote on a single model. When the models may vote on discrete possibilities, maximum voting is utilised. The choice is the proposal that has received the most votes. It is employed to solve classification issues. The option with the most votes is chosen. Each deep learning model casts a vote. Max voting comes in two flavours: harsh and soft. We attempt to predict five classes, T0 (no cancer signs), T1 (tumour $\leq$ 2cm), T2 (tumour $>$ 2cm $\leq$ 4cm), T3 (tumour $>$ 4cm), and T4 (tumour spread), using three

basic classifiers, VGG16 [44], ResNet34 [12], and Inception V3 [15]. After training the base classifier models, we attempt to predict a single point's class.

3.4 Models used in Ensemble

The following models were used in the first level prediction stage to implement the ensemble:

3.4.1 VGG16

Oxford University developed one of the most popular deep learning architectures, VGG16 [11]. It has 41 strata that have been disturbed: 13 convolutional layers (Conv.), 16 weight layers, and 3 Fully Connected layers. All Conv Layers in VGG16 use a tiny 3x3 kernel (filter) with stride 1. Conv. Max pooling layers always follow layers. Three fixed-size channels of 224x224 pixels make up the input for VGG16.

The three Fully Connected layers in VGG16 [45] have various depths. The soft-max layer controls the probability specified for the input image and serves as the output layer. If the weights are started randomly, VGG16 [44] requires extensive training, just like any pre-trained model. Consequently, transfer learning methods are generally used in Convolutional Neural Network (CCN) models. TL describes a technique where a model developed for one activity is used in some way for another comparable task.

3.4.2 ResNet34

Given that neural networks are models of the human brain and how it thinks, it stands to reason that deeper networks would be used to mimic the deeper thinking required to solve difficult problems. The issue of vanishing gradients is the fundamental issue facing deep networks [12].

ResNet34 [13] is a neural network that solves the issue of training deep learning networks using skip connections. It "skips" several convolutional layers in each basic network block, providing alternate paths for the original and derived data and enabling training to be completed more quickly. The following equation, which describes such skip connections, adds the outputs of the preceding blocks to the ones that follow:

$$y = F(x) + x \tag{1}$$

When F is the residual function, as in Equation (1), x is the input, and y is the output. Two convolution layers, a pooling layer (3x3 size), a (ReLU) activation function, and batch normalisation comprise each basic block. ResNet has significantly improved neural network performance by stacking more layers onto neural networks to provide a deeper architecture and, thus, deeper learning instead of shallower learning. The ResNet-34 [46], [47] (ResNet with 34 layers) comprises a fully connected layer, a max-pooling layer (3x3 size), a layer with average pooling, and 33 convolutional layers.

3.4.3 Inception V3

In 2014, Google proposed the GooLeNet network, which is a CNN. It employs the Inception network topology, which reduces the number of network parameters while increasing network depth. As a result, it is frequently used in image classification jobs. Because its primary component is the [48], [49], the GoogLeNet network is also known as the Inception network. The four primary

versions of Google Net are Inception v1 (2014), Inception v2 (2015), Inception v3 (2015), Inception v4 (2016), and Inception-ResNet (2016). The Inception module typically includes one maximum pooling and three varying-sized convolutions.

After the convolution operation, the channel is aggregated for the net output of the preceding layer, and the nonlinear fusion is then carried out. This method can avoid over-fitting while enhancing the network's expression and flexibility to various scales. The main component of Inception v3 is a network structure created by Keras that has already been trained on Image Net. Three channels and a preset image input size of 299*299 are used. Unlike the Inception v1 and v2 network structures, the Inception v3 network structure splits huge volume integrals into smaller convolutions using a convolution kernel splitting technique [22].

3.5 XGBoost

The sophisticated use of the gradient boosting method is called XGBoost. An extremely successful deep learning algorithm is XGBoost [34]. Compared to other gradient-boosting approaches, XGBoost is nearly ten times faster and has strong predictive power. Additionally, it contains a range of regularisations that lessens overfitting and enhances general performance. It is often referred to as the "regularised boosting" technique. Due to the following characteristics, XGBoost [37] is compared to other approaches as being superior:

1. Regularisation [50]:
    - Unlike XGBoost, the standard Gradient Boosting Machine (GBM) implementation does not regularise.
    - As a result, XGBoost also aids in lowering over-fitting.
2. Parallel Processing [51]:
    - In comparison to GBM, XGBoost uses parallel processing and is faster.
    - XGBoost also supports Hadoop implementation.
3. High Flexibility [52]:
    - With the help of XGBoost, users can create unique optimisation goals and assessment standards, giving the model an entirely new dimension.
4. Handling Missing Values [53]:
    - A built-in procedure in XGBoost can deal with missing values.
5. Tree Pruning [54]:
    - After splitting the tree to the maximum depth chosen, XGBoost begins to prune the tree backwards and removes the splits above without benefit.
6. Built-in Cross-Validation [55]:
    - Because the user can perform cross-validation at each stage of the boosting process, obtaining the precise ideal number of boosting iterations in a single run is simple with XGBoost.

3.6 Proposed Model

Stacking [56] is a sophisticated ensemble learning model. The fundamental idea behind stacking is that we base our future forecasts on the base data's derivative models. Now, the outcomes would also be comparable if the models were. To better understand the outcome, we purposefully chose various models, as those models may have learned some aspects of the data more effectively. Stacking [57] takes diverse weak learners into account. Stacking combines several weak learners by training a meta-model to produce predictions based on the numerous predictions that these weak learners returned.

The suggested SEDL framework has two layers, depicted in Figure 2. Each basic classifier in the first layer—VGG16, Inception V3, and ResNet34—was trained using training data. Different classifiers in the representation learning technique of stacking express heterogeneity for various features [58]. High diversity and high accuracy are two conditions the first layer's basic classifiers must meet to learn features from the raw input effectively. The three base classifiers are highly good at solving the nonlinear problem, but their modelling approaches are very dissimilar.

Inception V3, VGG16, and ResNet34 are used to make first-level predictions from the dataset, divided into training and test sets. The three models are each trained from scratch on the dataset to generate feature predictions for pdac CT images. The feature predictions from the first-level prediction models are combined and sent as input to the XGBoost-based second-level prediction. The new test data is used to evaluate the performance and progress of the algorithm training and adjust or optimise it for improved results. XGBoost will make the final prediction and classifications.

The training grounds for ensemble classification systems are developing various single classification methods to solve the same job and the subsequent aggregation of their results using a particular combiner. The suggested model uses transfer learning for image classification [59]. It holds that a model can effectively function as a generic model of the visual world if trained on a sizable and general enough dataset. By training a big model on a big dataset, we can use these learnt feature maps instead of starting from scratch. Take advantage of a prior network's representations learned to identify significant aspects in fresh data. On top of the pre-trained model, add a fresh classifier trained from scratch to reuse the feature maps previously learned for the dataset. Following that, the classier receives the features learned from the ensemble models. Transfer learning has made it possible to train deep neural networks even with little or no data by utilising the capacity to reuse previous models and their understanding of new issues [60]. Transfer learning has the advantages of using fewer data, requiring the classifier to train more quickly, and performing better on neural networks.

We ultimately integrated all three models to produce the first layer of the stacking model due to their similarities and differences, as well as the positive results from cross-validation. Strong generalisation skills are necessary for the meta-learner in the second layer to rectify the bias of various learning algorithms toward the training set and prevent the over-fitting effect through aggregation. Therefore, we used XGBoost [50] to benefit from its second-level meta-learner generalisation capacity. This approach employs the greatest likelihood method to estimate the parameters after assuming all data follow the logical distribution.

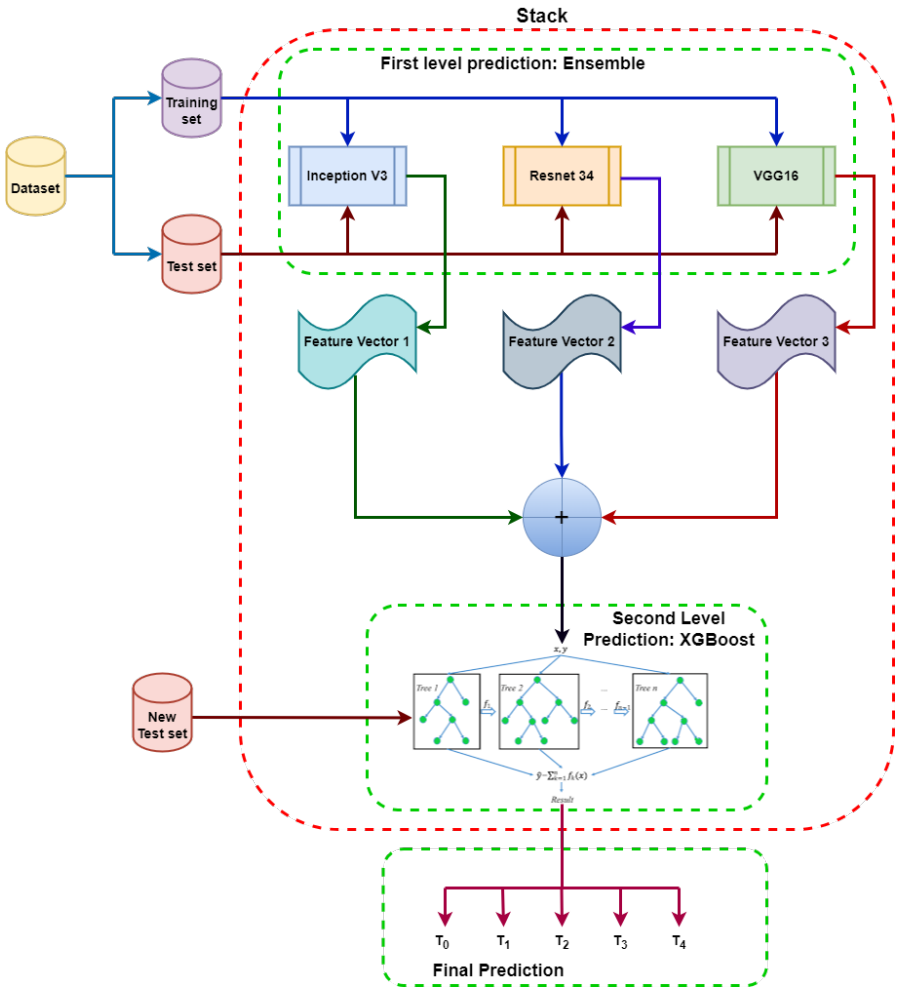


Figure 2: Proposed SEDL Model: Each model at the first level prediction produces its own independent set of the base feature vector. These features from each ensemble model are then fused to make the meta feature vector that will be input into XGBoost to learn from and evaluated using the new test set. The results are different classes depending on the prediction based on the features obtained.

