XMARCUS II: Domain Adaptation for Classification of Endoscopic and Robotic-Assisted Peripheral Nerve Tissue

Gunnar Beck Nelson BS; MS

Imperial College London
Department of Surgery & Cancer

Master of Research
Medical Robotics & Image Guided Intervention

Joseph Davids MSc, PhD
Hutan Ashrafian MBA, PhD

October 2024
London, United Kingdom

Keywords: Robotic Surgery, Artificial Intelligence (AI), Computer Vision, Deep Learning, Machine Learning, Surgical Training, Robotic Surgical Training, Endoscopic Surgery, Service-Oriented Architecture (SOA)

Dedication

To those who are close to me, ones who know me, actually know me.

Acknowledgements

Thank you to both Dr. Joseph Davids and Dr. Hutan Ashrafian, who, the fact I had to travel over oceans to be told I mattered, and the work I'm doing matters, hold greater weight on my own path towards healing and becoming the person I want to be. The space to be vulnerable as a student and be acknowledged that I am doing the work necessary means a lot.

Contents

List of Tables

List of Figures

Abstract

The role of advanced software applied for robotic surgery has wide-stream critical application to enhance surgical procedures, across multiple domains and reducing the patient length of stay. Efforts from large technological advances and mediums on the physical plane have led to intensive efforts and expensive hardware equipment. However, software is heavily underutilized as a critical point and crucial technology to detect components of anatomy in real time. This has lad to the movement of Artificial Intelligence for Surgery where collections of surgeons have proposed and attested time and time again for the impending need to categorize components of anatomy in real time to not only enhance the surgical experience, but also reduce the time required for residence training, and decrease the time for medical education, thus reducing vast amounts of costs needed to qualify surgeons for evolving practices. However, practices and techniques consistently evolve to advanced forms for surgery with the goal of minimizing the impact of damage done through surgery, hence minimally invasive surgery, and reducing the patient length of stay by 1 to 6 days. Robotic surgery, though been prevalent and around since the late 1990s, and with notable efforts of adapting manufacturing robots within the surgical domain in earlier decades, the very definition of robotics is malleable and subject to change with apparatuses and mechanics needed for enhanced dexterity and vision. With the advent of robotic surgery, such efforts have not only enhanced surgical dexterity but also vision with different intra-operative perspectives to be close to the internal tissue, thus minimizing the damage done during such events and various techniques.

The advent of advanced software to mimic the intelligence of humans, and make such predictions as humans, artificial intelligence, received a series of boom and bust cycles and experienced moments of AI winters due to vast amounts of wasted expenditures without real innovation to help people; which was also subject to the present hardware of the time and being unable to adapt and mend with the definition of artificial intelligence. However, software collectively has evolved overtime with brilliant experts creating expansive libraries which allow people to adapt such languages and format their own architectures to create software based on the confines of what is artificial intelligence, eventually to further define the discipline. Past

the AI winters, the access of advanced hardware which a personal computer can hold 1.65 million times the greater computational power than that of the first supercomputer which was completed in the 1970s the Cray-1 [99]. Now in this time, we have greater access to more power to mimic what is human. A key paradigm is in pattern recognition. In the evolving landscape of surgical care, the capacity for real-time tissue segmentation represents a pivotal advancement. In the realm of deep learning, this capability hinges on the integration of current knowledge about tissue interaction with the technology underpinning robotic and endoscopic surgery, aiming to minimize the learning curves associated with these procedures. However, the crux of the challenge lies within the domain of AI and computer vision, necessitating the training, testing, and deployment of models, given their rooted principles of deep learning stemming from semi-supervised learning.

Such models require extensive surgical video data for training to ensure accuracy and effectiveness before they can be applied in a clinical setting. However, the acquisition of such data is hampered by stringent regulations and the inherent specificity of surgical video data to individual surgical procedures. The video data by principle is constrained based on discipline and it's a hampered cost to extract such data, refine, and utilize for training, testing, and deploying a present definition supervised model. This specificity limits the utility of trained models to the type of surgery on which they were trained, rather than allowing for their application across a variety of surgical disciplines. Drawing a parallel with the fields of electrocardiography (ECG) and electroencephalography (EEG), where data extracted from these sources is utilized for diverse diagnostic purposes, there's an opportunity to revolutionize the approach to developing AI models for surgery[19][20][21]. Just as ECG and EEG data can inform a wide range of diagnoses and treatments, the ambition in robotic surgery is to create universal models capable of real-time tissue segmentation. These models would not be constrained by the type of surgical hardware used, mirroring the versatility seen in ECG and EEG applications[22][26][28]. The goal is to transcend the limitations imposed by the scarcity of open-source surgical data and the niche nature of surgical video data, moving towards the creation of AI models that are not only robust but also possess a high degree of accuracy due to their ability to learn from a diverse array of data sources. Therefore, the focus shifts from merely gathering a vast corpus of

surgical data, which is both time-consuming and costly, to the development of sophisticated AI models, and questioning the foundations for after acquiring such data to then. These models are designed to accurately segment anatomy in real-time across various surgical disciplines, heralding a new era in the field of robotic surgery where universal applicability, irrespective of the specific surgical procedure or hardware employed, becomes the norm. This approach not only promises to enhance the precision and efficiency of surgical care but also paves the way for broader application and innovation in the realm of robotic surgery and applied AI[27][43][49].

The project will include a systematic literature review for fundamental artificial intelligent practices in computer vision applied for endoscopic anatomy in both robotic and endoscopic surgery. How training, testing, and deploying artificial intelligent models with a lack of present surgical data will be conducted. Then applying, how the model can apply consistent real-time feedback. A case study will be conducted to build and compare the performance of such developed model between both robotic and general endoscopic data from Imperial College's Open Source Endoscopic Surgical Video data https://hamlyn.doc.ic.ac.uk//vision/ to create a new form of model which abides by the following principles: universal application, real time segmentation of anatomy, present information of the given segmented anatomy with a higher accuracy threshold.

Chapter 1

Introduction

1.1 Motivation and Objectives

Real time training, assessing, and diagnosing are essential for robotic surgery, but novel equipment presents challenges. In robotic surgery, surgeons must become fully adept in psychomotor skills and identifying endoscopic tissue in real time, requiring a year-long endeavor for surgical training. Meanwhile for complex surgical tasks in cardiac and neurosurgery, the current field of robotic surgery to enhance minimally invasive surgical care is limited by the hardware size, complexity of trocars, and tools with their respected angles, requiring different forms of hardware to train, assess, and diagnose patients, such as measuring electrical signals [10] [11] [15] [19]. Robotic surgery applied to cross domains of surgery, much like in the case of how measuring electrical signals (via ECG and EEG) is applied to multiple domains, is a highly underrated endeavor, largely due to constrained applications of the hardware robotic surgery applied for gastrointestinal and urology, with intense capital, time, and legal requirements to manufacture, distribute, and scale, intricate equipment[98][3][4]. The motivation for such project is to apply advanced software practices of artificial intelligence (AI), computer vision (CV), multi-modal classification rooted in questioning semi-supervised learning, foundational basis of deep learning, to help mitigate the present and consistent problem of limited available data to train novel models for the purpose of medical application [89][92][69].

1.2 Observations

To apply an example of universal hardware, is the measurement of electrical signals using ECG and EEG for both cardiac and brain interfacing to scan and classify different means of diagnosing patients. Both ECG and EEG provide a non-invasive diagnosis to record electrical activity classifications of waveforms and use cases of machine learning and deep learning are widely used for classification of diseases, emotions, and actions from users[83]. The hardware required for both equipment contain both electrodes, with a difference in interface for acquiring such data where ECG uses a printer, and EEG requires a computer system to display the electrical activity in real time, requiring specialized software for signal processing. Present research for ECG focuses on classifying the different patterns of peaks a patient would experience during an EEG scan to make predictions and classifications of heart rhythms and neuro-disease categorization [117][23][108]. Present research for analyzing electrode signals comes from the premise of applying deep neural networks to create predictions, and there is a vast amount of open-source data available. However, there is also the subject of bias and controlled bias with such studies and abide by a big majority of self-supervising principles for classification, training, and testing sets to normalize the notion of such practices, prior to applying logistic or backdrop propagation to mitigate loss[74]. There is limited use cases of unsupervised learning, but a paramount shift of direction to the necessary forms of unsupervised learning, yet are limited to hybrid learning models instead of true, unsupervised learning. Such algorithms for classification are widely applied decision trees and support vector machines and applying a confusion matrix[57][17]. However, since the majority of studies are applied in a closed environment versus real time information and segmentation. Though both devices extract the same type of data, their interface is different with ECG still relying on paper outputs versus EEG requiring amplifiers and a computer interface for the signaling.

Even though both machines operate with different interfaces, a keen investigation would be applying universal multimodal classification, which is hardware independence, with the goal of predicting the following, based on the peaks for potential adverse disease. For ECG, the direction of research would be aptitude to focus on predicting the probability of heart attacks

based on extracted data from cortisol levels obtained through ECG data . Meanwhile, though there has been vast work for multimodal classification of EEG signals, a keen avenue of research is focusing on probabilities for disease based on real time classification of signal[66][101][46]. For EEG data, predicting probabilities of stroke, Parkinson's, or dementia are the three parameters to investigate when extracting the levels of cortisol based on EEG data[85]. The case study which will be conducted on one ECG machine and one EEG machine for multimodal classification, after the architecture and accuracy of the models are developed, would determine the probability of events to aid in proper diagnosis, care, and plan of action. With the advent of deep learning-based networks, it is possible to design and develop the classification model based on local features along with spatial and temporal context of the physiological signals. However, when applying to real time practice, most studies are limited based on the scarcity of open-source medical data, and conforming practices[50][30][33]. From a specific case point, the foundations of deep learning are built on the premise of supervised learning . The issue is with the fundamentals of supervised learning, it's a requirement to deal with a certain amount of data to then make a prediction from next input. However, when applied to surgical data, there is only a finite amount of information available to train any given model for a specific task, and was suggested, based on present research, unsupervised learning, from a long standpoint, will supersede supervised learning[108][13][57][59]. As a result of the lack of surgical data to aid in prediction for real time classification, regardless of profession the fundamental basis of supervised learning and principles of backdrop propagation must be revisited to create universal models applied regardless of lack of data and hardware. Instead of treating on facet of information or specific domain, it is necessary for a holistic side of care to understand.

1.3 Aims & Objectives

From the following paper, a systematic literature review will be conducted will focus on the domains of robotic surgery, applied engineering principles of software engineering, with a foundation of combinatorics and theories applied for multimodal classification and hardware independence to push for unsupervised learning. Then focusing on building the universal model

from a local machine with the following specs (Intel i9, Titan NVIDIA RTX, 128 GB DDR5 RAM), using Intel's DeepLearning OpenVino Workbench for model creation and deployment, and then outlining the developed model architectures extended in Pytorch libraries and merging with C++ integration with 3 medical devices with the premise of real time classification (Intuitive Surgical's DaVinci Console, ECG Machine, and EEG Machine). After development, two case studies will be conducted to apply the universal models, and how they can be applied for real time classification and segmentation of incoming data will be conducted. For building the model The following paper, there will be 4 paradigms tackled:

1. creation of universal models applied for robotic surgery for real time tissue segmentation, applied in the DaVinci console with necessary benchwork case study on classifying tissue.

2. universal models integrated with electrode data for real time diagnosing and classification in cardiology and neurology, in ECG and EEG machines.

3. Ethics, legality, and patenting for methodology of advanced software. Full scale intricacies of development, deployment, distribution, scale, and maintenance of universal models.

4. For the goal is to bring about the universal application of AI models, to procure high accuracy outputs, with being universal regardless of hardware for robotic and general surgery.

1.4 Deliverables

The scope of this project encompasses a comprehensive exploration into advanced deep learning techniques for analyzing peripheral nerve tissue. This effort is driven by the necessity to enhance our understanding and capabilities in medical imaging, particularly in distinguishing between damaged and undamaged nerves. By leveraging a combination of annotated datasets, cutting-edge neural network architectures, and embedded systems integration, this research aims to push the boundaries of current methodologies. The deliverables outlined in this paper are carefully structured to not only address the immediate challenges in nerve tissue analysis

but also to contribute to the broader field of medical AI applications. This work is anticipated to offer valuable insights and tools for both the scientific community and clinical practitioners, ultimately leading to improved outcomes in surgical procedures and patient care.

The following paper, there will be 4 paradigms tackled:

1. Public annotated dataset of peripheral nerve tissue for both damaged and undamaged nerves from American Association of Neurological Surgeons and the Hamlyn Endoscopic Video Dataset.

2. Development and comparison of different deep learning model architectures with a CNN, 1-D CNN, and Vision Transformer Architecture all developed with input layers of the ReLU activation function and noise reduction using the Adam optimizer with the following annotated dataset

3. Comparison study of developed models with non-pre-trained models with Intel's Open Vino Deep Learning Workbench with the same trained dataset

4. Embedded Systems Service Oriented Architecture system design to integrate with custom deep learning models for the DaVinci surgical system via ROS.

The culmination of these deliverables represents a significant stride towards revolutionizing the application of AI in medical imaging. By developing a publicly accessible dataset, advancing the state-of-the-art in deep learning architectures, and integrating these models within real-world embedded systems, this project lays a robust foundation for future research and practical implementations [118][2]. The comparative analysis against pre-trained models underscores the efficacy of custom-trained architectures, while the integration with surgical systems like the DaVinci demonstrates the practical applicability of this research. Ultimately, these contributions are poised to drive innovation in medical technology, fostering a new era of precision and efficiency in the diagnosis[77]

Chapter 2

Systematic Literature Review

2.1 Identification

The need for a systematic literature review stems from the question of how universal multi-modal classification can be applied to aid, assess, and diagnose patients to enhance surgical care, regardless of hardware. For robotic surgery, with extensive manufacturing, the hardware must adapt to access the patient for minimally invasive surgery. Present ECG and EEG utilize the same form of data extraction, and apply modes of classification, but present research has only focused on holistically specific natures and observations of the organs such as heart rhythm and brain moods[51][53][54]. Yet there still needs to be a focus on extracting specific diseases or conditions to properly assess and diagnose. However, given the vital risk and nature of dealing with such organs, with in place protocols, prior to any surgery, there is a drawback on implementing novel forms of integration advanced hardware and software, unless deployed for education and training purposes [80][55][56]. The review is comprised of screening and data extraction of 43,468 articles from both medical and engineering databases, accompanying a systematic review, reporting with a provided synthesized summary of all studies, finally with a PRISMA checklist reporting to interpret the results and context of specific objectives and aim for future directions of research.