A broad range of hyperparameters offered by the XGBoost algorithm needed to be adjusted for a better classification model. A machine learning model's overall performance and behaviour can be enhanced by hyperparameter adjustment [61]. It is a particular parameter chosen before learning that occurs outside the model. If the loss function is not minimised, improper results may follow from a lack of hyperparameter tweaking. We aim to have as few errors as feasible produced by our model. Finding a collection of ideal hyperparameter values that maximises the model's performance, minimises loss, and generates superior outputs is the main goal of hyperparameter tweak-

ing. We had to adjust the XGBoost hyperparameters like model_colsample_bytree, model_gamma, model_learning_rate, model_max_depth, model_min_child_weight, model_n_estimators, model_reg_lambda, and model_subsample, to achieve better classification results.

4 Experiments

We used logic to create parallel ensemble models [62] that take advantage of the basic learners' independence (rather than the Sequential ensemble models, which employ logic to leverage the dependence between the base learners.). As a result, labelling errors generated by one model are distinct from those made by a different independent model. The ensemble model can then average out the mistakes as a result.

The data for the classification of pdac contains CT scans from 80 subjects. Two hundred twenty-two images were used, of which 0.80 was train data and 0.20 was test data. As shown in Fig: 3, there were different stages of pdac:

- $T0 = Image\_A\_stage$,
- $T1 = Image\_C\_stage$,
- $T2 = Stage\_between\_Image\_C\_and\_E$,
- $T3 = Image\_E\_stage$,
- $T4 = Image\_G\_stage$.

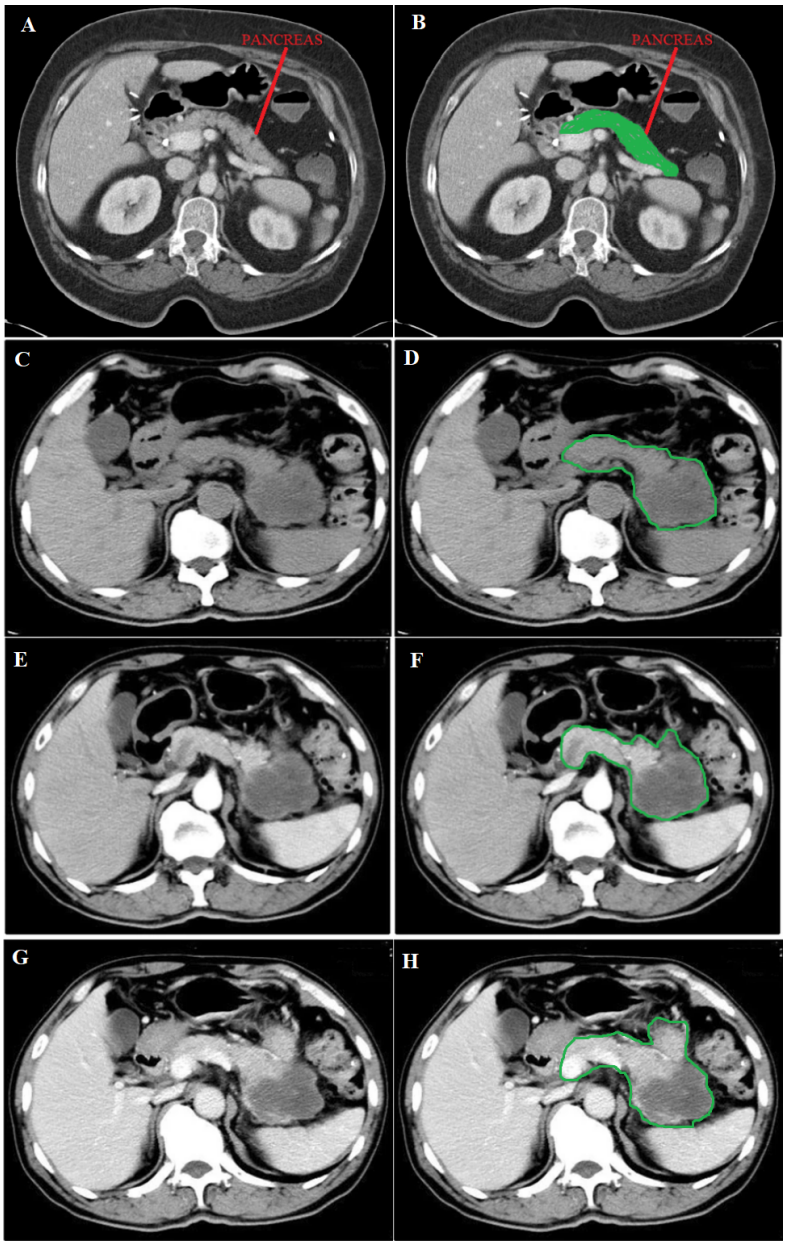


Figure 3: A - Normal image of the pancreas, C - Unenhanced phase, E - Arterial phase and D - Portal phase. B - in green shows the position of the normal pancreas, while D, F, and H - in green circulation show the spread of pdac.

Image A shows a Normal pancreas, while Image C shows an unenhanced stage of pdac. Image

E shows the arterial phase, a contrast-enhanced CT series, in which the contrast is still in the arteries and has not reached the organs and other soft tissues. Image G shows the portal phase, a contrast-enhanced CT series at its peak enhancement.

The steps followed to do the experiments on each base model are in the following Algorithm: 1, which shows how VGG16, ResNet34 and Inception V3 are trained and used in the first-level prediction classification phase.

Algorithm 1: Model algorithm: The model performs an ensemble classification using three models, that is, VGG16, ResNet34 and Inception V3. These steps are followed to train each model on the same dataset. A confusion matrix and accuracy score are produced for each model.

Input: Input: train_images.
Input: Input: test_labels.
Input: M(p): Model_parameters.
/* Read training and testing data. */
le = preprocessing.LabelEncoder(). /* Encode labels from text to integers. */
Split data into test and train datasets /* already split but assigning to the meaningful convention. */
Normalize images. /* Do not normalize masks, rescale to 0 to 1. */
Normalize pixel values to between 0 and 1
Sanity check, view few mages /* . */
encoder = OneHotEncoder. /* One hot encode y values for the neural network. */
for $M(p) \rightarrow set$ **do**
    // Parameters for model.
    $IMG\_HEIGHT = X\_train.shape[1]$
    $IMG\_WIDTH = X\_train.shape[2]$
    $IMG\_CHANNELS = X\_train.shape[3]$
    $num\_labels = 1$ // Binary.
    $input\_shape = (IMG\_HEIGHT, IMG\_WIDTH, IMG\_CHANNELS)$
    $batch\_size = 32$
end for
MAKE NEW MODEL // combining both feature extractor and x.
Train various models: VGG16, ResNet34, and Inception V3
Convert the history.history dict to a pandas DataFrame and save as csv. // For future plotting.
for $M(p) \rightarrow Check\_History$ **do**
    // Check history plots, one model at a time. $history = VGG16\_history$
    $history = ResNet34\_history$
    $history = Inception_V3\_history$
    /* plot the training and validation accuracy and loss at each epoch. */
    $acc = history.history['jacard\_coef']$
    $val\_acc = history.history['val\_jacard\_coef']$
end for
Load one model at a time for testing.
Calculate the Accuracy Score for all test images and the average.
return $Classification\_Report$

We then ensemble the predictions from base models to be the new train set for the final classifier

in the second level prediction phase, tested for XGBoost, LGBM and Random Forest, as shown in Algorithm: 2 steps.

Algorithm 2: XGBoost classifier: Steps followed to implement XGBoost, LGBM, or Random Forest as an image classifier at the second level prediction phase.

1 Import libraries. `/* testing of different models, XGBoost, Random Forest and LGBM. So the respective libraries are imported. */`
2 Send test data through the same feature extractor process.
3 Model average/sum Ensemble. `/* Simple sum of all outputs/predictions and argmax across all classes. */`
4 argmax across classes.
5 Inverse le transform. `/* to get the original label back. */`
6 Print overall accuracy.
7 Weighted average ensemble.
8 weighted_preds = np.tensordot(preds, weights, axes=((0),(0))) `/* Use tensordot to sum the products of all elements over specified axes. */`
9 Explore metrics for the ideal weighted ensemble model.
10 ideal_weighted_preds = np.tensordot(preds, ideal_weights, axes=((0),(0))) `/* Use tensordot to sum the products of all elements over specified axes. */`
11 Print confusion matrix.

5 Results

Compared to using a single model, ensemble learning enhances model results by combining the knowledge and approach of multiple models. For ensemble models to function, several weak learners must be combined into one strong learner, which helps by lowering bias, reducing variation, or increasing prediction accuracy.

5.1 XGBoost hyperparameters tuning

A wide variety of hyperparameters are available with the XGBoost algorithm [63]. For the XGBoost model to work better and to its full potential, we need to be able to adjust these hyperparameters. Optimising a machine learning model's hyperparameters is essential for enhancing its general behaviour and performance. It's a parameter chosen before the learning process and takes place independently of the model. Table: 1 shows the results after adjusting the parameters to have optimal results from the XGBoost classifier at second-level prediction after implementing the ensemble technique at first-level prediction. The best model learning rate was 0.085, with an accuracy of 0.9677. An accuracy of 0.9643 was achieved at a model n_estimator of 250. The model_colsample_bytree at 0.9 achieved better accuracy of 0.9686 than the default model_colsample_bytree of 1. The model subsample was best at 0.8 with an accuracy of 0.984. Model gamma at 4.5 was the best, with 0.989 accuracy.

Table 1: XGBoost parameter tuning at a second level prediction: XGBoost as the classifier had to be tested using different parameters to obtain optimal accuracy. Table: 1 shows, mlr (model_learning_rate), mcb (model_colsample_bytree), mg (model_gamma), mmd (model_max_depth), mmw (model_min_child_weight), mne (model_n_estimators), ms (model_subsample), Accuracy, Macro Avg (Macro Average) and Weighted Avg (Weighted Average

Constant Parameter	**Changing Parameter**	**Parameter value**	**Accuracy**	**Macro Avg**	**Weighted Avg**
$mcb = 0.8$, $mmd = 5$, $mmw = 1$, $ms = 0.8$	mlr	0.01	0.9345	0.9345	0.9298
		0.05	0.9442	0.9542	0.9542
		0.085	0.9677	0.9695	0.9676
		0.1	0.9475	0.9434	0.9428
		0.15	0.8876	0.8844	0.8830
	mne	800	0.9468	0.9453	0.9332
		550	0.9547	0.9567	0.9579
		250	0.9643	0.9665	0.9629
		100	0.9475	0.9444	0.9430
$mlr = 0.085$, $mmd = 5$, $mmw = 1$, $mne = 250$	mcb	0.6	0.9553	0.9555	0.9552
		0.8	0.9574	0.9574	0.9574
		0.9	0.9686	0.9678	0.9678
		1	0.9578	0.9572	0.9575
	ms	0.6	0.9276	0.9243	0.9231
		0.8	0.9840	0.9840	0.9822
		0.9	0.9733	0.9641	0.9624
		1	0.9693	0.9679	0.9661
$mcb = 0.9$, $mlr = 0.01$, $mmd = 5$, $mmw = 1$, $mne = 250$, $ms = 0.8$	mg	0	0.9445	0.9445	0.9398
		1	0.9442	0.9432	0.9428
		3	0.9713	0.9671	0.9660
		4.5	0.9895	0.9889	0.9880
		5	0.9856	0.9836	0.9839

If the loss function is not minimised, findings without proper hyperparameter adjustment may be unreliable. We want as few errors as possible to be produced by our model. Hyperparameter tuning aims to find ideal hyperparameter values that maximise model performance, minimise loss, and create superior outputs [64].

5.2 Averaging and Weighted Average

When averaging, multiple forecasts are made for each data point, similar to the max voting method. In this approach, the final prediction is made by averaging the results of all the models. Probabilities can be computed for classification issues using averaging.

Weighted Average is an extension of the averaging method. Different weights are assigned to each model, indicating the significance of each model for prediction. For instance, the responses from these two models will be given greater weight than those from the other model if two of the models are critics and one has no prior expertise in this subject. Table: 2 shows the averaging and weighted average for the three models tested as classifiers at the final stage. XGBoost had the best ensemble average and ensemble weighted average at 0.987 and 0.981, respectively. The averaging for the ensemble shows a better improvement of results than when using individual models at first-level prediction when implemented at the second-level prediction.

Table 2: Average score table showing the averaging of using a classifier at second level prediction after implementing individual models without ensemble at first level prediction. Avg Ensemble (Average Ensemble) and W Avg Ensemble (Weighted Average Ensemble) for different model classifiers at second level prediction are also shown after implementing ensemble at first leave prediction.

Classifier	**Inception V3 Averaging**	**VGG16 Averaging**	**ResNet-34 Averaging**	**Avg Ensemble**	**W Avg Ensemle**
XGBoost	0.942	0.979	0.964	0.987	0.981
LGBM	0.905	0.953	0.948	0.955	0.951
RF	0.940	0.942	0.939	0.952	0.949

5.3 Confusion Matrix

The confusion matrix is made up of five target classes. A $5x5$ matrix is used to evaluate how well a proposed SEDL classification model performs at second-level prediction, as shown in Figure: 4 using different models. XGBoost showed less misclassification than the other models. The deep learning model's predictions are compared to the target values in the matrix, providing a comprehensive understanding of our classification model's effectiveness and the types of errors it makes.

The confusion matrix for XGBoost shows two negative images, while two hundred and twenty images are positive, out of two hundred and twenty images. Each column in the confusion matrix represents the instances of that projected class. Each row of the confusion matrix displays an instance of the real class. It provides insight into errors that are being made and errors made by a classifier.

Figure 4: Confusion matrix for Random Forest, LGBM and XGBoost as the classifiers

5.4 Classification Report

The primary classification metrics at second-level prediction are provided for each class in the classification report as in Table: 3. The classification report [65] visualiser shows the model's precision, recall, F1 and support scores for the three models tested at second-level prediction after the ensemble is implemented at first-level prediction. In contrast to overall accuracy, which can hide functional deficiencies in one class of a multi-class problem, this provides a deeper insight into the classifier behaviour.

Precision can be thought of as a measure of a classifier's accuracy. Recall is a measure of a classifier's completeness; it is the classifier's ability to identify all positive instances correctly. The F1 score is a weighted harmonic mean of precision and recall, where 1.0 represents the best, and 0.0 represents the worst. The precision, recall, and f1-score are all greater than 0.98, implying that the SEDL model with XGBoost is a powerful classification tool. Support is the number of actual class occurrences in the specified dataset.

5.5 Training and Validation Loss

The training loss metric assesses how well a deep learning model fits the training data. In other words, it assesses the model's error on the training set. It should be noted that the training set is a subset of the dataset that was initially used to train the model. The training loss [66] is calculated by adding the sum of errors for each example in the training set. It is also important to note that the training loss is calculated after each batch, typically represented by plotting a training loss curve.

Validation loss [67], on the other hand, is a metric used to assess the performance of a deep learning model on the validation set. The validation set is a subset of the dataset set aside to test

Table 3: Classfication report for ensemble model showing Avg Ensemble (Average Ensemble), W Avg Ensemble (Weighted Average Ensemble), Accuracy, Macro avg (Macro average) and Weighted avg (Weighted average) for different models

Classifier	**Class**	**Precision**	**Recall**	**F1 Score**	**Support**
XGBoost					
	T0	0.98	0.98	0.98	61
	T1	1.00	0.98	0.99	41
	T2	1.00	0.98	0.99	42
	T3	1.00	1.00	1.00	37
	T4	0.95	1.00	0.98	41
		Metrics			
	Metric	**Precision**	**Recall**	**F1 Score**	**Support**
	Accuracy	-	-	0.99	222
	Macro avg	0.99	0.99	0.99	222
	Weighted Avg	0.99	0.99	0.99	222
LGBM					
	T0	0.78	0.97	0.86	61
	T1	1.00	0.83	0.91	41
	T2	1.00	0.95	0.98	42
	T3	1.00	0.78	0.88	37
	T4	0.95	1.00	0.98	41
		Metrics			
	Metric	**Precision**	**Recall**	**F1 Score**	**Support**
	Accuracy	-	-	0.91	222
	Macro avg	0.95	0.91	0.92	222
	Weighted Avg	0.93	0.91	0.92	222
RF					
	T0	0.71	1.00	0.83	61
	T1	1.00	0.59	0.74	41
	T2	1.00	0.95	0.98	42
	T3	1.00	0.84	0.91	37
	T4	1.00	1.00	1.00	41
		Metrics			
	Metric	**Precision**	**Recall**	**F1 Score**	**Support**
	Accuracy	-	-	0.89	222
	Macro avg	0.94	0.88	0.89	222
	Weighted Avg	0.92	0.89	0.89	222

Random Forest

LGBM

XGBoost

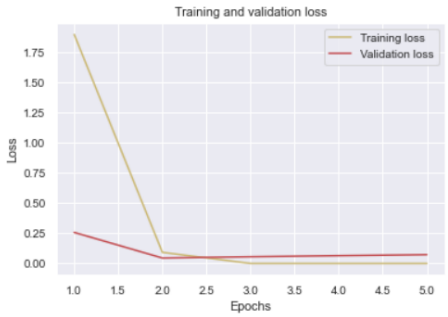


Figure 5: The graphs show the Validation and Testing Loss graphs for the three models used as learners at the first level of prediction.

the model's performance. The validation loss is calculated similarly to the training loss by adding the errors for each example in the validation set. Furthermore, the validation loss is calculated after each epoch, indicating whether the model requires additional tuning or adjustments. For the validation loss, we created a learning curve.

The second level classification by XGBoost displays good fit learning curves, which is the learning algorithm's goal, and exists between an overfit and underfit model. Figure 5shows training and validation loss for different models at second-level prediction. A training and validation loss identifies a good fit that decreases stability with a minimal gap between the two final loss values, as shown by the XGBoost graph.

The model's loss is almost always lower on the training dataset than on the validation dataset, which implies that there will be some disparity between the train and validation loss learning curves known as the "generalisation gap." The plot of learning curves for XGBoost shows a good fit, indicated by:

- *T*he training loss plot decreases to the point of stability.
- *T*he validation loss plot approaches stability and has a small gap with the training loss.

5.6 ROC

A graph depicting the effectiveness of a classification model at all classification thresholds is called a ROC curve (receiver operating characteristic curve). Figure 6 shows the ROC curve for the three models tested at second-level prediction. The True Positive and False Positive rates are plotted on this curve. The chart demonstrates the connection between the true and false positive rates. It was decided to compare each class to every other using the One-vs-Rest methodology. XGBoost had an AUC (Area under the ROC Curve) of 0.98, which is best than the other models.

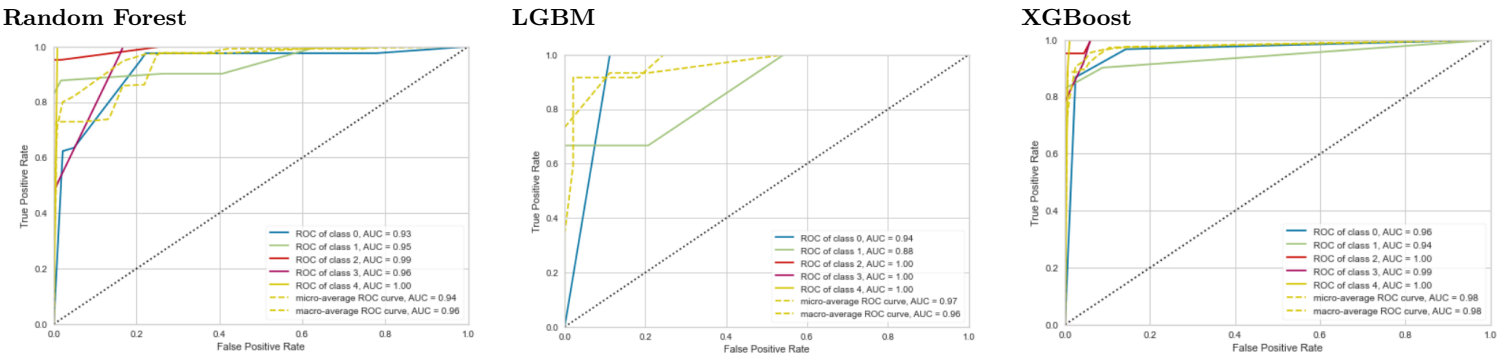


Figure 6: ROC curves for the Random Forest, LGBM and XGBoost implementation at second level prediction. AUC is the Area under the ROC Curve.

5.7 Discussion

The idea of stacked ensemble deep learning is used in our proposal of a unique deep learning framework for identifying pdac. AUC was 98.6%, F1 was 99.6%, and accuracy was 98.8% obtained from the studies. Using several neural network models trained on ImageNet [68], features from images are extracted in this method and fed into a classifier for prediction. They suggested an ensemble model incorporating all pre-trained models' outputs to achieve the best pneumonia recognition performance. This model surpassed individual models. On uncovered data from the Guangzhou Women and Children's Medical Centre dataset, their ensemble model achieved an accuracy of 96.4% and a recall of 99.62%. Our model SEDL performs end-to-end classification in two tiers by utilising each model's ability to anticipate any hidden features in the pdac CT images.