2.2 Screening & Data Extraction

To create a robust literature review, a deep understanding of modern approaches meshed with foundational basis of historical context, will provide insight for future directions of development. Focusing on the question of how to apply multimodal classification regardless of hardware and minimal data. The following systematic review conducted focuses on a modern first approach of novel software and hardware practices for the purposes of multimodal classification, for hardware independence, and a historical overview to understand the further fundamentals of combinatorics with discrete mathematics required for accurate real-time segmentation. Then seeing alternatives against the fundamental basis of deep learning, relying semi supervised learning to work towards unsupervised learning and autonomous integration.There separate literature reviews were conducted, encompassing the total articles with different search terms focusing in domains of medicine, engineering, and combinatorics. For the purposes of modern hardware and software practices, the systematic review comprises recent journal publications, within the previous decade of January 2014-January 2024. The literature review conducted screened for the following key terms with added context for “robotic surgery” + “ECG” + “EEG” + “robotic-assisted surgery” (”multimodal classification” OR ”deep learning” OR ”artificial intelligence” OR ”graph theory” OR ”modalities” OR ”categorization” OR ”segmentation” OR ”real-time feedback” OR ”real-time diagnosis” OR ”real-time segmentation” OR ”real-time high accuracy” OR ”high accuracy model” OR ”computer vision” OR ”machine learning” OR ”general artificial intelligence” OR ”classification” OR ”real-time classification”) with interchangeable (“hardware independence” OR “domain adaptation”) in IEEExplore, ACM Digital Library, Nature Biomedical, Nature Biotech, Nature Cardiovascular Research, Nature Imaging, Nature Machine Intelligence, Nature Medicine, Nature Robotics, Nature Unconventional Computing, Science Advances, Science Robotics, Science Translational Medicine, and Web of Science. Then for a foundational understanding between computer engineering principles to apply universal multimodal classification with hardware independence the following terms included were (PyTorch architecture” OR “PyTorch”) AND (”self-supervised learning” OR ”unsupervised learning” OR ”classification” OR ”segmentation” OR ”backdrop

propagation" OR "local and cloud based model configuration" OR "chain rule" OR "alternatives to supervised learning" OR "Intel DeepLearning Workbench" OR "computer vision applications" OR "computer vision artificial intelligence" OR "combinatorics for segmentation" OR "combinatorics for real-time classification" OR "EEG" OR "ECG" OR "robotic surgery") within the same journals. Deep learning is built on two paradigms of supervised learning and backdrop propagation, an extension of the chain rule, with the crux of supervised learning requiring vests amount of data to be able to train a model, forget the principle and cases of domains where there is very little data available. To investigate alternatives and specializing in unsupervised learning, the third systematic literature review focused on the key terms within the 10 year most recent span, January 2014-January 2024 of discrete mathematics and combinatorics to gain the further understanding principles to construct the logic of the universal model's architecture with the present problems of 2D dimensional data, how data extraction can be made to predict future outcomes. Since a wide plethora of research within the field of deep learning and its applications focus on supervised learning, the direction for advanced artificial intelligence to becomes fully autonomous with high output predictions requires a deeper understanding of 1. dealing with a lack of data required for any pre training of models, and further understanding of methodology to obtain the goal of unsupervised learning, or foundational basis to categorize the unknown in real time to aid in training, diagnosing, and assessing. The following terms for this part of the literature review include the terms ("combinatorics unsupervised learning" OR "autonomous segmentation" OR "segmentation and graph theory" OR "modalities and categorization" OR "alternatives to backdrop propagation" OR "self-supervised learning").

2.3 Systematic Review

2.3.1 Overview

The systematic review draws from three distinct but interrelated perspectives, focusing on the intersection of modern hardware and software practices in the domain of multimodal

classification. This research spans several critical areas, including the application of computer engineering principles to enable hardware-independent multimodal classification, with a particular emphasis on unsupervised learning techniques. The target domains for this analysis include robotic surgical devices, electroencephalography (EEG), and electrocardiography (ECG), where the complexity of medical data necessitates sophisticated approaches to classification and analysis. The systematic review process involved the identification and screening of thousands of research articles across numerous databases with the focus was on filtering relevant studies that directly contributed to the research goals, particularly those dealing with anatomy classification, real-time segmentation, and multimodal or domain-specific applications. The studies included in the systematic review provide a detailed landscape of the current state of research, identifying trends, challenges, and opportunities for future exploration.

2.3.2 Approach

The systematic review employs a three-tier search strategy across a wide array of databases, ensuring comprehensive coverage of the relevant literature. The search strategy was designed to capture the breadth and depth of research in multimodal classification, with a particular focus on studies that align with the objectives of the analysis. The primary criteria for including studies in this systematic review were their relevance to anatomy classification, unsupervised learning, real-time segmentation, multimodal applications, or domain-specific adaptations. Studies that met these criteria were considered for detailed review and synthesis. Additional considerations included the quality of the research as determined by methodological rigor and the impact of the study, as evidenced by citation metrics and relevance to current research trends. A significant number of studies were excluded from the analysis. The most common reasons for exclusion included a lack of focus on anatomy classification, emphasis on supervised rather than unsupervised learning, or an absence of multimodal or domain-specific applications. Studies that were narrowly focused on single modalities or that did not demonstrate a clear link to the objectives of the analysis were also excluded. The search process was systematic and comprehensive, covering a total of 14 databases in the first review, 5 databases in the second

review, and 3 databases in the third review. Each review involved the identification of relevant studies, removal of duplicates, and a detailed screening process to ensure that only the most relevant studies were included in the final analysis.

2.3.3 Analysis

The analysis is organized around several key themes, each representing a critical area of focus within the broader field of multimodal classification. The themes were identified based on commonalities observed across the included studies and are further divided into sub-themes to provide a granular understanding of the research landscape, and include the application of artificial intelligence in robotic surgery, the use of AI in EEG and ECG data analysis, and the challenges and opportunities presented by multimodal classification in medical imaging. Within each theme, the analysis explores the methods and approaches used by researchers, the challenges they faced, and the solutions proposed. One of the key findings of this systematic review is the identification of common challenges across the different domains. For instance, the application of unsupervised learning in robotic surgery and EEG/ECG data analysis is often hampered by the complexity and variability of medical data [6][7][8]. The analysis discusses potential solutions, including the use of more advanced machine learning techniques and the integration of domain-specific knowledge to improve the accuracy and reliability of classification models. In the context of robotic surgery, there is a growing body of research focused on the use of AI for skills assessment, which could revolutionize surgical training and practice [16][17][23]. Similarly, in the field of EEG and ECG analysis, there is an upward trend in the application of AI, although much of the research remains focused on supervised learning [25][29][30]. This suggests a need for more research into unsupervised methods that can handle the sparse and complex nature of medical data. The quantitative aspect of the analysis is equally important. For the first systematic review, out of 43,468 identified articles, 37,079 were deemed relevant to the medical domains of interest, focusing on unsupervised learning in robotic surgery, EEG, and ECG research. After a rigorous screening process, 9,314 articles were thoroughly evaluated, and 120 were accepted for detailed review. The second review started with 1,039 articles, of

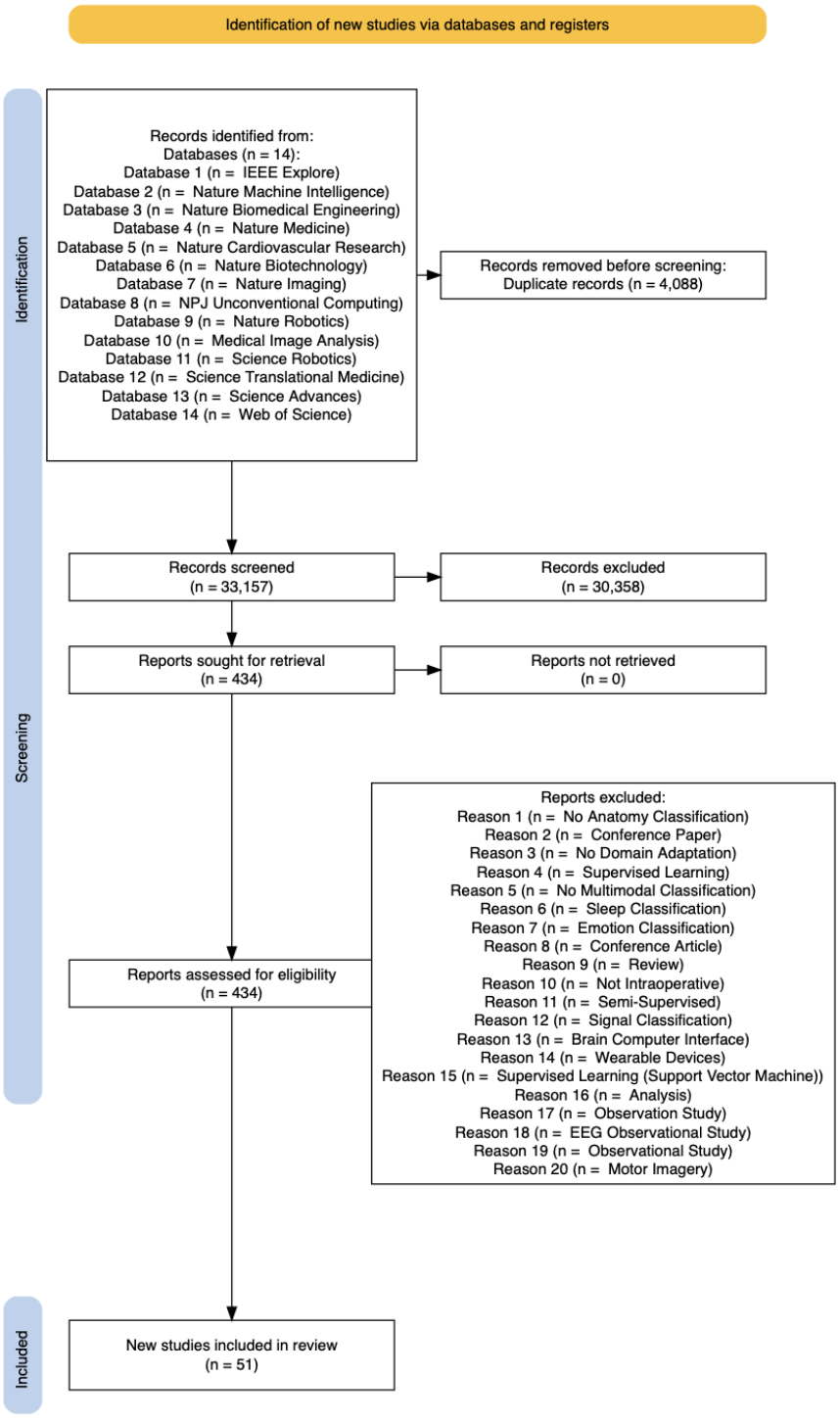


Figure 2.1: Meta Analysis from systematic Literature 1-Medicine

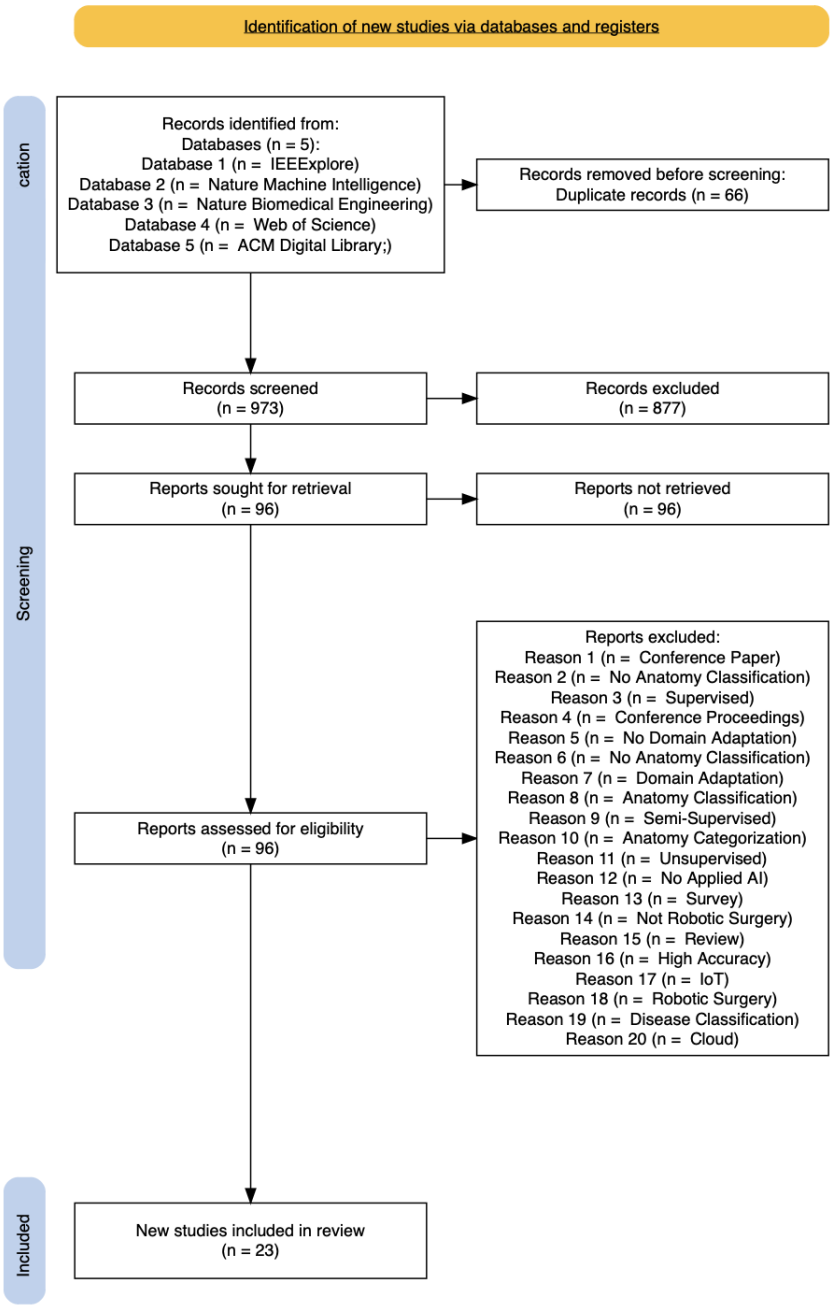


Figure 2.2: Meta Analysis from systematic Literature 1-Engineering

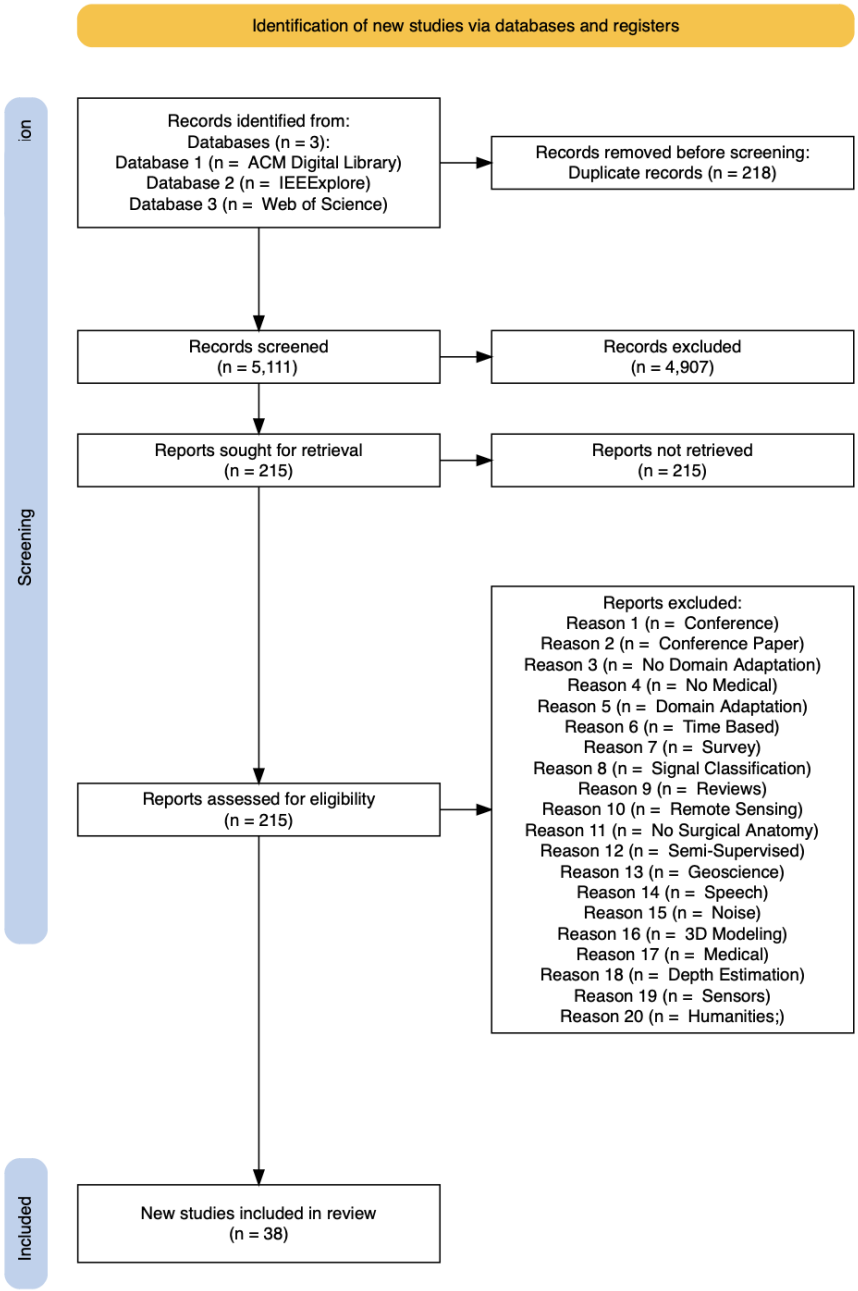


Figure 2.3: Meta Analysis from systematic Literature 1-Unsupervised Learning

which 76 met the inclusion criteria. The third review began with 5,329 articles, leading to the acceptance of 115 studies after screening.