According to [69], a stacked ensemble of the top three retrained models shows promising performance (accuracy: 0.941; 95% Confidence Interval (CI): [0.899, 0.985], Area Under the Curve (AUC): 0.995; 95% CI: [0.945, 1.00]). The ensemble approaches' accuracy (P =.759) and AUC (P =.831) do not differ statistically, according to one-way ANOVA analysis. The classification of our SEDL model was enhanced by the knowledge conveyed through modality-specific learning of pertinent features. The ensemble model decreased the prediction variance and sensitivity to changes in training data. The outcomes of combining them outperform current technology.

Using chest x-ray and Canny edge detected images, [70], demonstrate a deep ensemble learning for TB identification. This strategy increases the base classifiers' diversity of mistakes by adding a new feature to the TB detection classifiers. The initial X-ray images were used to extract the first set of features, and the edge-detected image was used to extract the second set of features. Two publicly accessible datasets were utilised to assess the suggested method. The findings demonstrate that the suggested ensemble technique achieved the best accuracy, sensitivity, and specificity values of 89.77%, 90.91%, and 88.64%. This suggests that the detection rate can be increased by utilising various characteristics retrieved from various images.

While the solution presented by [71] combines elements from AlexNet, GoogleNet, and ResNet.

The study's main contribution is creating an ensemble capable of performing TB classification by training these structures from scratch. On a combined dataset created using openly accessible standard datasets, the suggested approach is trained and evaluated. The ensemble outperforms most of the known approaches, with an accuracy of 88.24% and an area under the curve of 0.93.

Reasons that could have contributed to high accuracy, besides the advantages of combined power in ensemble techniques, include:

- We maintained very little correlation between the base classifiers utilised. This will ensure that those classifiers' faults are also diversified. The base classifiers are anticipated to work in tandem to get superior classification results. Only classifiers trained on similar features were combined in most studies analysed. The base classifiers' correlation error is high as a result of this.
- We maintained the original image size, as most researchers lowered the original image's size during training to save on computing costs. Even with the most potent GPU hardware, training a highly complicated model requires significant processing power compared to the original image size.
- For training, hundreds of thousands of images from each class must be collected. In doing so, a classifier will be more precise. The amount of training data that is readily available, however, is frequently subpar due to the small number of datasets. Because of this, scientists are looking for several solutions to create a reliable classifier.

6 Conclusion

Ensemble techniques show great potential. Ensemble methods often improve detection accuracy. An ensemble of several features might provide better detection results. An ensemble of different deep-learning techniques could also be considered because ensembles perform better if the errors of the base classifiers have a low correlation. Ensemble learning has proven effective and functional in many problem domains and with significant applications in medical imaging. Deep ensemble learning builds several classifiers or a set of basic learners and merges their output to lessen overall variance. Compared to using a single classifier or base learner, the accuracy is greatly increased when combined with a group of classifiers or base learners. A potent deep learning approach known as deep ensemble learning has demonstrated clear benefits in numerous applications. The generalisation capacity of an ensemble can be significantly higher than that of a single learner by utilising numerous learners.

The ensemble models, which apply pooling functions on top of various deep convolutional neural network architectures, work by minimising bias and variance to improve the accuracy of models. Modern medical image classification pipelines frequently combine unique architectures or models that have undergone varied training to maximise performance. Utilising the prediction data from several methodologies leads to greater inference quality and bias or error reduction.

The fact that the user cannot understand the knowledge acquired by ensembles is a fundamental flaw in existing ensemble methods. The direction of increasing the comprehensibility of ensembles is significant yet poorly unexplored. Another significant problem is that, even though diversity is known to be crucial in ensembles, there are currently no satisfactory diversity measurements. In our upcoming work, we attempt to overcome these challenges to increase ensemble learning's ability to contribute to more applications.

References

[1] G. Wang, "A perspective on deep imaging," *IEEE access*, vol. 4, pp. 8914–8924, 2016.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[3] A. Fourcade and R. Khonsari, "Deep learning in medical image analysis: A third eye for doctors," *Journal of stomatology, oral and maxillofacial surgery*, vol. 120, no. 4, pp. 279–288, 2019.

[4] E. Goceri and N. Goceri, "Deep learning in medical image analysis: Recent advances and future trends," 2017.

[5] J. Wang, H. Zhu, S.-H. Wang, and Y.-D. Zhang, "A review of deep learning on medical image analysis," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 351–380, 2021.

[6] G. Currie, K. E. Hawk, E. Rohren, A. Vial, and R. Klein, "Machine learning and deep learning in medical imaging: Intelligent imaging," *Journal of medical imaging and radiation sciences*, vol. 50, no. 4, pp. 477–487, 2019.

[7] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on mri," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.

[8] S. K. Zhou, H. Greenspan, C. Davatzikos, *et al.*, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, 2021.

[9] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 24 365–24 398, 2021.

[10] J.-G. Lee, S. Jun, Y.-W. Cho, *et al.*, "Deep learning in medical imaging: General overview," *Korean journal of radiology*, vol. 18, no. 4, pp. 570–584, 2017.

[11] T. Kaur and T. K. Gandhi, "Automated brain image classification based on vgg-16 and transfer learning," in *2019 International Conference on Information Technology (ICIT)*, IEEE, 2019, pp. 94–98.

[12] R. Godasu, D. Zeng, and K. Sutrave, "Transfer learning in medical image classification: Challenges and opportunities," 2020.

[13] Y. Alsabahi, L. Fan, and X. Feng, "Image classification method in dr image based on transfer learning," in *2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, 2018, pp. 1–4.

[14] Y. Dai, Y. Gao, and F. Liu, "Transmed: Transformers advance multi-modal medical image classification," *Diagnostics*, vol. 11, no. 8, p. 1384, 2021.

[15] H. Xie, X. Zeng, H. Lei, *et al.*, "Cross-attention multi-branch network for fundus diseases classification using slo images," *Medical Image Analysis*, vol. 71, p. 102 031, 2021.

[16] S. Rauschert, K. Raubenheimer, P. Melton, and R. Huang, "Machine learning and clinical epigenetics: A review of challenges for diagnosis and classification," *Clinical epigenetics*, vol. 12, no. 1, pp. 1–11, 2020.

[17] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: A new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, pp. 1071–1092, 2020.

[18] M. Siddiq, "Ml-based medical image analysis for anomaly detection in ct scans, x-rays, and mris," *Devotion Journal of Community Service*, vol. 2, no. 1, pp. 53–64, 2020.

[19] R. Goel and I. Adak, "Improved detection of brain tumor & classification using cnn," in *2021 12th international conference on computing communication and networking technologies (ICCCNT)*, IEEE, 2021, pp. 1–7.

[20] S. Sundaramurthy, C. Saravanabhavan, and P. Kshirsagar, "Prediction and classification of rheumatoid arthritis using ensemble machine learning approaches," in *2020 International Conference on Decision Aid Sciences and Application (DASA)*, IEEE, 2020, pp. 17–21.

[21] K. Zhang, B. Khosravi, S. Vahdati, *et al.*, "Mitigating bias in radiology machine learning: 2. model development," *Radiology: Artificial Intelligence*, vol. 4, no. 5, e220010, 2022.

[22] M. A. Ganaie, M. Hu, *et al.*, "Ensemble deep learning: A review," *arXiv preprint arXiv:2104.02395*, 2021.

[23] Y. Yang, Y. Hu, X. Zhang, and S. Wang, "Two-stage selective ensemble of cnn via deep tree training for medical image classification," *IEEE Transactions on Cybernetics*, 2021.

[24] D. Xue, X. Zhou, C. Li, *et al.*, "An application of transfer learning and ensemble learning techniques for cervical histopathology image classification," *IEEE Access*, vol. 8, pp. 104 603–104 618, 2020.

[25] D. Müller, I. Soto-Rey, and F. Kramer, "An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks," *arXiv preprint arXiv:2201.11440*, 2022.

[26] P. Pintelas and I. E. Livieris, *Special issue on ensemble learning and applications*, 2020.

[27] R. A. Pratiwi, S. Nurmaini, D. P. Rini, M. N. Rachmatullah, and A. Darmawahyuni, "Deep ensemble learning for skin lesions classification with convolutional neural network," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, p. 563, 2021.

[28] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (covid-19) classification using ct images by machine learning methods," *arXiv preprint arXiv:2003.09424*, 2020.

[29] U. Özkaya, Ş. Öztürk, S. Budak, F. Melgani, and K. Polat, "Classification of covid-19 in chest ct images using convolutional support vector machines," *arXiv preprint arXiv:2011.05746*, 2020.

[30] N. An, H. Ding, J. Yang, R. Au, and T. F. Ang, "Deep ensemble learning for alzheimer's disease classification," *Journal of biomedical informatics*, vol. 105, p. 103 411, 2020.

[31] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1249, 2018.

[32] I. Kandel, M. Castelli, and A. Popovič, "Comparing stacking ensemble techniques to improve musculoskeletal fracture image classification," *Journal of Imaging*, vol. 7, no. 6, p. 100, 2021.

[33] Y. Cao, T. A. Geddes, J. Y. H. Yang, and P. Yang, "Ensemble deep learning in bioinformatics," *Nature Machine Intelligence*, vol. 2, no. 9, pp. 500–508, 2020.

[34] S. Ramaneswaran, K. Srinivasan, P. Vincent, and C.-Y. Chang, "Hybrid inception v3 xgboost model for acute lymphoblastic leukemia classification," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.

[35] Ş. Öztürk, U. Özkaya, and M. Barstuğan, "Classification of coronavirus (covid-19) from x-ray and ct images using shrunken features," *International Journal of Imaging Systems and Technology*, vol. 31, no. 1, pp. 5–15, 2021.

[36] U. Özkaya, Ş. Öztürk, and M. Barstugan, "Coronavirus (covid-19) classification using deep features fusion and ranking technique," *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*, pp. 281–295, 2020.

[37] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of xgboost-based algorithm for breast cancer classification," *Machine Learning with Applications*, vol. 6, p. 100 154, 2021.

[38] J. Wu, P. Guo, Y. Cheng, H. Zhu, X.-B. Wang, and X. Shao, "Ensemble generalized multiclass support-vector-machine-based health evaluation of complex degradation systems," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 5, pp. 2230–2240, 2020.

[39] K. Gu, Z. Xia, and J. Qiao, "Stacked selective ensemble for pm 2.5 forecast," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 660–671, 2019.

[40] K. M. Alam, N. Siddique, H. Adeli, *et al.*, "A dynamic ensemble learning algorithm for neural networks," *Neural Computing and Applications*, vol. 32, no. 12, pp. 8675–8690, 2020.

[41] Y. Liu and F. Long, "Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning," in *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Springer, 2019, pp. 113–121.

[42] R. Chatterjee, A. Datta, and D. K. Sanyal, "Ensemble learning approach to motor imagery eeg signal classification," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*, Elsevier, 2019, pp. 183–208.

[43] R. Iranzad, X. Liu, W. A. Chaovalitwongse, *et al.*, "Gradient boosted trees for spatial data and its application to medical imaging data," *IISE Transactions on Healthcare Systems Engineering*, vol. 12, no. 3, pp. 165–179, 2022.

[44] Z.-P. Jiang, Y.-Y. Liu, Z.-E. Shao, and K.-W. Huang, "An improved vgg16 model for pneumonia image classification," *Applied Sciences*, vol. 11, no. 23, p. 11 185, 2021.

[45] M. Chhabra and R. Kumar, "An advanced vgg16 architecture-based deep learning model to detect pneumonia from medical images," in *Emergent Converging Technologies and Biomedical Systems*, Springer, 2022, pp. 457–471.

[46] M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, and U. R. Acharya, "Convolutional neural networks for multi-class brain disease detection using mri images," *Computerized Medical Imaging and Graphics*, vol. 78, p. 101 673, 2019.

[47] S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora, and R. B. Pachori, "Application of deep learning techniques for detection of covid-19 cases using chest x-ray images: A comprehensive study," *Biomedical Signal Processing and Control*, vol. 64, p. 102 365, 2021.

[48] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3d deep learning on medical images: A review," *Sensors*, vol. 20, no. 18, p. 5097, 2020.

[49] H. Li, S. Zhuang, D.-a. Li, J. Zhao, and Y. Ma, "Benign and malignant classification of mammogram images based on deep learning," *Biomedical Signal Processing and Control*, vol. 51, pp. 347–354, 2019.

[50] Z. Shilong *et al.*, "Machine learning model for sales forecasting by using xgboost," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, IEEE, 2021, pp. 480–483.

[51] J. Henriques, F. Caldeira, T. Cruz, and P. Simões, "Combining k-means and xgboost models for anomaly detection using log datasets," *Electronics*, vol. 9, no. 7, p. 1164, 2020.

[52] D. Zhang, H.-D. Chen, H. Zulfiqar, *et al.*, "Iblp: An xgboost-based predictor for identifying bioluminescent proteins," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.

[53] X. Zhang, C. Yan, C. Gao, B. A. Malin, and Y. Chen, "Predicting missing values in medical data via xgboost regression," *Journal of healthcare informatics research*, vol. 4, no. 4, pp. 383–394, 2020.

[54] O. Sagi and L. Rokach, "Approximating xgboost with an interpretable decision tree," *Information Sciences*, vol. 572, pp. 522–542, 2021.

[55] C. Wang, C. Deng, and S. Wang, "Imbalance-xgboost: Leveraging weighted and focal losses for binary label-imbalanced classification with xgboost," *Pattern Recognition Letters*, vol. 136, pp. 190–197, 2020.

[56] Y. Wang, D. Wang, N. Geng, Y. Wang, Y. Yin, and Y. Jin, "Stacking-based ensemble learning of decision trees for interpretable prostate cancer detection," *Applied Soft Computing*, vol. 77, pp. 188–204, 2019.

[57] A. U. Haq, J. P. Li, Z. Ali, *et al.*, "Stacking approach for accurate invasive ductal carcinoma classification," *Computers and Electrical Engineering*, vol. 100, p. 107 937, 2022.

[58] J. Tang, J. Liang, C. Han, Z. Li, and H. Huang, "Crash injury severity analysis using a two-layer stacking framework," *Accident Analysis & Prevention*, vol. 122, pp. 226–238, 2019.

[59] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the covid-19 infected patients using densenet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 15, pp. 5682–5689, 2021.

[60] M. Iman, H. R. Arabnia, and K. Rasheed, "A review of deep transfer learning and recent advancements," *Technologies*, vol. 11, no. 2, p. 40, 2023.

[61] B. Zhang, R. Rajan, L. Pineda, *et al.*, "On the importance of hyperparameter optimization for model-based reinforcement learning," in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2021, pp. 4015–4023.

[62] C. Wei, J. Zhang, T. Valiullin, W. Cao, Q. Wang, and H. Long, "Distributed and parallel ensemble classification for big data based on kullback-leibler random sample partition," in *International Conference on Algorithms and Architectures for Parallel Processing*, Springer, 2020, pp. 448–464.

[63] J. Sommer, D. Sarigiannis, and T. Parnell, "Learning to tune xgboost with xgboost," *arXiv preprint arXiv:1909.07218*, 2019.

[64] S. Putatunda and K. Rama, "A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of xgboost," in *Proceedings of the 2018 international conference on signal processing and machine learning*, 2018, pp. 6–10.

[65] A. Luque, A. Carrasco, A. Martin, and A. de Las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 2019.

[66] S. Deepak and P. Ameer, "Brain tumor classification using deep cnn features via transfer learning," *Computers in biology and medicine*, vol. 111, p. 103 345, 2019.

[67] G. S. Dhillon, P. Chaudhari, A. Ravichandran, and S. Soatto, "A baseline for few-shot image classification," *arXiv preprint arXiv:1909.02729*, 2019.

[68] V. Chouhan, S. K. Singh, A. Khamparia, *et al.*, "A novel transfer learning based approach for pneumonia detection in chest x-ray images," *Applied Sciences*, vol. 10, no. 2, p. 559, 2020.

[69] S. Rajaraman and S. K. Antani, "Modality-specific deep learning model ensembles toward improving tb detection in chest radiographs," *IEEE Access*, vol. 8, pp. 27 318–27 326, 2020.

[70] M. H. A. Hijazi, S. K. T. Hwa, A. Bade, R. Yaakob, and M. S. Jeffree, "Ensemble deep learning for tuberculosis detection using chest x-ray and canny edge detected images," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 4, p. 429, 2019.

[71] R. Hooda, A. Mittal, and S. Sofat, "Automated tb classification using ensemble of deep architectures," *Multimedia Tools and Applications*, vol. 78, pp. 31 515–31 532, 2019.

5. 4 Conclusion

In conclusion, the accurate classification of CT pancreas images holds immense potential for revolutionising the diagnosis and treatment of pancreatic diseases. This field has made significant strides in enhancing our understanding of pancreatic conditions, enabling earlier detection, personalised treatment approaches, and improved patient outcomes through advanced deep-learning techniques and image analysis algorithms. As research and technology evolve, we can anticipate further refinement of classification models, leading to more robust and efficient tools that empower healthcare professionals to make timely and precise clinical decisions. Ultimately, the classification of CT pancreas images is a pivotal component in the ongoing mission to combat pancreatic diseases and elevate the standards of modern medical practice.

Chapter 6

Future Work and Conclusion

6. 1 Introduction

There is still much room for improvement in the design [52] of PDAC overall survival prediction models. Ensemble modelling has been examined to increase the prediction models' accuracy and effectiveness. The models used convolutional neural networks (CNN), widely employed in computer vision and deep learning technologies. While various CNN-based algorithms can be integrated into health products and adhere to medical imaging requirements, the typical CNN algorithm still has several flaws and needs improvement. The encoder-decoder architecture is covered in this chapter, along with its limitations and why attention mechanisms can be used to solve some of the issues.

The initial driving force behind deep learning attention research was the bottleneck issue in the traditional encoder-decoder system [53]. The key problem is that a neural network needs to condense all the essential data from a given text into a fixed vector. As the size of the input images grows, [54], research demonstrated that the traditional encoder-performance decoders rapidly decline. The attention mechanism removes the information bottleneck caused by the fixed-size context vector, which enables additional information to be distributed throughout the network.

6. 2 Future work

Humans must choose and assess the most pertinent information in their working memory at any time in a complex world. Attention mechanisms are used to accomplish this, which entails disregarding other perceivable information to selectively focus on a certain portion of it when and where it is needed. Making it possible for people to swiftly choose the most important information from a large

amount of information enhances the efficiency and accuracy of perceptual information processing [52]. The attention mechanism in humans is a sophisticated cognitive ability. People may ignore some of the major information when given it while focusing on secondary information. This capability of self-selection is known as "attention". The attention mechanism of the neural network enables the selection of specific traits by focusing on a subset of inputs.

6. 2.1 Attention Networks

The goal of attention mechanisms in deep learning is to identify dependencies between various input data items. Attention mechanisms provide distinct input elements with varying weights or priorities rather than treating the information equally. This improves performance and interpretability by enabling the model to focus on the most important data for a particular task.

Saliency-based attention [55], influenced by outside stimuli, and max-pooling and gating, which advance more appropriate values to the following phase, are the two categories into which the human attention process falls. Top-down conscious attention [56], which is task-focused and has a preset goal, is known as "focused attention". It is the foundation for most deep-learning attention mechanisms and allows humans to focus actively and intentionally on a specific object. Except for particular expressions, it is typically called "focused attention".

An attention model, sometimes called an attention mechanism, is a neural method networks use to process data. Segmenting difficult problems into manageable attentional zones and processing them sequentially aids neural networks in completing challenging tasks. The attention mechanism enables neural networks to tackle intuitive and difficult tasks like translation and subtitle generation. This is similar to how the human brain solves big problems by breaking them down into simpler tasks and focusing on them one at a time. Until it classifies the entire dataset, the neural network concentrates on particular features of a complicated input [52].

The attention mechanism can be attributed to the interpretability of deep learning, performance enhancements, and intelligible neural architecture behaviour. In some cases, such as when attention weight is visualised, it might offer an intuitive explanation.

The concept of attention [57], seen as a feature of numerous and various perceptual and cognitive operations, has pervaded most areas of perception and cognition study during the past few decades. Hence, attention reflects this organisation to the extent that these mechanisms are specialised and decentralised. The executive control procedures aid in establishing the system's priorities and these methods are widely communicated. The brain's selection mechanisms are active at every stage, from

sensory processing to awareness and decision-making. Its effects can be quantified through conscious reflection, electrophysiology, and brain imaging, and attention has come to be used broadly to describe how the brain regulates its information processing. For a long time, attention mechanisms have been investigated from various angles.

Natural Language Processing (NLP) saw the emergence of Attentional Networks (NANs) [58] in the middle of 2014, resulting in substantial advancements in neural machine comprehension, sentiment analysis, machine translation, question-answering, and transfer learning. Many domains, including computer vision, reinforcement learning, and robotics, have used these networks. Regarding citation counts and inventions, the most pertinent collection of papers includes recurrent neural networks (RNNs) search, transformer, Memory Networks, "show, attend, and tell," and RAM.