2.4 Reporting

The reporting section synthesizes the findings from the three systematic literature reviews, offering a comprehensive view of the current state of multimodal classification research across medicine, engineering, and mathematics. This interdisciplinary approach provides a rich source of insights that can inform future research and development in the field. The holistic examination of the research reveals three main perspectives: the medical application of multimodal classification, the engineering challenges and solutions, and the mathematical models underpinning the algorithms used in classification [34][36][38][40][42]. The integration of these perspectives is crucial for advancing the state of the art in multimodal classification, particularly in the context of limited data availability and the need for hardware-independent solutions [31][32][33]. Across the three reviews, a total of 9314 articles were screened in the first review, leading to the acceptance of 120 studies. The second review screened 1039 articles, with 76 accepted for review. The third review involved 5329 articles, of which 115 were included in the final analysis. These accepted studies represent the cutting edge of research in multimodal classification, offering valuable insights into the current trends, challenges, and opportunities in the field.The findings of this systematic review have significant implications for future research. The identification of gaps in the literature, particularly the need for more research into unsupervised learning methods, and developing robust deep learning architectures, points to important areas for further exploration[59][60][61][64][119][68]. Additionally, the emphasis on hardware-independent solutions and the integration of multimodal data across different domains suggests a need for more interdisciplinary research that can bridge the gap between theory and practice[71][74][75][76][77].

Chapter 3

Methodology

3.1 Experimental Setup

The experimental setup for this study involves a comprehensive workflow designed to develop, compare, and deploy deep learning models for segmenting peripheral nerve tissue. The process begins with the annotation of data using Intel's Computer Vision Annotation Tool (CVAT), to then model develpment within Jupyter Notebook for initial model development and analysis. The workflow then leverages Intel OpenVINO Deep Learning Workbench for optimizing the models. Three models are developed focusing on their aptitude for classifying components of anatomy. Models are compared in terms of their benchmarking performance, specifically focusing on metrics like F1 score, accuracy, and confusion matrix.Following the model optimization, the workflow includes a pipeline for selecting the model with the highest accuracy. This model is then prepared for deployment using Google Cloud Vertex AI, where a RESTful API endpoint is created for real-time model inference. The deployment process also involves considerations for load balancing, pipeline automation, and error handling to ensure robustness and scalability. Finally, the workflow concludes with documentation and continuous integration/continuous deployment (CI/CD) practices, ensuring that the developed model is efficiently integrated and maintained within the production environment [83][84][85].

1. Public annotated dataset of peripheral nerve tissue for both damaged and undamaged nerves from American Association of Neurological Surgeons and the Hamlyn Endoscopic Video Dataset.

2. Development and comparison of different deep learning model architectures with a CNN, 1-D CNN, and Vision Transformer Architecture all developed with input layers of the sigmoid activation function and noise reduction using the Adam optimizer with the following annotated dataset

3. Comparison study of developed models with non-pre-trained models with Intel's Open Vino Deep Learning Workbench with the same trained dataset

4. Embedded Systems Service Oriented Architecture system design to integrate with custom deep learning models for the DaVinci surgical system via ROS.

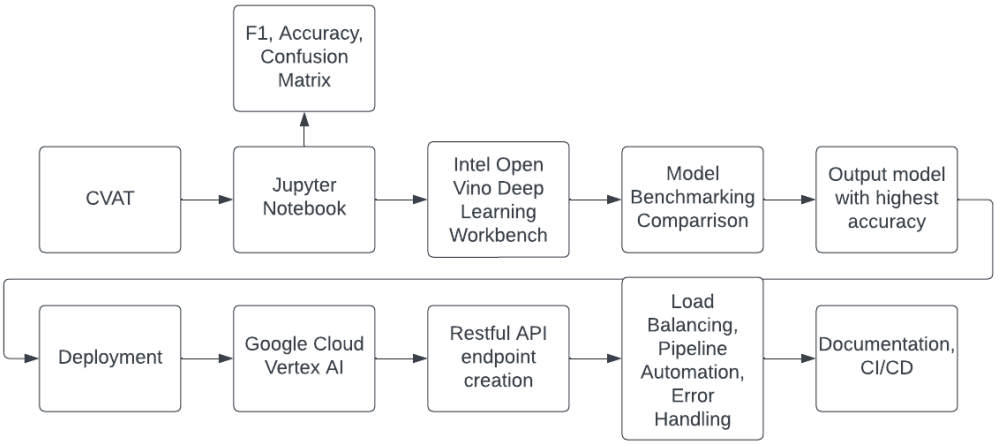


Figure 3.1: Pipeline for Model Creation and Dataflow

3.2 Data Collection and Pre-processing

The data collection and preprocessing phase for this study involved the careful extraction and segmentation of peripheral nerve tissue from a diverse array of surgical videos. These videos were sourced from open-source platforms, including the American Association of Neurological Surgeons (AANA) and the Hamlyn Endoscopic Video Dataset. The dataset is comprehensive,

spanning various surgical approaches such as robotic, endoscopic, and open surgeries, each containing instances of peripheral nerve tissue, both intact and damaged[92][97][100]. A critical tool in this process is Intel's Computer Vision Annotation Tool (CVAT), which was employed to perform detailed instance segmentation and tracking within the videos. CVAT allows for the creation of high-quality instance segmentation masks, crucial for accurately delineating peripheral nerve tissues amidst the complex anatomical environment. This tool also supports the annotation of videos at a frame rate of 60fps, ensuring that even subtle movements and changes in the tissue structure are captured and labeled precisely. The high frame rate is particularly beneficial in surgical contexts where rapid tissue deformation and instrument interactions occur, necessitating meticulous tracking to maintain annotation accuracy [86][87][89].

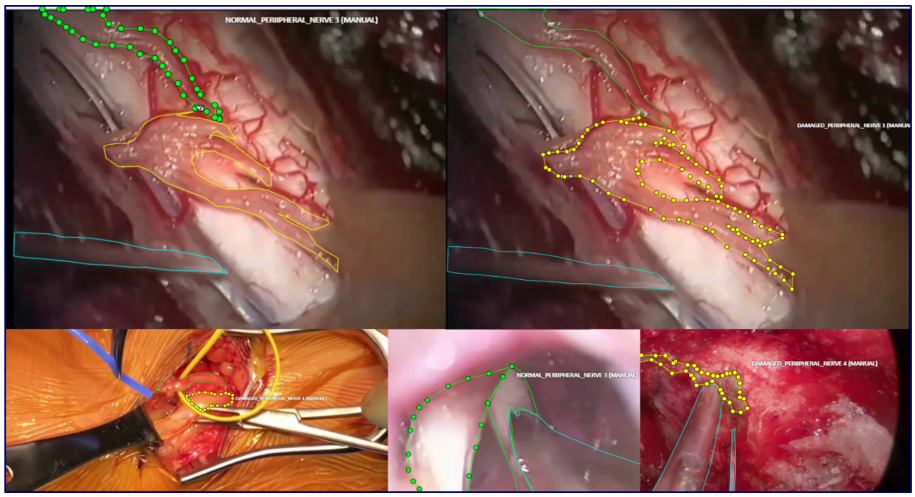


Figure 3.2: CVAT annotation of peripheral nerve tissue from the AANA and Hamlyn Endoscopic Video Dataset

Additionally, the educational surgical videos provided specific instructions for identifying when peripheral nerves were damaged versus undamaged. These instructions were meticulously followed during the annotation process to ensure accurate and consistent labeling. For instance, in the videos, damaged nerves were often indicated by changes in coloration, texture, or structural integrity, such as fraying or swelling, which were carefully marked in CVAT. In contrast, undamaged nerves were noted for their consistent, smooth appearance and uniform color, which were also accurately annotated. These annotations provided critical context for training the machine learning models, as the distinction between damaged and undamaged nerve tissue is essential for developing tools that can assist in real-time surgical decision-making[104][58][73]. The seg-

mentation process currently leverages physical indicators to delineate anatomical structures. While this method has proven effective to a degree, it faces challenges, particularly in differentiating between similar tissue types and identifying minor pathological changes. Consequently, the annotations generated through CVAT, enriched by the precise instructions from the educational videos, provide a robust dataset that can be used to train advanced machine learning models[104][106][114]. These models aim to improve upon traditional segmentation methods by integrating the precise, frame-by-frame annotations produced during the CVAT process. Given the diversity of the dataset, which includes various types of surgical procedures and a range of anatomical variations, this methodology strives to develop a robust model capable of accurately identifying and segmenting peripheral nerve tissue across different contexts[117][1][5]. The ultimate goal is to create a system that can reliably support surgeons by providing real-time, precise anatomical segmentation during surgery, enhancing both the safety and efficacy of surgical interventions.

3.3 Model Development

The development of the segmentation model for this study is centered around leveraging PyTorch libraries to preprocess data and implement various neural network architectures. These architectures aim to accurately segment damaged and undamaged peripheral nerves as well as surgical tools. The comparative study involves three different model architectures, each designed to approach the segmentation task from a unique perspective within the Jupyter Notebook environment. In order to train necessary model architectures, the preprocessing stage is critical and involves the handling of three primary parameters: damaged peripheral nerves, undamaged peripheral nerves, and surgical tools. Using the annotations generated with Intel's CVAT, the data was preprocessed with PyTorch to ensure consistency and quality across the dataset. This included steps such as normalization, data augmentation to enhance model robustness, and careful splitting of data into training, validation, and test sets to maintain the integrity of the evaluation process. To determine the most effective approach for this segmentation task, three different neural network architectures were implemented and compared, all

using PyTorch libraries.

1. 1. Standard Convolutional Neural Network (CNN) with U-Net: The U-Net architecture is well-suited for medical image segmentation, thanks to its encoder-decoder structure that captures both context and fine details necessary for accurate segmentation. U-Net is particularly effective in scenarios requiring precise boundary delineation, such as differentiating between damaged and undamaged nerve tissue [9][12][14].

2. 2. 1D Convolutional Neural Network (1D CNN): A 1D CNN was explored for its utility in processing sequential data, such as time-series signals extracted from video data. Although not traditionally used for image segmentation, 1D CNNs can effectively analyze patterns over time, aiding in identifying transitions between damaged and undamaged tissues or tracking surgical tools.

3. 3. Vision Transformer (ViT) Model Architecture: Vision Transformers (ViTs) are known for their ability to capture global dependencies across an image. Given the complex interactions between nerve tissues and surgical tools, a ViT model was included to enhance segmentation accuracy, especially for capturing long-range dependencies and structural details. maintenance of universal models [18][37][41].

In terms of activation function and optimizer, the ReLU activation function is widely used in neural networks due to its ability to introduce non-linearity while maintaining computational efficiency. For anatomy segmentation, ReLU helps in learning complex patterns in the data by enabling effective gradient propagation during training, essential for capturing intricate anatomical details [94][98][101]. Then, the Adam optimizer was chosen for its adaptive learning rate and efficiency in handling sparse gradients. It combines the benefits of AdaGrad and RMSProp, making it particularly effective for medical image segmentation tasks, where sensitivity to subtle changes in the data, such as those between damaged and undamaged tissues, is critical. Adam's fast convergence helps achieve high performance with fewer iterations, making it ideal for the complex dataset used in this study [90][91][93]. Each model was evaluated based

on its ability to accurately segment damaged and undamaged nerve tissues and identify surgical tools within the diverse dataset. Key performance metrics, including a confusion matrix, Intersection over Union (IoU), F1, training, and validation loss scores were used to compare the effectiveness of each architecture [67][62][66]. The results of this comparative study will guide the selection of the most effective model architecture, which will be further refined for real-time surgical applications, ultimately supporting surgeons by providing accurate, real-time segmentation during procedures.

3.3.1 Convolution Neural Networks for Anatomy Segmentation

Convolutional Neural Networks (CNNs) are a cornerstone of modern image processing, particularly in the domain of medical imaging and anatomy segmentation [102][105][81]. At the core of CNNs is their ability to automatically learn and extract hierarchical features from raw image data, which is crucial for tasks like identifying and segmenting anatomical structures [107][109][112]. A CNN typically comprises multiple layers, starting with convolutional layers that apply a series of filters to the input image. These filters, or kernels, are designed to detect specific features such as edges, textures, and patterns, which are fundamental to understanding the content of the image. As the input image passes through successive convolutional layers, the network progressively captures more complex and abstract representations of the data, enabling it to distinguish between different anatomical structures, such as nerves, muscles, or tissues. Pooling layers, often interspersed between convolutional layers, play a critical role in reducing the spatial dimensions of the feature maps generated by the convolutions. This reduction process, known as downsampling, helps in managing the computational complexity and prevents the network from overfitting. By focusing on the most salient features, pooling layers ensure that the CNN retains essential information while discarding less important details, which is particularly useful in anatomy segmentation where precise localization and identification of structures are necessary. The final layers of a CNN typically consist of fully connected layers that take the high-level features extracted by the previous layers and make predictions or classifications. In the context of anatomy segmentation, these predictions might

involve labeling each pixel in the image according to its corresponding anatomical structure, such as differentiating between damaged and undamaged tissues or identifying the presence of surgical tools.

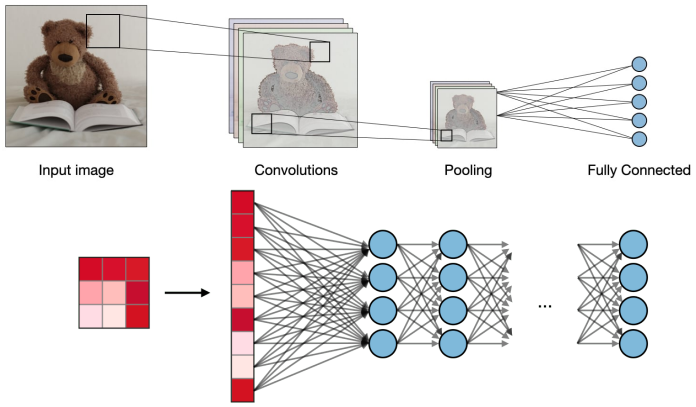


Figure 3.3: A fully connected Convolution Neural Network

One of the fundamental advantages of CNNs in anatomy segmentation is their ability to handle the variability inherent in medical images. Variations in lighting, orientation, tissue appearance, and even the presence of noise can complicate image analysis, but CNNs are robust to these changes due to their hierarchical feature learning process. Additionally, CNNs can be trained on large datasets, enabling them to generalize well to new, unseen data, which is crucial in medical applications where accuracy and reliability are paramount. In summary, the fundamental structure of CNNs—comprising convolutional, pooling, and fully connected layers—makes them exceptionally powerful for the task of anatomy segmentation. By automatically learning from raw data and progressively capturing complex features, CNNs enable precise and detailed segmentation of medical images, facilitating better visualization and analysis of anatomical structures. This capability has made CNNs an indispensable tool in the development of advanced medical imaging technologies, supporting a wide range of applications from diagnostic imaging to surgical planning.

3.3.2 Activation Function and Optimizer

In the design and training of neural networks for anatomy segmentation, the choice of activation function and optimizer plays a pivotal role in determining the performance and efficiency of

the model. The Rectified Linear Unit (ReLU) activation function and the Adam optimizer are widely adopted in this domain due to their robustness and effectiveness.

The ReLU activation function, as shown in Figure 3.5, is defined as $R(x) = x$ if $x \geq 0$, and $R(x) = 0$ otherwise. This simple yet powerful non-linearity introduces sparsity into the network by setting all negative input values to zero, which helps in mitigating the vanishing gradient problem—a common issue in deep networks where gradients diminish as they propagate backward through the layers. By allowing only positive inputs to pass through, ReLU ensures that the network remains computationally efficient and capable of learning complex features from the data. This is particularly important in anatomy segmentation, where the model needs to accurately capture fine details and subtle variations in medical images, such as differentiating between damaged and undamaged tissues or identifying the boundaries of anatomical structures.

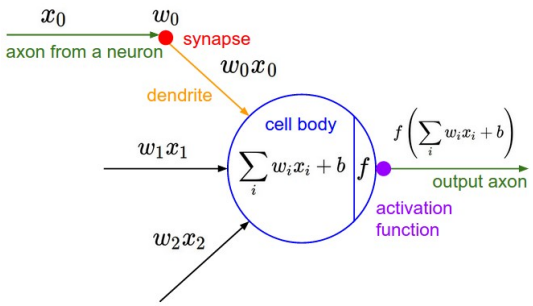


Figure 3.4: Activation function for a following neuron in a neural network

Complementing the ReLU activation function is the Adam optimizer, depicted in Figure 3.6, which stands for Adaptive Moment Estimation. Adam is an advanced gradient-based optimization algorithm that combines the benefits of both the AdaGrad and RMSProp algorithms. It computes adaptive learning rates for each parameter by keeping track of the first and second moments of the gradients. This allows Adam to adaptively adjust the learning rate during training, which accelerates convergence and improves performance, especially in models dealing with high-dimensional data like medical images. The equation governing the Adam optimizer, $\theta = \theta - \alpha \frac{m_t}{\sqrt{v_t}+\epsilon}$, highlights its mechanism of adjusting the learning rate based on the running averages of the gradients (m_t) and their squared values (v_t), ensuring stable and efficient learning.

$$R(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

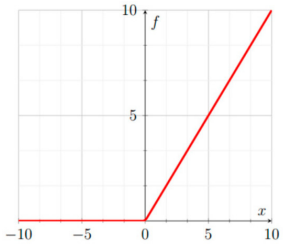


Figure 3.5: The RelU Activation Function

In the context of anatomy segmentation, the combination of ReLU and Adam is particularly effective. ReLU enables the network to model complex, non-linear relationships in the data, which is essential for segmenting intricate anatomical structures. Meanwhile, Adam's adaptive learning rate ensures that the model can quickly converge to a high-performing solution without requiring extensive hyperparameter tuning. This synergy between the ReLU activation function and the Adam optimizer allows for the development of robust, high-accuracy models that can handle the diverse challenges presented by medical image segmentation, ultimately supporting better clinical outcomes.

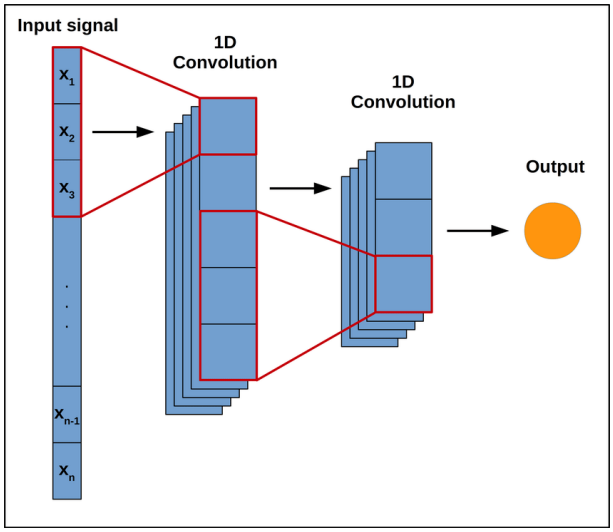
$$\theta = \theta - \alpha \frac{m_t}{\sqrt{v_t + \epsilon}}$$

Figure 3.6: Adam Optimizer

3.3.3 Different Dimensions for CNNs

Convolutional Neural Networks (CNNs) can be designed to operate across different dimensions, depending on the nature of the data being processed. Traditionally, CNNs are employed in two-dimensional (2D) space for image data, where they excel at capturing spatial hierarchies through convolutional operations across the width and height of an image. However, CNNs can also be extended into one-dimensional (1D) or three-dimensional (3D) spaces to accommodate various types of data. For instance, 1D CNNs are particularly well-suited for processing sequential data or time series, where the convolutional operations are performed along a single axis, typically time. In the context of surgical anatomy, 1D CNNs can be leveraged to analyze time series data that reflects the dynamics of surgical procedures. This might include monitoring the sequential

movement of surgical tools, changes in tissue properties over time, or even the progression of a surgical operation. The equation illustrated in Figure 3.7 demonstrates the operation of a 1D convolution, where the output at each step is calculated by applying a weighted sum over a sequence of input values. By sliding the convolutional filter across the time dimension, the 1D CNN captures temporal patterns and trends that are crucial for understanding the flow and impact of surgical interventions.



$$\text{out}(N_i, C_{\text{out}_j}) = \text{bias}(C_{\text{out}_j}) + \sum_{k=0}^{C_{in}-1} \text{weight}(C_{\text{out}_j}, k) \star \text{input}(N_i, k)$$

Figure 3.7: 1D Convolution Neural Network

The application of 1D CNNs in surgical anatomy is particularly valuable when analyzing continuous data streams, such as video sequences or sensor readings during surgery. These networks can detect subtle temporal patterns, such as the onset of tissue damage, changes in tool pressure, or variations in surgical techniques. By integrating this temporal information, 1D CNNs enhance the model's ability to make informed predictions about the state of the anatomy over time, offering a dynamic perspective that complements the static analysis provided by 2D CNNs. Moreover, the use of 1D CNNs in conjunction with 2D or 3D CNNs allows for a more comprehensive analysis of surgical anatomy, combining spatial and temporal insights. This multi-dimensional approach provides a richer understanding of the surgical process, enabling better decision-making and potentially leading to improved surgical outcomes. In summary, while 2D CNNs focus on spatial relationships within images, 1D CNNs provide the neces-

sary tools to analyze the temporal sequences crucial for monitoring and understanding surgical procedures in real-time.

3.3.4 Vision Transformers

Vision Transformers (ViTs) represent a significant shift in the field of image processing, offering a novel approach that leverages the power of transformer models, traditionally used in natural language processing, for visual tasks. Unlike Convolutional Neural Networks (CNNs) that rely on convolutional layers to capture local spatial relationships, Vision Transformers operate by dividing an image into fixed-size patches and treating each patch as a sequence element, similar to the tokens in a text sequence. As illustrated in Figure 3.9, these patches, along with their positional embeddings, are fed into a transformer encoder, which processes them through a series of multi-head attention layers and feedforward networks. The key advantage of Vision Transformers lies in their ability to capture long-range dependencies within an image. While CNNs are inherently localized, focusing on nearby pixel relationships, ViTs can learn to correlate information across the entire image, making them particularly effective in scenarios where global context is crucial. In the context of anatomy segmentation, this global perspective allows ViTs to better understand the overall structure and relationships between different anatomical components, which is essential for accurately identifying and segmenting complex structures. The transformer encoder, as depicted on the right side of Figure 3.9, consists of multiple layers

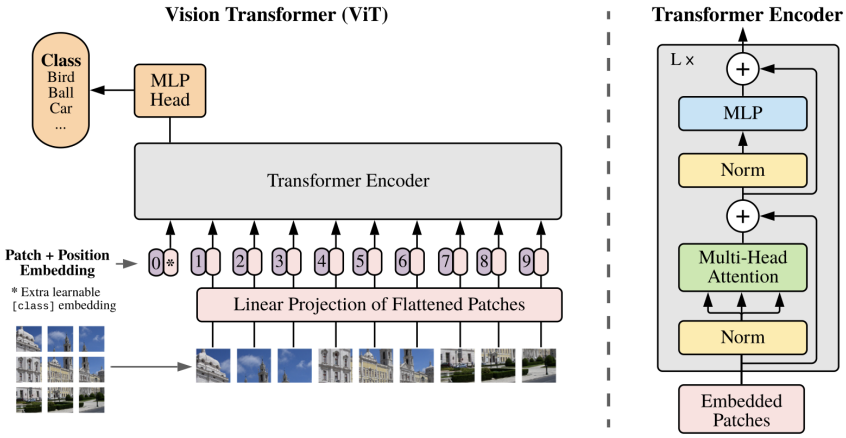


Figure 3.8: Vision Transformer Architecture

of multi-head self-attention mechanisms, which enable the model to weigh the importance of each patch relative to others in the image. This allows the ViT to capture subtle patterns and dependencies that might be overlooked by more localized methods. Furthermore, the use of positional embeddings ensures that the model retains information about the spatial arrangement of the patches, which is crucial for maintaining the integrity of the anatomical structures during segmentation. ViTs are particularly well-suited for tasks involving high-resolution medical images, where understanding the global context can lead to more accurate and reliable segmentation results. By leveraging the transformer architecture, ViTs can effectively model complex relationships within the image, leading to improved performance in tasks like identifying pathological regions, segmenting organs, or distinguishing between different types of tissues. In summary, Vision Transformers provide a powerful alternative to traditional CNNs, offering a global approach to image analysis that is particularly beneficial in the field of anatomy segmentation. Their ability to capture long-range dependencies and global context makes them a valuable tool in the development of advanced medical imaging technologies, potentially leading to more accurate and efficient diagnostic and treatment processes.

3.4 Training, Validation, and Deployment

The process of training, validating, and deploying a model is critical to ensuring its accuracy, generalizability, and effectiveness in real-world applications. In the context of anatomy segmentation, these steps involve rigorous testing and optimization to ensure that the model can reliably differentiate between various anatomical structures and identify pathological changes. The training phase involves feeding the model with labeled data and allowing it to learn the underlying patterns and features necessary for accurate segmentation. During this phase, the model's weights are adjusted iteratively based on the error between the predicted output and the actual labels. For this study, the Intersection over Union (IoU) metric, as depicted in Figure 3.10, was employed to evaluate the model's performance during training. The IoU measures the overlap between the predicted segmentation mask and the ground truth, providing a clear indication of the model's accuracy. A higher IoU indicates better performance, as it signifies

a greater degree of overlap between the prediction and the actual label. Training was conducted using a combination of data augmentation techniques to increase the diversity of the training set, thereby improving the model's ability to generalize to new data. Techniques such as rotation, flipping, and scaling were applied to the images, ensuring that the model could handle variations in the input data. Additionally, the training process was optimized using the Adam optimizer, which adaptively adjusts the learning rate, enabling faster convergence and better handling of sparse gradients. The validation of the model was conducted using

$$\textit{Intersection over Union} \left(\textit{IoU}\right) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \qquad \textit{Intersection over Union} \left(\textit{IoU}\right) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A}| \cup |\mathrm{B}|}$$

Figure 3.9: Intersection over Union

Intel's OpenVINO (Open Visual Inference and Neural Network Optimization) framework, a powerful toolkit designed to optimize and deploy deep learning models for inference on a variety of hardware platforms. OpenVINO was utilized to convert the trained model into an optimized format, enhancing its performance during inference while maintaining accuracy. The framework supports multiple hardware configurations, ensuring that the model can be validated efficiently across different environments. OpenVINO's optimization capabilities include reducing the model's size and increasing its inference speed, which are crucial for real-time applications such as anatomy segmentation in surgical settings. During validation, the model was benchmarked using various performance metrics, including the Intersection over Union (IoU), F1 Score, and Confusion Matrix, as part of the Intel OpenVINO Deep Learning Workbench. This comprehensive validation process ensured that the model could deliver high accuracy and reliability when deployed in a real-world scenario.