6. 2.2 Attention Network Researches

Philosophers, psychologists, neuroscientists, and computer scientists have studied attention concepts and functions for decades. This property has been extensively investigated in deep neural networks over the last six years. Table: 6.1 shows some research implementing attention networks. In future work, we look forward to exploring how PDAC survival prediction models can be improved using this technique, maybe from the angle of designing hybrid ensemble and attention network variants to harness the advantages of the two.

In recent years, there has been an increase in interest in using attention networks for medical picture processing. It has been demonstrated that attention networks are efficient for several tasks, including:

6. 2.2.0.1 Image segmentation: By enabling the model to concentrate on the pertinent areas of the image, attention networks can be utilised to increase the accuracy of image segmentation tasks. Attention networks [59], for instance, have been used to separate breast cancer lesions in mammograms and brain tumours in MRI scans.

6. 2.2.0.2 Image classification: By enabling the model to learn the relative relevance of various features in the image, attention networks can increase the accuracy of image classification tasks. For instance [60], attention networks have been used to categorise diabetic retinopathy in retinal images and skin lesions in dermoscopy images.

Table 6.1: Proposed attention network and the various application domains tested on different datasets

Attention Network	Application	Dataset	Paper
Densely connected attention network	diagnosing COVID-19	2019 (COVID-19) based on chest CT images	[61]
Adaptive Attention Network	COVID-19 Detection	DLAI3 dataset,1 COVIDx dataset [46], and COVIDGR_E	[62]
Cross-Attention Networks	Thoracic Disease Image Classification	Chest X-Ray14 Dataset, CheXpert Dataset	[63]
Progressive Alternating Attention Networks	Medical Image Segmentation	2018 Data Science Bowl (DSB) Challenge, ISIC 2018 Challenge, Kvasir-Instruments	[64]
Self-Attention Networks	Image Recognition	ImageNet	[65]
Graph Attention Networks	Segment Labeling in Coronary Artery	71 CCTA images	[66]
Dense Convolutional Attention Network	Medical Image Analysis	MICCAI2015 and Kvasir	[67]
	Hyperspectral Image Classification	University of Pavia (UP)	[68]
	image super-resolution	Set5	[69]
High-order Attention Network	Medical Image Segmentation	REFUGE, Drishti-GS1 for optic disc/cup segmentation, DRIVE for blood vessel segmentation, and LUNA for lung segmentation	[70]
pyramid attention network	medical image fusion	The Whole Brain Atlas	[71]
channel attention networks	intracranial vessel segmentation	intracranial arterial blood vessel	[72]
Triple attention network	medical image segmentation	retinal blood vessels (DRIVE and STARE), cells (ISBI 2012), cutaneous melanoma (ISIC 2017), and intracranial blood vessels	[73]
Feature fusion attention network	medical image segmentation	ISIC 2017, CHAOS	[74]

The ability of attention networks to efficiently capture and emphasise important elements within images has attracted a lot of attention in medical imaging research. These networks can better execute various medical imaging tasks like segmentation, classification, and illness diagnosis by

utilising attention mechanisms to concentrate on particular regions or patterns of interest. There is still more to learn about using attention networks in medical imaging. The outcomes of recent research, however, indicate that attention networks may have the ability to increase the accuracy of tasks involving medical picture processing considerably. Attention networks will probably be used in more medical imaging applications as this field develops.

The following are some study ideas about how to use attention networks in medical imaging:

1. Exploring the use of attention networks for different medical image analysis tasks: This research proposal will investigate the use of attention networks in several medical image-processing tasks, including object recognition, picture segmentation, and image classification. This study aims to uncover the parameters that influence attention networks' performance in medical imaging and to ascertain which attention network types perform best for certain tasks.

2. Developing new attention networks for medical image analysis: New attention networks will be created to examine medical images as part of this research project. To increase the accuracy of activities involving medical image analysis, this research aims to create attention networks that are more efficient than current attention networks.

3. Improving the interpretability of attention networks for medical image analysis: This research plan aims to make medical image processing more comprehensible. To help radiologists and other medical practitioners understand how attention networks decide what to pay attention to, this study aims to develop techniques that can be used to explain the attention weights produced by these networks.

These are only a handful of the numerous study proposals that may be made for using attention networks in imaging for medical purposes. As this field of study develops, attention networks for medical image processing will probably be used in even more creative and useful ways.

The following are some specific research queries that these proposals might address:

- Which attention network types can be applied to the processing of medical images?
- What aspects of attention networks for medical imaging are successful?
- How can new attention networks be created, especially for processing medical images?
- For medical image analysis, how may the interpretability of attention networks be enhanced?

Other approaches could be used to answer these research issues, such as:

- Various attention networks and medical image processing tasks are being tested in studies.
- Creating new attention network theoretical models.
- Evaluating the requirements and preferences of radiologists and other medical specialists through surveys.

The findings of this study may significantly impact the field of medical image analysis. We could make it possible to employ these potent tools to detect diseases early, treat patients more successfully, and enhance patient outcomes by increasing the accuracy and interpretability of attention networks.

6. 3 Conclusion

One of the most important steps in the medical imaging workflow is manually segmenting target structures and organs at risk. Its drawbacks include that it can take a doctor many hours to complete each patient, and inter- and intra-observer variability can occur. These problems are addressed by automatic segmentation, which uses computer techniques. Methods for precise and effective auto-segmentation have been developed due to deep learning and computer vision developments.

Literature to support and direct the commissioning and clinical deployment of these algorithms is becoming more widely available as deep learning applications and auto-segmentation in medical imaging get more attention and adoption. There isn't any prospective data to support the widespread hypotheses that workflow advantages and time savings are related to deep learning auto-segmentation. Compared to the CT-based models in this study, which may be more broadly applicable to typical medical imaging practises, this study evaluates PDAC patients' survival using CT images based on deep learning segmentation and classification techniques.

The best image segmenting method is by hand, and expert manual segmentation is regarded as the gold standard. However, manual segmentation takes a long time, necessitates continual expert supervision, and is subject to wide differences since various experts see things differently. In addition, the background noise may lessen the contrast of the segmentation targets, making it vulnerable to human error. Deep learning algorithms, widely employed for biomedical image segmentation, can overcome manual segmentation drawbacks and speed up the process while enhancing reproducibility and reducing workload.

Convolutional neural networks (CNNs) extract significant characteristics from inputs using various kernels. The segmentation of cells, pancreatic vessels, and PDAC tumours are only a few examples of biomedical image challenges for which CNNs have been used. Due to its great performance and

accuracy, the U-Net model among the several CNN architectures is frequently chosen for biomedical image segmentation.

UNet, a fully convolutional neural network based on ResNet 34, is proposed as an encoder and point of departure for boundary class separation in computer vision applications. Deep learning architectures such as UNet, ResNet34, and Watershed can be used to perform complex PDAC image segmentation tasks on computer vision datasets with high-quality results. Image segmentation divides an image into smaller fragments, whereas masks and edge classes are used to create binary images with 0.89 accuracy. Deep learning frameworks TensorFlow and Keras are used to build U-Net architectures [75]. UNet aims to capture the characteristics of both context and localisation to produce higher-resolution output on the input image using a continuous shrink layer and the upsampling operator.

The UNet architecture generates a high-resolution, comprehensive PDAC image with labelled pixels. Two convolutions are used to shrink the grid size and create a modern GPU to quickly compute the segmentation of images with a size of 512 * 512 to improve the efficiency, speed, and simplicity of the image segmentation process. Batch normalisation is included in the model implementation because it consistently produces the best results [76]. The functionality of the PyTorch-based U-Net architecture and some alternative model structures will be investigated in future publications.

Using UNet only shows an improved accuracy of 0.90, possibly because of the different features and sizes of data used between the two experiments. Using attention, UNet proved to be the best, obtaining 94% accuracy. We will consider implementing Watershed for a combined model with Attention UNet in future work.

Pancreatic tumour auto-segmentation has appeared to be more difficult [77]. A square-window CNN-based method in multi-parametric MRI produced an average Dice of 0.73; however, very noticeably, this was comparable to a Dice between two different observer outlines. Another intriguing study performed auto-segmentation using contrast-enhanced endoscopic ultrasound images of pancreatic tumours and a U-Net [77]. Intersection over Union (IoU) was used to assess the outcomes rather than dice, and the mean IoU value was 0.77, with minimum and highest values of 0.39 and 0.91, respectively [78]. This demonstrates that applying deep learning yields promising outcomes, but further research is required to achieve consistent accuracy sufficient for therapeutic application. More accurate results have been reported to be made possible by advancements in motion management, image quality, and network design.

We can nevertheless make some attempts to improve segmentation performance despite the limitations of data preparation and execution. First, we may enhance the image quality by fine-tuning gross

tumour volume (GTV) contours, eradicating image artefacts, and perfecting image registration during pre-processing. Second, other image modalities can be added to the multi-modality to increase the segmentation accuracy even more. For instance, magnetic resonance imaging (MRI), which offers superior soft tissue contrast, multi-planar imaging capabilities, and benefits for dose management, is quickly becoming standard practice for PDAC.

We also investigated the factors that can help select a model that best suits any segmentation task and has specific requirements. A Model Analyser (MA) and an Intelligent Automation Model implement these factors in the Inference Engine. The performance metrics for segmentation models are accuracy and mean IoU. The authors discussed selecting the best U-Net model for a new dataset, considering the problem's complexity, memory requirements, training time, number of features and parameters, training data size, and accuracy or interpretability.

Performance is the most important metric for selecting a deep-learning task model. However, the model must also meet additional criteria such as memory requirements, training and prediction time, interpretability, and data format. The MA can use a validation dataset to test the best model for new data from a few options, providing metrics for comparison and making a final decision. A thorough examination of the needs and anticipated outcomes is required to make an informed decision. Pre-trained models shorten model training and error time while providing a benchmark for comparing models to improve accuracy. The interpretability of future work will be emphasised. Intelligent Automation (IA) can interpret internal decision processes at some levels.

SEDL is a pipeline for classifying pancreatic CT medical images. It combines several forecasting models, including Inception V3, VGG16, ResNet34, and a stacking ensemble. The ensemble model at the second level of prediction, XGBoost, receives its input train set. According to the findings, SEDL performed better, with a 98.4% ensemble accuracy.

Ensemble learning is a deep learning method that reduces bias and variance by pooling functions on top of various deep convolutional neural network architectures [79]. It is effective and functional in many problem domains, with important applications in medical imaging. Deep ensemble learning combines the output of several classifiers or a group of basic learners to reduce overall variance and increase accuracy. It can also be significantly higher than for a single learner when many learners are used. This study employs cutting-edge deep learning techniques to classify PDAC using CT medical imaging modalities. The VGG16-XGBoost hybrid model, VGG16, is proposed as the backbone feature extractor and Extreme Gradient Boosting as the image classifier.