The deployment of the validated model was based on a Service-Oriented Architecture (SOA) designed for an embedded system within the DaVinci Research Kit. This architecture, as shown in Figures 3.1 and 5.1, facilitates the seamless integration of the model into a complex surgical environment. The SOA framework enables the model to interact with other components of the DaVinci system through standardized communication protocols, ensuring flexibility, scalability, and reliability. The deployment process involved creating RESTful API endpoints that allow

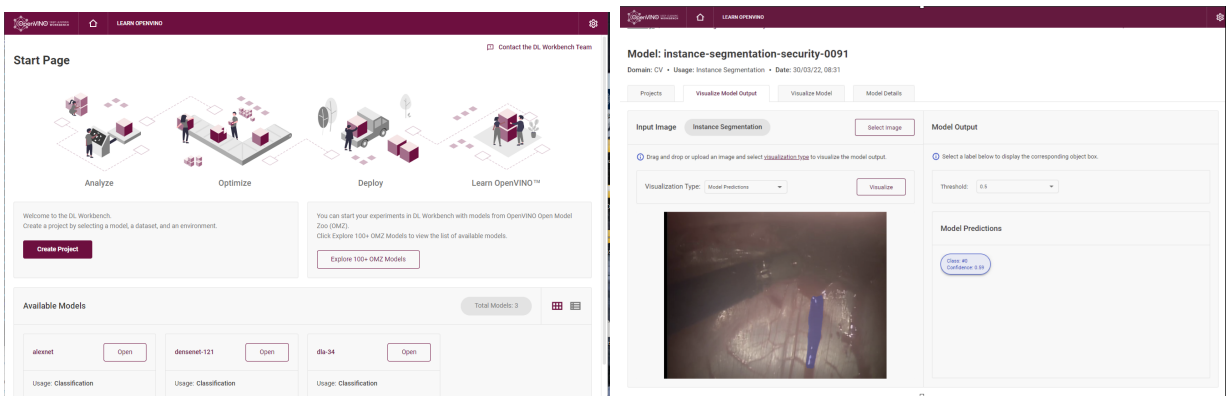


Figure 3.10: Open Vino Deep Learning Workbench with an instance segmentation of bleeding for an in vivo robotic assisted diaphram disection

for real-time interaction between the model and the DaVinci surgical system. The model was integrated into the DaVinci system using Google Cloud's Vertex AI platform, which provided the infrastructure for hosting and managing the model in the cloud. The SOA framework also supports features such as load balancing, pipeline automation, error handling, and continuous integration/continuous deployment (CI/CD), ensuring that the model remains up-to-date and functional throughout its lifecycle. In this context, the deployed model plays a crucial role in providing real-time segmentation of anatomical structures during surgical procedures. It enhances the capabilities of the DaVinci system by offering precise and reliable guidance to surgeons, ultimately improving surgical outcomes. In summary, the training, validation, and deployment phases are integral to the successful development and application of anatomy segmentation models. By carefully optimizing each stage and leveraging advanced tools like Intel's OpenVINO and a Service-Oriented Architecture, the resulting model is not only accurate but also robust, scalable, and ready for real-world surgical applications.

Chapter 4

Results

4.1 Model Performance

The following results focus on the development and application of a customized dataset tailored for domain adaptation within the context of surgical disciplines. Given the unique challenges presented by surgical data, it is imperative to employ deep learning practices that are specifically designed to predict outcomes with high accuracy in such specialized environments. The customized dataset has been meticulously curated to capture the nuances of surgical anatomy, enabling the models to learn from and adapt to the specific characteristics of surgical images, such as varying tissue types, lighting conditions, and the dynamic nature of endoscopic procedures. In particular, domain adaptation plays a critical role in enhancing the performance of deep learning models by allowing them to generalize across different surgical disciplines. This involves adjusting the model to understand and interpret data from various surgical domains—be it general surgery, endoscopy, or robotic surgery—while maintaining a high level of accuracy. The adaptation process ensures that the models are not only trained on a single type of data but are also capable of applying their learned knowledge to new, unseen surgical scenarios. This capability is crucial for real-time applications where the model must be versatile and adaptable to different types of surgeries without requiring extensive retraining. Furthermore, the implementation of deep learning practices tailored for surgical data prediction is essential

to ensure that the models can handle the complex and often unpredictable nature of surgical environments. The models must be capable of accurately classifying anatomical components, detecting anomalies, and providing reliable predictions in real-time. This requires not only sophisticated model architectures but also robust training and validation processes that take into account the variability and intricacies of surgical data. The comparison between these models and five non-pre-trained models, tested on Intel's Open Deep Learning Workbench, underscores the significance of domain adaptation and specialized training. The non-pre-trained models, which lack the customized adaptation to surgical data, often fall short in accuracy and generalization. This comparison highlights the importance of domain-specific training and the necessity of fine-tuning models to the particular demands of surgical data. The results from these comparisons provide valuable insights into the potential of advanced software to classify components of anatomy in real time, regardless of the surgical domain. Moving forward, the focus will be on further refining these models to enhance their performance. This includes improving data annotation processes, refining pre-processing techniques, and optimizing the alignment of models to the specific requirements of surgical tasks. Future work will also explore the integration of these models into real-time surgical systems, ensuring that they can be employed effectively across various surgical specialties, including general surgery, endoscopy, and robotic-assisted procedures.

4.1.1 Results Vision Transformer

The transformer model, widely recognized for its ability to handle sequential data, has emerged as a formidable tool in the domain of medical image classification, particularly in the context of surgical applications. Its performance in this area has been notable, achieving an F1 score that consistently surpassed that of the 1D-CNN across multiple categories. The model's strength lies in its sophisticated ability to capture and process complex patterns and relationships within the data, a capability that becomes particularly advantageous when dealing with intricate anatomical components. Unlike traditional models that may struggle with the nuanced variations present in medical imagery, the transformer model excels at identifying and classifying struc-

tures that are often difficult to discern, especially when these structures overlap or present in complex configurations. Central to the transformer model's success is its self-attention mechanism. This feature allows the model to dynamically focus on the most relevant parts of the input data, effectively filtering out noise and honing in on critical features that contribute to accurate classification. This ability to selectively attend to different parts of the data is particularly beneficial in medical contexts, where the importance of specific image features can vary significantly depending on the anatomical region being analyzed. The robustness of the transformer model is further evidenced by its confusion matrix, which reveals a lower incidence of misclassifications compared to the 1D-CNN, particularly in challenging scenarios. These scenarios often involve similar anatomical features, such as subtle variations in endoscopic camera angles, fluctuating lighting conditions, and the contouring of images during video recordings.

Metric	**Value**
F1 score	0.986
Accuracy	0.992
Intersection Over Union	0.894
Training Loss	0.012
Validation Loss	0.015

Table 4.1: Model Performance Metrics for Vision Transformer

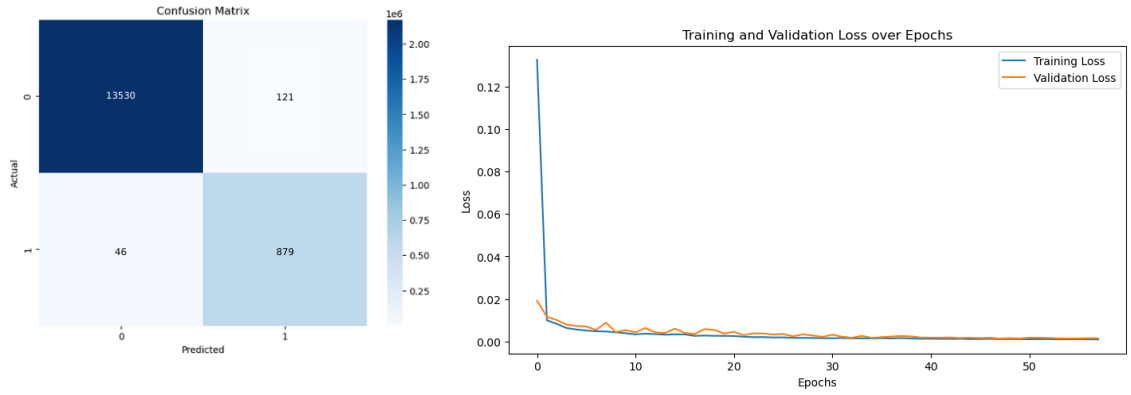


Figure 4.1: Performance of Visual Transformer Architecture

These factors introduce a level of complexity that can confound less sophisticated models, but the transformer's architecture is uniquely equipped to manage these challenges. By effectively navigating the inherent complexities of surgical data, the model not only improves classification accuracy but also enhances the reliability of its predictions in real-time surgical settings.

Backtesting of the transformer model further underscores its generalization capabilities. When applied across various subsets of the dataset, the model maintained high accuracy and F1 scores, demonstrating its versatility and robustness across different surgical disciplines. This consistent performance across diverse datasets indicates that the transformer model is not merely tailored to specific scenarios but is adaptable to a wide range of surgical applications. This makes it a reliable tool for predicting outcomes in various surgical contexts, from routine procedures to more complex interventions. However, the superior performance of the transformer model does come with certain trade-offs. Benchmarking revealed that, although it delivers unparalleled accuracy and classification precision, the model is also more computationally intensive compared to the 1D-CNN. The increased computational demands necessitate careful management, such as implementing early stopping techniques to prevent overfitting and reduce processing time. This trade-off between computational cost and predictive accuracy is a crucial consideration for its deployment in real-time surgical settings. In environments where both speed and precision are critical, the decision to use the transformer model must balance its advanced capabilities with the practical constraints of the operating room. Overall, the transformer model represents a significant advancement in the field of medical image classification, offering a powerful combination of accuracy, adaptability, and precision. Its ability to effectively handle the complexities of surgical data, coupled with its strong generalization across various surgical disciplines, positions it as a valuable tool in enhancing the quality of care in minimally invasive surgeries. As the model continues to evolve, its integration into real-time surgical applications holds the potential to significantly improve patient outcomes, making it an indispensable asset in modern surgical practice.

4.1.2 Results 2D CNN

The standard Convolutional Neural Network (CNN), particularly influenced by the UNET architecture, has established itself as a robust tool in the realm of medical image classification, demonstrating competitive performance across several key metrics. In terms of the F1 score, the CNN achieved results that were highly competitive with those of the transformer model,

especially in cases where the preservation and understanding of spatial relationships within the data were critical. This ability to capture and leverage local features allowed the CNN to effectively classify distinct anatomical components, making it a valuable asset in situations where the clear delineation of structures is essential. When evaluating the accuracy of the CNN, it was observed that the model performed admirably, though its accuracy was slightly lower than that of the transformer model. This difference suggests that while the CNN is proficient at recognizing and classifying various anatomical structures, there may be room for further optimization. Enhancing the model's performance could involve additional tuning or the application of data augmentation techniques, which could help elevate its accuracy to levels comparable with more advanced architectures like transformers. Such improvements could make the CNN even more reliable in medical contexts where precision is paramount.

Metric	**Value**
F1 score	0.843
Accuracy	0.980
Intersection Over Union	0.764
Training Loss	0.038
Validation Loss	0.041

Table 4.2: Model Performance Metrics for 2D Convolutional Neural Network (CNN)

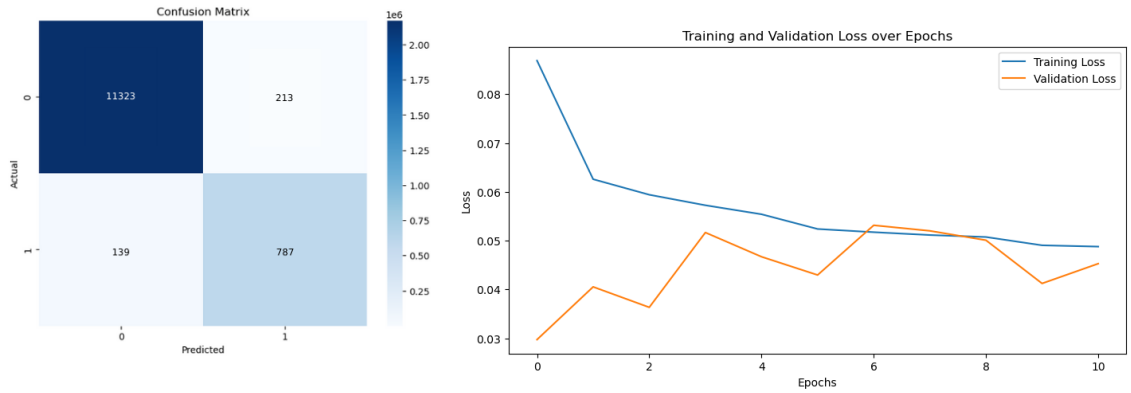


Figure 4.2: Performance of 2D Convolution Neural Network

The confusion matrix for the CNN provided further insights into its performance across different classes. Generally, the CNN performed well, managing to correctly classify most categories. However, there were instances of misclassifications, particularly in cases where anatomical components were visually similar. These misclassifications highlight a potential area for

enhancement, indicating that the model might benefit from integrating additional contextual information. By doing so, the CNN could improve its ability to distinguish between closely related structures, thus reducing the likelihood of errors in classification. Backtesting the CNN revealed its consistency across different datasets, reinforcing its reliability as a model capable of maintaining performance across various surgical disciplines. This consistency is particularly important in real-time applications, where the ability to generalize across different types of data is crucial. Although the CNN's performance in backtesting was strong, it did not quite match the transformer model in more complex scenarios. This discrepancy underscores the CNN's limitations when handling more intricate and overlapping anatomical structures, suggesting that while the CNN is a reliable tool, it may not be the best choice for the most challenging classification tasks without further modifications. In terms of benchmarking, the CNN demonstrated commendable efficiency, particularly in speed. While it did not achieve the same level of accuracy as the transformer model, its lower computational requirements present a significant advantage. This makes the CNN a viable option in environments where processing power is limited or where faster inference times are crucial. The balance between computational cost and classification performance positions the CNN as an attractive choice for deployment in settings where resources are constrained but reliable and quick results are necessary. Comparing the CNN with non-pretrained segmentation models further emphasizes its strengths. The influence of the UNET architecture within the CNN allows it to effectively perform tasks that require precise segmentation and classification, particularly when pretrained models are not available or applicable. This ability to operate effectively without relying on pretrained weights broadens the CNN's applicability across different medical imaging tasks, making it a versatile tool in the arsenal of medical AI technologies.

4.1.3 Results 1D CNN

The 1D Convolutional Neural Network (1D-CNN) exhibited a respectable performance in the classification of surgical anatomy, though it fell short compared to the more advanced transformer model and the standard CNN. The F1 score achieved by the 1D-CNN indicated a

balanced, albeit less robust, performance in terms of precision and recall across the test data. While the model demonstrated an ability to classify surgical anatomy components, its effectiveness varied significantly depending on the specific anatomy in question. This variability highlights both the potential and the limitations of 1D convolutional layers, particularly in their capacity to capture temporal dependencies within the dataset. However, these layers may not be as effective in handling the complex spatial relationships that are better managed by more sophisticated models.In terms of accuracy, the 1D-CNN showed solid results, but it did not reach the high levels observed in the transformer model or the standard CNN. This lower accuracy suggests that while the 1D-CNN is capable of recognizing and predicting anatomical components, it may struggle with more intricate or overlapping structures that require a deeper understanding of spatial context. The model's performance, though commendable in simpler scenarios, indicates that additional tuning, data augmentation, or even architectural enhancements might be necessary to bring its accuracy closer to that of its more advanced counterparts.

Metric	**Value**
F1 score	0.717
Accuracy	0.925
Intersection Over Union	0.654
Training Loss	0.025
Validation Loss	0.275

Table 4.3: Model Performance Metrics for 1D Convolutional Neural Network (CNN)

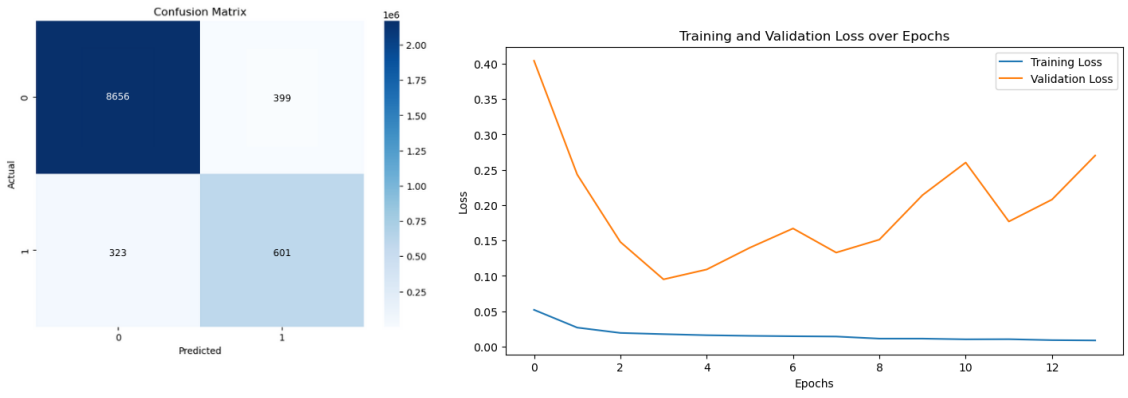


Figure 4.3: Performance of 1D Convolution Neural Network

The confusion matrix for the 1D-CNN provided further evidence of its relative shortcomings.

While the model was generally accurate, it exhibited a higher rate of misclassification in certain classes, particularly those involving anatomically similar structures. These errors underscore a critical limitation of the 1D-CNN: its reduced capacity to distinguish between closely related anatomical features, a task where both the transformer and standard CNN performed significantly better. This finding suggests that further refinement, such as incorporating additional contextual information or improving data preprocessing techniques, could help mitigate these issues, though the model's inherent architectural constraints may still limit its ultimate performance. Backtesting the 1D-CNN against a holdout dataset revealed a degree of robustness, but also confirmed its limitations. While the model maintained reasonable predictive capabilities across different surgical scenarios, it struggled to generalize as effectively as the transformer model and the standard CNN. This indicates that the 1D-CNN, while useful, may not be the best choice for applications requiring high levels of generalization across diverse datasets, particularly those involving complex or varied anatomical features. Benchmarking against other models highlighted the 1D-CNN's efficiency in terms of speed and computational requirements. Its simpler architecture allowed for quicker inference times, making it a viable option in real-time surgical environments where computational resources are limited. However, this advantage in speed comes at the cost of lower accuracy and classification precision. The 1D-CNN's reduced computational demands make it a practical choice for certain scenarios, but its performance trade-offs must be carefully considered, especially in contexts where accuracy is critical. In conclusion, while the 1D-CNN presents a balanced and efficient model for specific tasks, its performance is noticeably inferior to that of the transformer model and the standard CNN. Its ability to handle less complex data with greater speed may be valuable in resource-constrained environments, but its limitations in accuracy and generalization suggest that it is less suited for tasks requiring the highest levels of precision. For real-time surgical applications, where both accuracy and speed are crucial, the 1D-CNN may serve as a supplementary tool, but it is unlikely to replace more advanced models in scenarios that demand the highest standards of performance. The results of this study focus on the application of deep learning techniques to a customized dataset tailored for domain adaptation within various surgical disciplines. The primary objective was to develop and evaluate models capable of accurately predicting and

classifying components of anatomy in real-time, regardless of the specific surgical domain. The study involved the use of multiple deep learning architectures, including a one-dimensional convolution neural network (1D-CNN), a transformer model, and a standard convolution neural network (CNN), each evaluated on key performance metrics including F1 score, accuracy, confusion matrix, backtesting, and benchmarking. Additionally, these models were compared with five non-pre-trained segmentation models using Intel's Open Deep Learning Workbench.

4.2 Comparison with Existing Methods with Intel Open Vino Deep Learning Workbench

We compare the performance of our deep learning models with five non-pre-trained instance segmentation models configured using Intel's OpenVINO Deep Learning Workbench. The instance segmentation models used for this comparison were:

Model Name	Number of Parameters
instance-segmentation-security-1040	13,567,300
instance-segmentation-security-0002	48,373,200
instance-segmentation-security-1039	9,336,200
instance-segmentation-security-0091	63,115,000
instance-segmentation-security-0228	23,598,000

Table 4.4: Intel Vino Open Instance Segmentation Models from Intel's OpenVINO Library

These models were tested to evaluate their performance against the custom-built architectures, including the 1D Convolutional Neural Network (1D-CNN), the standard CNN, and the Vision Transformer. The parameters utilized in the configuration of these models underscore the critical importance of model complexity and parameter tuning in achieving higher accuracy. Each of the custom-built models—1D-CNN, standard CNN, and Vision Transformer—was constructed with a considerable number of parameters, totaling 77,717,506 across the three models. These parameters were distributed across 8 hidden layers, with an output size of 3 and a hidden size of 10,224. This extensive parameterization is a key factor that contributes to the superior performance of these models when compared to the non-pre-trained models. The higher number of parameters enables the models to capture more intricate patterns and relationships within

the data, which is especially crucial in the complex and variable context of surgical anatomy classification. More parameters allow the models to learn more sophisticated features, which in turn leads to better generalization and higher accuracy. In contrast, the non-pre-trained instance segmentation models, while effective in certain scenarios, lack the depth and breadth of parameters necessary to achieve the same level of precision. These models, configured with fewer parameters, are limited in their ability to adapt to the specific nuances of the surgical data, which often involves highly intricate anatomical structures. The comparison clearly illustrates that while these models can serve as a baseline, the more complex architectures, equipped with a larger number of parameters, are far better suited to the demands of real-time surgical applications. Moreover, the inclusion of multiple hidden layers and a large hidden size in the custom-built models plays a pivotal role in enhancing their capacity to process and classify complex anatomical images accurately. The layered architecture allows for the progressive refinement of features, leading to a more nuanced understanding of the input data. This is particularly evident in the Vision Transformer, which leverages its extensive parameter space to focus on relevant aspects of the image data, ultimately resulting in superior classification performance.

4.3 Summary of Results

The results of this study demonstrate the significant potential of advanced software, specifically deep learning models, in classifying anatomical components in real-time across various surgical domains. The transformer model, with its advanced architecture, proved to be the most effective in handling the complexities of surgical data, achieving the highest scores in most performance metrics. However, the 1D-CNN and standard CNN also showed strong performance, particularly in scenarios where computational efficiency and speed were critical. The comparison with non-pretrained models further highlights the advantages of using sophisticated, pretrained architectures for surgical applications. The deep learning models developed in this study not only outperformed these alternatives but also showcased the potential for future improvements through model alignment, data annotation, and preprocessing refinement.

In summary, the comparison between the non-pre-trained models and the more advanced, parameter-rich models highlights the importance of a well-configured architecture. The ability to adjust and optimize the number of parameters, hidden layers, and other critical settings directly correlates with the model's accuracy and effectiveness in real-world applications. As demonstrated by the performance of the 1D-CNN, standard CNN, and Vision Transformer, a greater number of parameters is essential for achieving higher accuracy, especially in complex domains such as surgical image classification. Future work will focus on these areas, aiming to enhance the models' accuracy and generalizability even further. This will involve refining data preprocessing techniques, improving model alignment with surgical data, and exploring new approaches to data annotation to better capture the nuances of different surgical disciplines. The ultimate goal is to develop models that can be seamlessly integrated into both general endoscopic and robotic surgery, providing real-time, accurate classification of anatomical components regardless of the specific surgical context.

Chapter 5

Discussion

5.1 Interpretation of Results

Based on the results obtained from our advanced models, it is clear that the Transformer model architecture significantly outperformed the other models, including the standard CNN and the 1D-CNN, as well as the non-pretrained instance segmentation models from Intel's OpenVINO library. The Transformer model's ability to handle complex patterns and relationships within the surgical data was lead to superior accuracy, F1 scores, and overall classification performance. This architectural advantage was particularly evident in scenarios involving intricate and overlapping anatomical structures, where the other models, including the instance segmentation models, struggled to maintain the same level of precision. The framework outlined below provides a pathway for implementing these models within a customized architecture, ensuring seamless integration with Intel's OpenVINO deep learning workbench. Specifically, the Transformer model's self-attention mechanism allowed it to focus on the most relevant parts of the input data, a feature that was less pronounced in the other architectures. This capability enabled the Transformer model to excel in real-time applications, where the ability to quickly and accurately process incoming data is critical. To achieve optimal deployment, a hybrid system design should be considered, which balances the benefits of both local and cloud-based models. While the Transformer model demands more computational resources, its superior performance

justifies the investment in higher-capacity hardware or cloud infrastructure. In contrast, the non-pretrained instance segmentation models, while effective in certain scenarios, were limited by their relatively simple architectures and fewer parameters. These models, configured with fewer parameters, could not match the depth of analysis provided by the Transformer, particularly in the highly variable and complex context of surgical anatomy classification. Furthermore, the deployment strategy must emphasize the customization of models to fit specific surgical applications. The Transformer model, with its extensive parameterization—totaling 77,717,506 parameters across its architecture—demonstrated that a higher number of parameters directly correlates with improved model performance. This level of detail allowed the Transformer to capture subtle differences between anatomical structures, outperforming both the standard CNN and the 1D-CNN, as well as the instance segmentation models, in accuracy and reliability. Moreover, the inclusion of multiple hidden layers and a large hidden size in the Transformer model plays a pivotal role in enhancing its capacity to process and classify complex anatomical images accurately. The layered architecture allows for the progressive refinement of features, leading to a more nuanced understanding of the input data. This is particularly evident in the Transformer model's ability to generalize across different surgical disciplines, a capability that was less robust in the other models tested. In summary, the results clearly demonstrate that the Transformer model architecture is the most effective solution for real-time surgical applications, outperforming both the custom-built models (standard CNN and 1D-CNN) and the instance segmentation models from Intel's OpenVINO library. The ability to adjust and optimize the number of parameters, hidden layers, and other critical settings in the Transformer architecture directly correlates with the model's accuracy and effectiveness in complex, real-world scenarios.

5.2 Implications for Surgical Training and Medical Practice

The integration of custom deep learning models into robotic-assisted surgery is crucial as the complexity and precision demanded by modern surgical procedures continue to grow. Standard, off-the-shelf models often fall short in the highly specialized environment of surgery, where real-time decision-making and the ability to adapt to unique patient anatomies are critical. Custom models, designed specifically for the intricacies of surgical data, can significantly enhance the functionality and effectiveness of robotic surgical systems like the da Vinci surgical console. One platform that has been central to the development of such models is the da Vinci Research Kit (dVRK), an open-source platform that operates on the Robot Operating System (ROS). The dVRK allows researchers to experiment with new algorithms and custom deep learning models, providing a flexible environment for innovation. However, while the dVRK offers substantial opportunities for research, it also comes with notable limitations that can hinder the transition from research to clinical application.A significant limitation of the dVRK is its reliance on a specific operating system: Ubuntu 12.04. This older version of Ubuntu is no longer supported with updates, which introduces potential security vulnerabilities and compatibility issues with newer software and hardware. The dependency on this outdated operating system restricts the dVRK's ability to incorporate the latest advancements in deep learning frameworks and tools, which are often optimized for newer operating systems. As a result, researchers working with the dVRK must navigate these constraints, potentially limiting the complexity and performance of the models they can develop. In addition to the OS limitations, the hardware constraints of the dVRK also pose challenges. The computational resources available within the dVRK are often less powerful than those in fully integrated surgical systems. This limitation can restrict the size and sophistication of the models that can be tested on the platform. For instance, while the dVRK is an excellent tool for prototyping, the models that perform well in this environment may not scale effectively to the clinical-grade da Vinci surgical console, which requires models to operate efficiently under more demanding conditions. Despite these limitations, the development and comparison of custom models within the dVRK have yielded

valuable insights. In this study, custom models including a Transformer-based architecture, a standard CNN, and a 1D-CNN were developed and rigorously tested within the dVRK environment. The Transformer model, in particular, demonstrated superior performance in handling complex surgical data, outclassing both the standard CNN and 1D-CNN in terms of accuracy, F1 score, and the ability to generalize across different surgical tasks. These models were also compared against five non-pretrained instance segmentation models available from Intel's OpenVINO library, specifically configured for tasks within the dVRK. While these non-pretrained models were effective in some scenarios, they lacked the sophistication and adaptability of the custom models, particularly the Transformer model. The custom models were better suited to the unique demands of robotic surgery, offering higher accuracy and more reliable performance in real-time applications. The limitations of the dVRK and its reliance on Ubuntu 12.04 highlight the need for a more robust and flexible deployment strategy for custom models. To overcome these challenges, transitioning to a service-oriented architecture (SOA) within the da Vinci surgical console could be a solution. In an SOA, custom models can operate as independent services that can be dynamically deployed, updated, and scaled as needed. This would not only mitigate the constraints posed by the dVRK's hardware and software limitations but also enable the deployment of more advanced models like the Transformer architecture in clinical settings. Moreover, the SOA approach would facilitate the integration of multiple models specialized for different tasks within the surgical workflow.

5.3 Service Oriented Architecture System Design for daVinci Research Kit

The Service-Oriented Architecture (SOA) design shown in the diagram represents a highly modular and scalable approach to deploying custom deep learning models within the da Vinci surgical system. The design is segmented into several key areas, each responsible for different aspects of data handling, model deployment, system reliability, and interaction with the ROS-based da Vinci surgical console. Let's explore these components in detail and their interaction

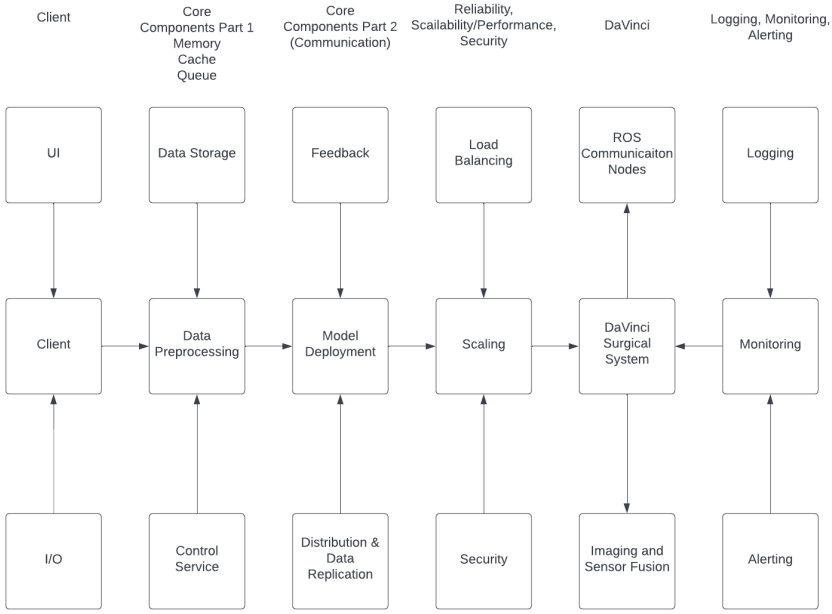


Figure 5.1: Service Oriented Architecture for the deploying on the daVinci Research Kit with the ROS system.

1. **Client Interface and Input/Output (I/O) Handling**

   - **Client:** The client interface is responsible for initiating and controlling the interaction with the surgical system. This could involve user inputs from the surgeon or automated commands based on pre-programmed instructions. Beign domain agnostic, dpending on what UI is implemented, the client can receive real time feedback either form the daVinci surgical console or other forms of robotic surgical console providing real time 3D imaging.
   - **I/O:** Continuous input/output component manages all external input and output operations, ensuring that data from various sensors and devices is correctly captured and routed to the appropriate services.

2. **Core Components Part 1: Data Preprocessing and Storage**

   - **Data Storage:** Surgical data, including imaging and sensor readings, are stored in a centralized database. This storage solution must be robust and capable of handling large volumes of data generated during surgery.