When choosing a model for a deep learning task, performance may appear to be the most obvious

metric. However, performance alone will not help you choose the best model for the job. Additional criteria include memory requirements, training and prediction time, interpretability, and data format, which the model must meet to make a more confident decision by incorporating a broader range of factors. When choosing the best model for new data from a few options is difficult, the MA can test them on a validation dataset. This provides metrics for the MA to compare each model and make a final decision.

Machine learning (ML) algorithms have been used to predict pancreatic ductal adenocarcinoma cancer (PDAC) patients' overall survival. According to research, machine learning algorithms can classify, predict, and estimate the patient's state, surgical treatment, optimal resource use, individualised therapy, drug prescription, and better patient management. This thesis presented cutting-edge techniques for cancer survival prediction and recommendations for implementing these techniques [80]. Machine learning (ML) and deep learning (DL) techniques have been used to predict the overall survival of pancreatic cancer patients. Prediction is based on genomics, proteomic data, clinical factors, and pathological images [81]. Identifying challenges in experimental design, data collection, and analysis or validation results is necessary. A model is required to aid in the calculation of individualised survival predictions and treatment decisions. DL techniques such as feature extraction, segmentation, and classification can provide pathologists with a useful prediction tool to help them make informed decisions.

To improve survival prediction for PDAC patients, deep learning techniques such as PyTorch and TensorFlow are used. CNN and FCN are the most commonly used neural networks in deep learning (DL) to create an efficient machine learning-based image processing solution [82]. The U-Net model is widely used in image processing tasks and has proven accurate. An end-to-end model based on DL techniques is proposed to predict the overall survival rate of PDAC patients.

Dual-energy computed tomography (DECT) is a new technique for concurrently acquiring CT images with two different peak energies. To enhance head and neck cancer image quality, tumour-soft tissue boundary assessment, and invasion of crucial structures, iodine could be specifically removed from contrast-enhanced images. Linearly combining dual CT reduces metal artefacts in pseudo-monochromatic images, which is particularly beneficial for head and neck cancer. Theoretically integrating multiple imaging modalities in our network could improve tumour boundary characterisation and offer a reliable target volume estimation.

The issues facing medical image segmentation are presented in this section, which makes it necessary to advance and develop U-Net-based deep learning techniques. First, to diagnose diseases, medical image processing demands the highest level of accuracy. The terms "pixel-level" and "voxel-level"

segmentation are used in medical imaging. In most cases, it is challenging to determine in an image where one cell or organ divides into another. Additionally, even after a given degree of accuracy is achieved using the appropriate deep learning model, the data extracted from the image is typically pre-processed, the relevant network is constructed, and it continues to run by modifying the parameters.

The standards for them, as well as for annotations and the performance of CT or MRI scanners, are not standardised. Medical images are also acquired using a variety of medical equipment. Because of this, trained models for deep learning are only appropriate in certain situations. However, a deep network with poor generalisation may quickly identify false signals in the analysed medical images. Additionally, there is a noticeable difference in size between the positive and negative samples that are always present, which may have a stronger influence on segmentation. U-Net, however, could afford a strategy that would perform better while minimising overfitting. Third, interpretable deep learning models used to analyse medical images are crucial, but there isn't much faith in their anticipated findings.

A CNN with low interpretability is U-Net. The patient's physiological state and an appropriate disease diagnosis may be reflected through segmentation in medical imaging. It is challenging for professional doctors to believe in and accept segmentation for clinical application when it lacks interpretability and confidence. Even though images are the primary tool for disease diagnosis, adding other components has also made it more complex. Understanding and managing these trade-offs makes achieving the interpretability and confidence of medical image segmentation difficult.

Deep learning techniques are used in real-world situations, and their implications for medical imaging are significant [82]. They can categorise, classify, and count illness patterns in medical image analysis, allowing for the expansion of analytical objectives and the generation of therapy prediction models. Deep learning is rapidly progressing, and it has the potential to aid in cancer diagnosis and treatment decision-making. Deep learning and machine learning can advance the field of radiomics in the coming years, regardless of the patient population or type of tumour.

Neural networks can manage the flow of information, features, and resources in real-time, which improves learning. These mechanisms help the network deal with long-term dependencies by filtering out irrelevant stimuli [57]. Although artificial intelligence has advanced rapidly in recent years, particularly in deep learning, interpretability remains a major concern. The attention layer in a neural network model allows one to reason about the model behind its predictions, but it is frequently criticised for being too opaque. Recent research suggests that attention may not be a reliable means of interpreting deep neural networks, making it an intriguing area for future research [52].

Bibliography

[1] J. Ferlay, M. Ervik, F. Lam, *et al.*, *Global cancer observatory: Cancer tomorrow. lyon, france: International agency for research on cancer; 2020*, 2021.

[2] C. Feig, A. Gopinathan, A. Neesse, D. S. Chan, N. Cook, and D. A. Tuveson, "The pancreas cancer microenvironment", *Clinical cancer research*, vol. 18, no. 16, pp. 4266–4276, 2012.

[3] A. Haleem, M. Javaid, R. P. Singh, and R. Suman, "Medical 4.0 technologies for healthcare: Features, capabilities, and applications", *Internet of Things and Cyber-Physical Systems*, 2022.

[4] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "Ai in health and medicine", *Nature medicine*, vol. 28, no. 1, pp. 31–38, 2022.

[5] N. H. Compare, *The centers for medicare, and medicaid services*, 2015.

[6] Y. Wu, B. Chen, A. Zeng, D. Pan, R. Wang, and S. Zhao, "Skin cancer classification with deep learning: A systematic review", *Frontiers in Oncology*, vol. 12, 2022.

[7] A. S. Panayides, A. Amini, N. D. Filipovic, *et al.*, "Ai in medical imaging informatics: Current challenges and future directions", *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 1837–1857, 2020.

[8] S. Deepa, B. A. Devi, *et al.*, "A survey on artificial intelligence approaches for medical image classification", *Indian Journal of Science and Technology*, vol. 4, no. 11, pp. 1583–1595, 2011.

[9] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep learning in medical image analysis", *Deep Learning in Medical Image Analysis: Challenges and Applications*, pp. 3–21, 2020.

[10] M. Hadden, A. Mittal, J. Samra, H. Zreiqat, S. Sahni, and Y. Ramaswamy, "Mechanically stressed cancer microenvironment: Role in pancreatic cancer progression", *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, vol. 1874, no. 2, p. 188 418, 2020.

[11] S. Li and K. Xie, "Ductal metaplasia in pancreas", *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, p. 188 698, 2022.

[12] Y. Nakanuma, Y. Kakuda, and K. Uesaka, "Characterization of intraductal papillary neoplasm of the bile duct with respect to the histopathologic similarities to pancreatic intraductal papillary mucinous neoplasm", *Gut and Liver*, vol. 13, no. 6, p. 617, 2019.

[13] J. G. Reiter, M. Baretti, J. M. Gerold, *et al.*, "An analysis of genetic heterogeneity in untreated cancers", *Nature Reviews Cancer*, vol. 19, no. 11, pp. 639–650, 2019.

[14] R. Sharma, "A comparative examination of colorectal cancer burden in european union, 1990–2019: Estimates from global burden of disease 2019 study", *International Journal of Clinical Oncology*, vol. 27, no. 8, pp. 1309–1320, 2022.

[15] J. Yu, X. Yang, W. He, and W. Ye, "Burden of pancreatic cancer along with attributable risk factors in europe between 1990 and 2019, and projections until 2039", *International journal of cancer*, vol. 149, no. 5, pp. 993–1001, 2021.

[16] J. D. Mizrahi, R. Surana, J. W. Valle, and R. T. Shroff, "Pancreatic cancer", *The Lancet*, vol. 395, no. 10242, pp. 2008–2020, 2020.

[17] J.-X. Hu, C.-F. Zhao, W.-B. Chen, *et al.*, "Pancreatic cancer: A review of epidemiology, trend, and risk factors", *World journal of gastroenterology*, vol. 27, no. 27, p. 4298, 2021.

[18] R. H. Hruban, M. M. Gaida, E. Thompson, *et al.*, "Why is pancreatic cancer so deadly? the pathologist's view", *The Journal of pathology*, vol. 248, no. 2, pp. 131–141, 2019.

[19] M. T. Roth, D. B. Cardin, and J. D. Berlin, "Recent advances in the treatment of pancreatic cancer", *F1000Research*, vol. 9, 2020.

[20] A. M. Chekroud, J. Bondar, J. Delgadillo, *et al.*, "The promise of machine learning in predicting treatment outcomes in psychiatry", *World Psychiatry*, vol. 20, no. 2, pp. 154–170, 2021.

[21] V. Kotu and B. Deshpande, *Data science: concepts and practice*. Morgan Kaufmann, 2018.

[22] P. Mounika and S. G. Rao, "Machine learning and deep learning models for diagnosis of parkinson's disease: A performance analysis", in *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, IEEE, 2021, pp. 381–388.

[23] S. Fuchs, A. Carpenter, M. Carroll, and K. Hale, "A hierarchical adaptation framework for adaptive training systems", in *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems: 6th International Conference, FAC 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011. Proceedings 6*, Springer, 2011, pp. 413–421.

[24] A. Plaat, W. Kosters, and M. Preuss, "High-accuracy model-based reinforcement learning, a survey", *Artificial Intelligence Review*, pp. 1–33, 2023.

[25] G. M. Yagli, D. Yang, and D. Srinivasan, "Automatic hourly solar forecasting using machine learning models", *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 487–498, 2019.

[26] A. Moore and T. Donahue, "Pancreatic cancer", *Jama*, vol. 322, no. 14, pp. 1426–1426, 2019.

[27] A. Rehman and F. G. Khan, "A deep learning based review on abdominal images", *Multimedia Tools and Applications*, vol. 80, pp. 30 321–30 352, 2021.

[28] G. Capurso, M. Traini, M. Piciucchi, M. Signoretti, and P. G. Arcidiacono, "Exocrine pancreatic insufficiency: Prevalence, diagnosis, and management", *Clinical and experimental Gastroenterology*, pp. 129–139, 2019.

[29] C. L. Faber, J. D. Deem, C. A. Campos, G. J. Taborsky, and G. J. Morton, "Cns control of the endocrine pancreas", *Diabetologia*, vol. 63, pp. 2086–2094, 2020.

[30] D. S. Longnecker, "Anatomy and histology of the pancreas (version 1.0)", *Pancreapedia: The Exocrine Pancreas Knowledge Base*, 2014.