- **Data Preprocessing:** Before data can be fed into the deep learning models, it undergoes preprocessing. This step includes tasks such as normalization, augmentation, and feature extraction, ensuring that the data is in the optimal format for analysis.
- **Control Service:** This service manages the flow of data through the system, coordinating the preprocessing and ensuring that the data is prepared for model deployment.

3. **Core Components Part 2: Model Deployment and Communication**

- **Feedback:** The feedback loop ensures that the outputs from the models are continually evaluated against the expected outcomes. This real-time feedback is crucial for adjusting model behavior during surgery, improving accuracy and performance.
- **Model Deployment:** The core service where the deep learning models are deployed and executed. This service must be highly optimized to allow for real-time inference, critical in a surgical environment.
- **Distribution & Data Replication:** Ensures that the model outputs and any processed data are replicated across the system to maintain consistency and reliability, even in the face of hardware or network failures.

4. **Reliability, Scalability, and Security**

- **Load Balancing:** To ensure the system remains responsive during complex surgical procedures, particularly when running multiple deep learning models concurrently, load balancing is a critical component. In the context of ROS and the da Vinci surgical console, load balancing distributes the computational tasks—such as real-time image processing, sensor data analysis, and model inference—across multiple CPUs and GPUs. This distribution is essential to prevent any single processing unit from becoming overwhelmed, which could lead to delays or interruptions during surgery. For example, if the system is processing high-resolution images from multiple cameras while running a deep learning model for tissue classification, the load balancer

will allocate these tasks to different GPUs or CPUs based on their current load, ensuring that each unit operates within its optimal capacity. This dynamic allocation is continuously adjusted based on real-time monitoring of system performance, ensuring that computational resources are used efficiently and that the surgical process is not hindered by processing delays.

- **Scaling:** As the surgical procedure progresses, the computational demands on the system may increase, particularly when more complex tasks or additional data streams are introduced. The system's ability to scale resources dynamically is therefore essential. In the context of the SOA integrated with ROS, scaling involves deploying additional computational resources, such as spinning up more virtual machines or allocating more GPUs, to handle the increased workload. For instance, if the procedure transitions from a routine task to a more complex one that requires intensive data analysis and real-time model inference, the system can automatically allocate additional GPUs to manage this increased demand [24][35]. Conversely, when the workload decreases, the system can scale down by releasing excess resources, thereby optimizing performance and reducing operational costs. This scalability is particularly important in ensuring that the deep learning models run smoothly without exhausting system resources, even during the most demanding surgical procedures.

- **Security:** Given the sensitive nature of surgical data, robust security protocols are embedded throughout the SOA, especially when interacting with the ROS environment. These protocols include end-to-end encryption of all data transmitted between components to protect against unauthorized access. Secure communication channels are established using protocols such as TLS (Transport Layer Security) to ensure that data exchanged between the ROS nodes, the SOA services, and the da Vinci surgical console remains confidential and tamper-proof. Additionally, rigorous access controls are implemented to restrict access to the system based on role and need, ensuring that only authorized personnel can interact with the deep learning models and the surgical data. For instance, different levels of access may be granted to

surgeons, system administrators, and developers, each with permissions tailored to their specific role [69][**?**]. Furthermore, continuous security monitoring is employed to detect and respond to potential threats in real-time, ensuring the integrity of the surgical process and safeguarding patient data.

5. **Da Vinci Surgical System and ROS Integration**

- **ROS Communication Nodes:** These nodes are the heart of the interaction between the SOA and the da Vinci surgical console. ROS (Robot Operating System) provides the middleware that connects the deep learning models with the robotic components of the da Vinci system. The communication nodes ensure that data flows seamlessly between the SOA and the ROS environment, allowing for real-time model inference and control of the surgical instruments.
- **Imaging and Sensor Fusion:** This component aggregates data from multiple sensors, including cameras and haptic feedback devices, to create a comprehensive understanding of the surgical environment. The fused data is then processed by the deep learning models to provide actionable insights during surgery.

6. **Logging, Monitoring, and Alerting**

- **Logging:** Continuous logging of all operations is critical for both auditing purposes and post-surgical analysis. This includes recording model inferences, system decisions, and any anomalies encountered during the procedure.
- **Monitoring:** Real-time monitoring of the entire system ensures that any issues are detected and addressed immediately. This includes tracking the performance of the deep learning models, system latency, and resource utilization.
- **Alerting:** If any component of the system fails or operates outside of expected parameters, alerts are generated. These alerts can be sent to the surgical team, allowing them to take corrective action quickly.

Interaction with ROS in the da Vinci Surgical Console

The interaction between this SOA and the ROS-based da Vinci surgical console is fundamental to the success of the system. ROS serves as the middleware that facilitates communication between the software components (deep learning models) and the hardware components (surgical instruments). The following points highlight how this interaction is managed:

- **Data Flow:** Surgical data captured by the ROS nodes is sent to the SOA for preprocessing, model inference, and feedback. The processed data, along with model outputs, is then sent back to the ROS environment, where it is used to guide surgical instruments or provide feedback to the surgeon.
- **Real-Time Inference:** The ROS nodes interact with the model deployment service in the SOA to ensure that deep learning models can operate in real-time. This requires highly efficient communication protocols and minimal latency to ensure that the surgeon's actions are accurately supported by the AI models.
- **System Synchronization:** ROS nodes are responsible for synchronizing the timing between the SOA services and the surgical instruments. This ensures that data is processed in the correct order and that the models' inferences are applied at the right moment during the surgery.

5.4 Integration with Current and Future Robotic Systems

The integration of AI models into robotic systems for surgical applications offers practical opportunities for improving the precision and accuracy of complex procedures. AI-driven models like Transformers have the ability to process complex patterns, allowing for more nuanced analysis of the anatomical structures and tools involved in surgery. This capability is key for future

integration, particularly in robotic surgery where decision-making must be immediate and accurate [88][?][?]. While there are still challenges to address, including computational demands and regulatory compliance, the long-term integration of such models can lead to more advanced and adaptive robotic systems [113][115][116]. Thus also considering the context of hardware being adapted across multiple domains and the necessary surgical view, while understanding risk of implementing novel technologies within high risk domains compared to other domains.

5.4.1 Clinical Explainability and Interpretability of Models

- **Clinical Applications:** AI models built for real-time anatomical segmentation, especially the Transformer-based architectures, have practical applications in both robotic and surgery. Their ability to process high-dimensional data from video feeds, sensors, and patient-specific imaging can assist in real-time decision-making during complex procedures. For instance, these models can help identify anatomical landmarks, and provide alerts in critical moments, thereby supporting surgeons during minimally invasive or robotic-assisted surgeries. As these models evolve, they can be customized for specific surgical disciplines such as cardiology or neurosurgery, where precision is critical.

- **Model Interpretability:** For the successful adoption of AI models in clinical environments, ensuring interpretability is essential. Surgeons and clinical staff need to understand how models make their decisions, especially during critical surgical procedures. The self-attention mechanism of Transformer models provides an opportunity to visualize which areas of the data the model is focusing on during inference, enabling clinicians to interpret the model's output. Additionally, implementing explainability methods like gradient-based saliency maps can offer insight into how the model prioritizes features, increasing trust in AI-assisted systems.

- **Expansion of Multimodal Models:** The future of AI models in robotic surgery likely lies in multimodal integration, where data from various sources such as imaging, video feeds, and robotic sensor data are combined to provide more comprehensive support during surgeries. The expansion of such models will allow for better generalization across

different surgical scenarios, as well as the ability to predict outcomes based on more complete patient profiles. This could enable highly specialized models that can adapt in real time to the specific needs of each surgery, enhancing both patient safety and surgical outcomes.

5.4.2 Ethical Considerations and Regulatory Compliance

Ethical deployment of advanced AI models, particularly those capable of predictive analytics across multiple domains, involves critical considerations around data usage, legal compliance, and international regulatory standards. These considerations are especially important for ensuring that the use of AI in robotic surgery is both ethically sound and aligned with global legal frameworks.

- **Data Allocation:** The ability to categorize components of anatomy in real time during surgery requires access to large volumes of surgical data for training AI models. Most current prediction models, such as those using segmentation masks for surgical tools, are limited by the quantity and diversity of available data. Therefore, the collection of vast amounts of surgical data is crucial for developing more reliable and accurate deep learning models. In addition, synthetic data generation, using simulation environments such as Unreal Engine 5 or Tesla's AI simulation, can augment real-world data and help train models to handle various scenarios. However, synthetic data must be validated against real-world clinical data to ensure it reflects true surgical conditions and is suitable for deployment

- **International Compliance:** The integration of AI models is subject to varying regulatory standards and adoption across different countries. As the use of robotic surgery expands globally, models must be adaptable not only to different hardware configurations but also to different regulatory requirements in various medical fields, such as cardiology or neurosurgery. Regularity practices, and implementation for novel technology in general have different regulations based on countries and procedures, while also considering

the cultural significance of what newer forms of advanced software can encompass. International competition in the AI-driven medical technology space, particularly between Western and Eastern countries, adds complexity to the regulatory landscape, coinciding with competition, capital allocation, and intensity of such capital for the basic definitions of AI and robotic surgery.

- **Capital Allocation & Distribution:** Developing and deploying AI-driven robotic surgery systems requires significant capital investment, particularly due to the costs associated with research, model training, hardware infrastructure, and regulatory approval. Stable, recurring investment streams, such as those from venture capital, convertible bonds, or even public funding, are necessary to sustain the development of these technologies. Financial planning should focus not only on covering the high initial costs but also on creating sustainable revenue models to support the long-term success and scalability of AI-assisted robotic systems.

5.4.3 Future Clinical Trials and Assessment Strategies

- **Clinical Trials Design:** To ensure the safe and effective integration of AI models into surgical practice, clinical trials must be designed to rigorously evaluate model performance in real-world settings. These trials should be randomized and controlled, covering a wide range of surgical procedures and patient demographics. The trials must also assess how well the models integrate into the existing surgical workflow, focusing on whether they enhance decision-making or create additional complexities for the surgical team.

- **Assessment Strategies:** Evaluating the performance of AI models in clinical settings requires a multi-faceted approach. Quantitative metrics such as accuracy, precision, recall, and F1 scores will provide a clear understanding of the model's technical capabilities. However, qualitative assessments, such as feedback from surgeons on usability and integration, are equally important. These assessments will help ensure that the model's outputs are not only accurate but also actionable and easily interpretable in high-pressure environments, reduce the learning curve for in-house residents, while causing less tissue

damage during surgery.

- **Outcome Evaluation:** Beyond immediate surgical outcomes, the long-term effects of AI integration on patient recovery, complication rates, and overall surgical efficiency must be closely monitored. Evaluating how AI-driven systems influence these factors over time will be critical in determining their true impact on healthcare. Continuous monitoring of patient outcomes, to even reducing the patient length of stay, will also provide valuable data to further refine and improve the models, ensuring that they continue to meet clinical needs and provide tangible benefits in surgical practice.

5.4.4 Educational Impact and Training Integration

- **Credentialing and Certification for Surgical Training (Robotic & General):** As AI models become more integrated into surgical procedures, new training and certification standards will be required to ensure that surgeons are adequately prepared to use these systems. For a big components of credentialing and certification for novel technology, is having to adapt to such frameworks and being cognizant of heuristics, such as adverse events occurring during procedures. In robotic surgery for instance, the most common form of adverse events pertain to broken pieces falling into patients or intraoperative malfunctioning electric cauterization. Training programs will need to focus on both technical proficiency with the AI systems and the ability to interpret and act upon AI-generated insights during surgery. Certification bodies should develop formal credentialing processes to ensure that surgeons are trained in the use of AI tools and understand their limitations.

- **Residence Training:** Surgical residency programs will play a crucial role in preparing future surgeons to work with AI-assisted systems. These programs must integrate AI tools into their curricula, providing residents with hands-on experience in using AI models during both simulated and real-world procedures. By gaining familiarity with AI-driven systems early in their careers, future surgeons will be better equipped to leverage these technologies effectively in clinical practice.

- **Towards Real-Time Intra-operative Procedures:** AI models, particularly those capable of real-time processing, hold the potential to assist surgeons during intra-operative procedures by providing actionable insights as surgery unfolds. Training surgeons to work effectively with these real-time systems is essential for ensuring that the AI model's outputs can be seamlessly integrated into the flow of the operation. This will allow surgeons to make more informed decisions in real time, potentially improving outcomes and reducing the likelihood of complications.

5.5 Conclusion of Results

The evaluation of deep learning models within the da Vinci surgical system has yielded significant insights, particularly highlighting the Transformer model architecture's superior performance. Among the models tested, including the standard CNN, 1D-CNN, and non-pretrained instance segmentation models from Intel's OpenVINO library, the Transformer model consistently outperformed its counterparts. This was evident across various metrics such as accuracy, F1 scores, and overall classification performance, particularly in scenarios that demanded a high degree of precision in identifying and interpreting complex and overlapping anatomical structures. The success of the Transformer model can be largely attributed to its sophisticated self-attention mechanism, which allows the model to dynamically focus on the most relevant aspects of the input data. This capability is less developed in the other architectures tested, giving the Transformer a distinct advantage in real-time surgical applications where the rapid and accurate processing of incoming data is critical. The model's extensive parameterization, with 77,717,506 parameters and multiple hidden layers, provides the depth required to capture subtle variations in anatomical structures, enabling the model to deliver high reliability and accuracy even in the most challenging surgical scenarios. These findings underline the importance of deploying advanced architectures like the Transformer within a robust and flexible framework. The Service-Oriented Architecture (SOA) designed for this purpose, integrated with the ROS-based da Vinci surgical console, provides a powerful solution for the deployment of such models. The SOA's modularity allows for the seamless integration of the Transformer

model and other AI tools, supporting the dynamic and scalable resource allocation necessary to meet the varying computational demands of surgical procedures. In addition to its technical advantages, the SOA framework is designed with security and reliability at its core, ensuring that sensitive patient data is protected and that the surgical process remains uninterrupted by technical failures. The ability to scale resources dynamically and balance computational loads across multiple processing units is particularly crucial in maintaining system responsiveness during complex procedures, where delays or interruptions could have serious consequences. This level of resilience is essential for clinical deployment, where the stakes are high and the margin for error is minimal. The implications of these results extend beyond the immediate application of the Transformer model in robotic surgery. The successful integration of such advanced AI models into the surgical workflow represents a significant step forward in the broader field of AI-assisted medicine. It opens the door to more sophisticated and personalized surgical care, where AI-driven insights can enhance the surgeon's decision-making process, reduce the risk of errors, and ultimately improve patient outcomes.