[31] P. C. Freeny and T. L. Lawson, *Radiology of the pancreas*. Springer Science & Business Media, 2012.

[32] S. Covantsev, C. Chicu, N. Mazuruc, and O. Belic, "Pancreatic ductal anatomy: More than meets the eye", *Surgical and Radiologic Anatomy*, vol. 44, no. 9, pp. 1231–1238, 2022.

[33] S. P. Papadakos, N. Dedes, N. Gkolemi, N. Machairas, and S. Theocharis, "The eph/ephrin system in pancreatic ductal adenocarcinoma (pdac): From pathogenesis to treatment", *International Journal of Molecular Sciences*, vol. 24, no. 3, p. 3015, 2023.

[34] L. A. Brosens, M. Mino-Kenudson, and L. D. Wood, "Non-neoplastic and neoplastic pathology of the pancreas", *Gastrointestinal and Liver Pathology*, pp. 455–488, 2024.

[35] S. S. Madaminovich, "Pancreatic cancer: Signs and symptoms", *Journal of Innovation, Creativity and Art*, vol. 2, no. 1, pp. 29–38, 2023.

[36] A. A. Connor and S. Gallinger, "Pancreatic cancer evolution and heterogeneity: Integrating omics and clinical data", *Nature Reviews Cancer*, vol. 22, no. 3, pp. 131–142, 2022.

[37] T. Takikawa, K. Kikuta, S. Hamada, *et al.*, "Clinical features and prognostic impact of asymptomatic pancreatic cancer", *Scientific reports*, vol. 12, no. 1, p. 4262, 2022.

[38] P. X. Mouratidis and G. Ter Haar, "Latest advances in the use of therapeutic focused ultrasound in the treatment of pancreatic cancer", *Cancers*, vol. 14, no. 3, p. 638, 2022.

[39] N. Selvarajoo, J. Stanslas, M. K. Islam, S. R. Sagineedu, H. K. Lian, and J. C. W. Lim, "Pharmacological modulation of apoptosis and autophagy in pancreatic cancer treatment", *Mini Reviews in Medicinal Chemistry*, vol. 22, no. 20, pp. 2581–2595, 2022.

[40] A. Rosenzweig, J. Berlin, S. Chari, *et al.*, "Management of patients with pancreatic cancer using the "right track" model", *The Oncologist*, oyad080, 2023.

[41] I. T. Gram, S.-Y. Park, L. R. Wilkens, L. Le Marchand, and V. W. Setiawan, "Smoking and pancreatic cancer: A sex-specific analysis in the multiethnic cohort study", *Cancer Causes & Control*, vol. 34, no. 1, pp. 89–100, 2023.

[42] G. Le Cosquer, C. Maulat, B. Bournet, P. Cordelier, E. Buscail, and L. Buscail, "Pancreatic cancer in chronic pancreatitis: Pathogenesis and diagnostic approach", *Cancers*, vol. 15, no. 3, p. 761, 2023.

[43] J. Yang, R. Xu, C. Wang, J. Qiu, B. Ren, and L. You, "Early screening and diagnosis strategies of pancreatic cancer: A comprehensive review", *Cancer Communications*, vol. 41, no. 12, pp. 1257–1274, 2021.

[44] L. D. Wood, M. I. Canto, E. M. Jaffee, and D. M. Simeone, "Pancreatic cancer: Pathogenesis, screening, diagnosis and treatment", *Gastroenterology*, 2022.

[45] H.-W. Dai, T.-E. Wang, C.-H. Chu, T.-P. Liu, S.-C. Shih, and W.-C. Lin, "Endoscopic ultrasound for detecting small pancreatic tumor missed by computed tomography", *Journal of Cancer Research and Practice*, vol. 3, no. 4, pp. 136–139, 2016.

[46] A. Săftoiu and P. Vilmann, "Role of endoscopic ultrasound in the diagnosis and staging of pancreatic cancer", *Journal of Clinical Ultrasound*, vol. 37, no. 1, pp. 1–17, 2009.

[47] J. Fouimtizi, A. Hosni, L. Jroundi, A. Slaoui, A. Koutani, and A. I. A. Andaloussi, "Renal pelvic rupture: A case report of an unexpected cause", *International Journal of Surgery Case Reports*, vol. 85, p. 106 176, 2021.

[48] R. Akshai, S. Rohit Krishnan, G. Swetha, and B. Venkatesh, "Detection of lymph nodes using centre of mass and moment analysis", in *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Springer, 2019, pp. 239–246.

[49] C. Bertin, A.-L. Pelletier, M. P. Vullierme, *et al.*, "Pancreas divisum is not a cause of pancreatitis by itself but acts as a partner of genetic mutations", *Official journal of the American College of Gastroenterology| ACG*, vol. 107, no. 2, pp. 311–317, 2012.

[50] W. Park, A. Chawla, and E. M. O'Reilly, "Pancreatic cancer: A review", *Jama*, vol. 326, no. 9, pp. 851–862, 2021.

[51] V. Tonini and M. Zanni, "Pancreatic cancer in 2021: What you need to know to win", *World Journal of Gastroenterology*, vol. 27, no. 35, p. 5851, 2021.

[52] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning", *Neurocomputing*, vol. 452, pp. 48–62, 2021.

[53] A. Santana and E. Colombini, "Neural attention models in deep learning: Survey and taxonomy", *arXiv preprint arXiv:2112.05909*, 2021.

[54] R. Gupta, D. Mehrotra, and R. K. Tyagi, "Computational complexity of fractal image compression algorithm", *IET Image Processing*, vol. 14, no. 17, pp. 4425–4434, 2020.

[55] X. Zhou, T. Tong, Z. Zhong, H. Fan, and Z. Li, "Saliency-cce: Exploiting colour contextual extractor and saliency-based biomedical image segmentation", *Computers in Biology and Medicine*, p. 106 551, 2023.

[56] W. C. Sullivan and D. Li, "Nature and attention", *Nature and psychology: Biological, cognitive, developmental, and social pathways to well-being*, pp. 7–30, 2021.

[57] A. de Santana Correia and E. L. Colombini, "Attention, please! a survey of neural attention models in deep learning", *Artificial Intelligence Review*, vol. 55, no. 8, pp. 6037–6124, 2022.

[58] H.-y. Lee, S.-W. Li, and N. T. Vu, "Meta learning for natural language processing: A survey", *arXiv preprint arXiv:2205.01500*, 2022.

[59] Y. Shen, N. Wu, J. Phang, *et al.*, "An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization", *Medical image analysis*, vol. 68, p. 101 908, 2021.

[60] A. A. Adegun, S. Viriri, and R. O. Ogundokun, "Deep learning approach for medical image analysis", *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–9, 2021.

[61] Y. Fu, P. Xue, and E. Dong, "Densely connected attention network for diagnosing covid-19 based on chest ct", *Computers in Biology and Medicine*, vol. 137, p. 104 857, 2021.

[62] Z. Lin, Z. He, S. Xie, *et al.*, "Aanet: Adaptive attention network for covid-19 detection from chest x-ray images", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4781–4792, 2021.

[63] R. Hou, H. Chang, B. Ma, S. Shan, and X. Chen, "Cross attention network for few-shot classification", *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[64] A. Srivastava, S. Chanda, D. Jha, *et al.*, "Paanet: Progressive alternating attention for automatic medical image segmentation", in *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, IEEE, 2021, pp. 1–4.

[65] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 076–10 085.

[66] N. Hampe, J. M. Wolterink, C. Collet, N. Planken, and I. Išgum, "Graph attention networks for segment labeling in coronary artery trees", in *Medical Imaging 2021: Image Processing*, SPIE, vol. 11596, 2021, pp. 410–416.

[67] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, "Dense convolutional network and its application in medical image analysis", *BioMed Research International*, vol. 2022, 2022.

[68] J. Yin, C. Qi, W. Huang, Q. Chen, and J. Qu, "Multibranch 3d-dense attention network for hyperspectral image classification", *IEEE Access*, vol. 10, pp. 71 886–71 898, 2022.

[69] F. Bai, W. Lu, Y. Huang, L. Zha, and J. Yang, "Densely convolutional attention network for image super-resolution", *Neurocomputing*, vol. 368, pp. 25–33, 2019.

[70] F. Ding, G. Yang, J. Wu, *et al.*, "High-order attention networks for medical image segmentation", in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, Springer, 2020, pp. 253–262.

[71] J. Fu, W. Li, J. Du, and Y. Huang, "A multiscale residual pyramid attention network for medical image fusion", *Biomedical Signal Processing and Control*, vol. 66, p. 102 488, 2021.

[72] J. Ni, J. Wu, H. Wang, *et al.*, "Global channel attention networks for intracranial vessel segmentation", *Computers in biology and medicine*, vol. 118, p. 103 639, 2020.

[73] Y. Li, J. Yang, J. Ni, A. Elazab, and J. Wu, "Ta-net: Triple attention network for medical image segmentation", *Computers in Biology and Medicine*, vol. 137, p. 104 836, 2021.

[74] J. Yu, D. Yang, and H. Zhao, "Ffanet: Feature fusion attention network to medical image segmentation", *Biomedical Signal Processing and Control*, vol. 69, p. 102 912, 2021.

[75] L. Ivanovsky, V. Khryashchev, V. Pavlov, and A. Ostrovskaya, "Building detection on aerial images using u-net neural networks", in *2019 24th Conference of Open Innovations Association (FRUCT)*, IEEE, 2019, pp. 116–122.

[76] Y. Li, N. Wang, J. Shi, X. Hou, and J. Liu, "Adaptive batch normalization for practical domain adaptation", *Pattern Recognition*, vol. 80, pp. 109–117, 2018.

[77] R. R. Savjani, M. Lauria, S. Bose, J. Deng, Y. Yuan, and V. Andrearczyk, "Automated tumor segmentation in radiotherapy", in *Seminars in radiation oncology*, Elsevier, vol. 32, 2022, pp. 319–329.

[78] A. Qureshi, S. Lim, S. Y. Suh, *et al.*, "Deep-learning-based segmentation of extraocular muscles from magnetic resonance images", *Bioengineering*, vol. 10, no. 6, p. 699, 2023.

[79] D. Müller, I. Soto-Rey, and F. Kramer, "An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks", *Ieee Access*, vol. 10, pp. 66 467–66 480, 2022.

[80] J. Bell, "What is machine learning?", *Machine Learning and the City: Applications in Architecture and Urban Design*, pp. 207–216, 2022.

[81] M. Cui and D. Y. Zhang, "Artificial intelligence and computational pathology", *Laboratory Investigation*, vol. 101, no. 4, pp. 412–422, 2021.

[82] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future", *Classification in BioApps: Automation of Decision Making*, pp. 323–350, 2018.