Chapter 6

Conclusion

The research focusing on the critical role of applying advanced software for robotic surgical procedures, underscores the critical role that deep learning models play in advancing surgical precision and efficiency, particularly through real-time categorization and segmentation of anatomical structures. The application of these models within various surgical domains, including robotic-assisted and general endoscopic surgeries, has shown significant promise in enhancing intraoperative decision-making and improving patient outcomes. The study demonstrates that advanced deep learning models, especially the Vision Transformer, excel in handling complex and overlapping anatomical structures. The Vision Transformer's architecture allowed it to achieve higher accuracy and more reliable performance metrics compared to other models such as the 1D-CNN and standard CNN. This model's effectiveness in real-time segmentation tasks highlights its potential for broad application across different surgical disciplines, where precise and timely information is crucial for successful outcomes. However, it is important to note that the 1D-CNN and standard CNN models also showed strong performance, particularly in scenarios requiring high computational efficiency and speed. These models provide valuable alternatives in contexts where resource constraints or specific surgical environments may dictate the need for lighter models[78][80][82]. The comparison with non-pretrained models from Intel's OpenVINO library further emphasizes the importance of using well-configured, pretrained architectures tailored to the demands of surgical applications. A key insight from this research

is the necessity of optimizing deep learning model architectures to achieve high accuracy in real-world surgical settings. The ability to fine-tune parameters, such as the number of hidden layers and the configuration of the model's architecture, directly correlates with performance outcomes. The study advocates for a model-centric approach, focusing on the development of deep learning models that can generalize across various surgical procedures by leveraging diverse data sources. This approach is expected to improve the models' ability to perform real-time segmentation tasks across different surgical disciplines, enhancing the overall quality of care. One of the major challenges identified in this research is the difficulty in obtaining large, diverse datasets of surgical video required for training these models [39][44][45][46]. The specificity of existing datasets limits the generalist ability of the models, making it challenging to apply them across different types of surgery [47][48][50]. To address this issue, the study suggests adopting strategies similar to those used in other medical fields, where data from multiple sources is integrated to develop more versatile and robust models.The systematic literature review highlights the current state of deep learning applications in computer vision, specifically in the context of endoscopic anatomy in both robotic and general surgery. While there has been considerable progress, there remains a significant gap between the capabilities of current models and the needs of the surgical community [52][63][65]. The lack of comprehensive, open-source surgical datasets continues to be a barrier to the development and widespread deployment of these models in clinical practice [70][72][79]. To address these challenges, this research proposes a case study comparing models trained on both robotic and general endoscopic data from Imperial College's Open Source Endoscopic Surgical Video database. The goal is to demonstrate the feasibility of developing a universal deep learning model capable of real-time tissue segmentation with high accuracy across various surgical disciplines. The results of this case study are expected to provide valuable insights into the creation of robust, accurate models that can be implemented in a wide range of surgical settings. By focusing on the development of universal deep learning models for surgery, the study opens the door to broader applications in the medical field. These models could be adapted for use in other minimally invasive procedures, such as endoscopy or laparoscopy, thereby extending their impact[95][96][103]. Additionally, by providing real-time feedback and decision support during surgery, these models could shorten

the training period required for surgeons and reduce the need for extensive hands-on experience, making advanced surgical techniques more accessible to a broader range of healthcare providers. Furthermore, this research highlights the importance of a collaborative approach in the development of deep learning models for surgery. The complexity of surgical procedures and the variability of human anatomy necessitate collaboration between surgeons, engineers, and computer scientists to ensure that the models are technically sound, clinically relevant, and practical for real-world application [108][110][111]. By focusing on model development tailored to real-time tissue segmentation across various surgical disciplines, this study provides a roadmap for improving surgical precision, efficiency, and patient outcomes. However, realizing this potential will require addressing significant challenges, including the scarcity of surgical data, the specificity of models, and the collaborative efforts necessary to develop and deploy these models effectively. The focus shifts from merely gathering a vast corpus of surgical data, which is both time-consuming and costly, to the development of sophisticated AI models, and questing the foundations for after acquiring such data[94][98][101]. These models are designed to accurately segment anatomy in real-time across various surgical disciplines, heralding a new era in the field of robotic surgery where universal applicability, irrespective of the specific surgical procedure or hardware employed, becomes the norm. This approach not only promises to enhance the precision and efficiency of surgical care but also paves the way for broader application and innovation in the realm of robotic surgery and applied artificial intelligence.

References

[1] A. Aghabiglou and E. M. Eksioglu. Projection-based cascaded u-net model for mr image reconstruction. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE*, 207, 2021.

[2] T. Alpay, S. Magg, P. Broze, and D. Speck. Multimodal video retrieval with clip: a user study. *INFORMATION RETRIEVAL JOURNAL*, 26(1), 2023.

[3] M. Antico, F. Sasazawa, Y. Takeda, A. T. Jaiprakash, M. L. Wille, A. K. Pandey, R. Crawford, G. Carneiro, and D. Fontanarosa. Bayesian cnn for segmentation uncertainty inference on 4d ultrasound images of the femoral cartilage for guidance in robotic knee arthroscopy. *IEEE Access*, 8:223961–223975, 2020.

[4] S Azadi, IC Green, A Arnold, M Truong, J Potts, and MA Martino. Robotic surgery: The impact of simulation and other innovative platforms on performance and training. *JOURNAL OF MINIMALLY INVASIVE GYNECOLOGY*, 28(3):490–495, 2021.

[5] I. Bag, E. Bilgir, I. S. Bayrakdar, O. Baydar, F. M. Atak, Ö Çelik, and K. Orhan. An artificial intelligence study: automatic description of anatomic landmarks on panoramic radiographs in the pediatric population. *BMC ORAL HEALTH*, 23(1), 2023.

[6] B. Bayramli, J. Hur, and H. T. Lu. Raft-msf: Self-supervised monocular scene flow using recurrent optimizer. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 131(11):2757–2769, 2023.

[7] S. Bechtle, N. Das, and F. Meier. Multimodal learning of keypoint predictive models for visual object manipulation. *IEEE TRANSACTIONS ON ROBOTICS*, 39(2):1212–1224, 2023.

[8] M. Cao, J. L. Yuan, H. L. Yu, B. M. Zhang, and C. J. Wang. Self-supervised short text classification with heterogeneous graph neural networks. *EXPERT SYSTEMS*, 40(6), 2023.

[9] Rui Cao, Scott D. Nelson, Samuel Davis, Yu Liang, Yilin Luo, Yide Zhang, Brooke Crawford, and Lihong V. Wang. Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy. *Nature Biomedical Engineering*, 7(2):124–134, 2023.

[10] J Cartucho, S Tukra, YP Li, DS Elson, and S Giannarou. Visionblender: a tool to efficiently generate computer vision datasets for robotic surgery. *COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING-IMAGING AND VISUALIZATION*, 9(4):331–338, 2021.

[11] S Chakraborty, S Pillai, and IEEE. *AI/ML enabled decision making in facilitating Robotic surgery.* 2022.

[12] K. X. Chen, T. Asada, N. Ienaga, K. Miura, K. Sakashita, T. Sunami, H. Kadone, M. Yamazaki, and Y. Kuroda. Two-stage video-based convolutional neural networks for adult spinal deformity classification. *FRONTIERS IN NEUROSCIENCE*, 17, 2023.

[13] L. Chen, W. Tang, N. W. John, T. R. Wan, and J. J. Zhang. De-smokegcn: Generative cooperative networks for joint surgical smoke detection and removal. *IEEE Transactions on Medical Imaging*, 39(5):1615–1625, 2020.

[14] Y. Chen, J. H. Yan, Y. S. Gao, Y. Zhang, Y. Liu, and M. W. Shi. Hyperspectral image ground-object identification method based on spectral segment fusion combination and depth residual network. *AIP ADVANCES*, 13(8), 2023.

[15] E. Colleoni and D. Stoyanov. Robotic instrument segmentation with image-to-image translation. *IEEE Robotics and Automation Letters*, 6(2):935–942, 2021.

[16] D. V. Dong, K. Nagasubramanian, R. D. Wang, U. K. Frei, T. Z. Jubery, T. Luebberstedt, and B. Ganapathysubramanian. Self-supervised maize kernel classification and segmentation for embryo identification. *FRONTIERS IN PLANT SCIENCE*, 14, 2023.

[17] G. N. Dong, C. M. Pun, and Z. Zhang. Deep collaborative multi-modal learning for unsupervised kinship estimation. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, 16:4197–4210, 2021.

[18] A. Dorzhigulov and V. Saxena. Spiking cmos-nvm mixed-signal neuromorphic convnet with circuit- and training-optimized temporal subsampling. *FRONTIERS IN NEUROSCIENCE*, 17, 2023.

[19] M El-Ahmar, F Peters, M Green, M Dietrich, M Ristig, L Moikow, and JP Ritz. Robotic colorectal resection in combination with a multimodal enhanced recovery program-results of the first 100 cases. *INTERNATIONAL JOURNAL OF COLORECTAL DISEASE*, 38(1), 2023.

[20] F. Falezza, N. Piccinelli, G. De Rossi, A. Roberti, G. Kronreif, F. Setti, P. Fiorini, and R. Muradore. Modeling of surgical procedures using statecharts for semi-autonomous robotic surgery. *IEEE Transactions on Medical Robotics and Bionics*, 3(4):888–899, 2021.

[21] W. Fan, Z. Zheng, W. Zeng, Y. Chen, H. Q. Zeng, H. Shi, and X. Luo. Robotically surgical vessel localization using robust hybrid video motion magnification. *IEEE Robotics and Automation Letters*, 6(2):1567–1573, 2021.

[22] Ge Fang, Marco C K Chow, Justin D L Ho, Zhuoliang He, Kui Wang, T C Ng, James K H Tsoi, Po-Ling Chan, Hing-Chiu Chang, Danny Tat-Ming Chan, Yun-hui Liu, F Christopher Holsinger, Jason Ying-Kuen Chan, and Ka-Wai Kwok. Soft robotic manipulator for intraoperative mri-guided transoral laser microsurgery. Report, 2021 2021.

[23] A. Fedorov, E. Geenjaar, L. Wu, T. Sylvain, T. P. DeRamus, M. Luck, M. Misiura, G. Mittapalle, R. D. Hjelm, S. M. Plis, and V. D. Calhoun. Self-supervised multimodal learning for group inferences from mri data: Discovering disorder-relevant brain regions and multimodal links. *NEUROIMAGE*, 285, 2024.

[24] I. R. B. Godoy, R. P. Silva, T. C. Rodrigues, A. Y. Skaf, A. D. Pochini, and A. F. Yamada. Automatic mri segmentation of pectoralis major muscle using deep learning. *SCIENTIFIC REPORTS*, 12(1), 2022.

[25] B. B. Han, Y. M. Wei, Q. Y. Wang, and S. S. Wan. Colm¡sup¿2¡/sup¿s: Contrastive self-supervised learning on attributed multiplex graph network with multi-scale information. *CAAI TRANSACTIONS ON INTELLIGENCE TECHNOLOGY*, 8(4):1464–1479, 2023.

[26] Anders E Hansen, Jonas R Henriksen, Rasmus I Jølck, Frederikke P Fliedner, Linda M Bruun, Jonas Scherman, Andreas I Jensen, Per Munck af Rosenschöld, Lilah Moorman, Sorel Kurbegovic, Steen R de Blanck, Klaus R Larsen, Paul F Clementsen, Anders N Christensen, Mads H Clausen, Wenbo Wang, Paul Kempen, Merete Christensen, Niels-Erik Viby, Gitte Persson, Rasmus Larsen, Knut Conradsen, Fintan J McEvoy, Andreas Kjaer, Thomas Eriksen, and Thomas L Andresen. H e a l t h a n d m e d i c i n e multimodal soft tissue markers for bridging high-resolution diagnostic imaging with therapeutic intervention. Report, 2020 2020.

[27] M. A. Hassan, B. W. Weyers, J. Bec, F. Fereidouni, J. Qi, D. Gui, A. F. Bewley, M. Abouyared, D. G. Farwell, A. C. Birkeland, and L. Marcu. Anatomy-specific classification model using label-free flim to aid intraoperative surgical guidance of head and neck cancer. *IEEE Transactions on Biomedical Engineering*, 70(10):2863–2873, 2023.

[28] MA Hassan, B Weyers, J Bec, JY Qi, D Gui, A Bewley, M Abouyared, G Farwell, A Birkeland, and L Marcu. *FLIm-Based in Vivo Classification of Residual Cancer in the Surgical Cavity During Transoral Robotic Surgery*, volume 14228. 2023.

[29] M. Y. He, Q. Q. Zhao, and H. Zhang. Multi-sample dual-decoder graph autoencoder. *METHODS*, 211:31–41, 2023.

[30] A. S. Hervella, J. Rouco, J. Novo, and M. Ortega. Multimodal image encoding pre-training for diabetic retinopathy grading. *COMPUTERS IN BIOLOGY AND MEDICINE*, 143, 2022.

[31] S. J. Hu, F. Bonardi, S. Bouchafa, and D. Sidibé. Multi-modal unsupervised domain adaptation for semantic image segmentation. *PATTERN RECOGNITION*, 137, 2023.

[32] J. H. Huang, F. Y. Lei, J. J. Jiang, X. Zeng, R. J. Ma, and Q. Y. Dai. Multi-order hypergraph convolutional networks integrated with self-supervised learning. *COMPLEX INTELLIGENT SYSTEMS*, 9(4):4389–4401, 2023.

[33] J. Z. Ji, H. Jia, Y. T. Ren, and M. L. Lei. Supervised contrastive learning with structure inference for graph classification. *IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING*, 10(3):1684–1695, 2023.

[34] Q. Y. Jin, S. M. Li, X. F. Du, M. Z. Yuan, M. N. Wang, and Z. J. Song. Density-based one-shot active learning for image segmentation. *ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE*, 126, 2023.

[35] A. Jungo, O. Scheidegger, M. Reyes, and F. Balsiger. pymia: A python package for data handling and evaluation in deep learning-based medical image analysis. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE*, 198, 2021.

[36] Y. Kang, K. Liu, Z. Y. Cao, and J. C. Zhang. Self-supervised node classification with strategy and actively selected labeled set. *ENTROPY*, 25(1), 2023.

[37] G. Keser, I. S. Bayrakdar, F. N. Pekiner, Ö Çelik, and K. Orhan. A deep learning approach for masseter muscle segmentation on ultrasonography. *JOURNAL OF ULTRASONOGRAPHY*, 22(91):e204–e208, 2022.

[38] K. M. Kim, M. Y. Lee, H. S. Won, M. J. Kim, Y. Kim, and S. Lee. Multi-stage prompt tuning for political perspective detection in low-resource settings. *APPLIED SCIENCES-BASEL*, 13(10), 2023.

[39] S. Kim, S. Yun, J. Lee, G. Chang, W. Roh, D. N. Sohn, J. T. Lee, H. Park, and S. Kim. Self-supervised multimodal graph convolutional network for collaborative filtering. *INFORMATION SCIENCES*, 653, 2024.

[40] J. Kishore and S. Mukherjee. Auto cnn classifier based on knowledge transferred from self-supervised model. *APPLIED INTELLIGENCE*, 53(19):22086–22104, 2023.

[41] J. D. Kong, Y. T. He, X. M. Zhu, P. F. Shao, Y. Xu, Y. Chen, J. L. Coatrieux, and G. Y. Yang. Bkc-net: Bi-knowledge contrastive learning for renal tumor diagnosis on 3d ct images. *KNOWLEDGE-BASED SYSTEMS*, 252, 2022.

[42] A. K. Koupai, M. J. Bocus, R. Santos-Rodriguez, R. J. Piechocki, and R. McConville. Self-supervised multimodal fusion transformer for passive activity recognition. *IET WIRELESS SENSOR SYSTEMS*, 12(5):149–160, 2022.

[43] Y. E. Lee, H. M. Husin, M. P. Forte, S. W. Lee, and K. J. Kuchenbecker. Learning to estimate palpation forces in robotic surgery from visual-inertial data. *IEEE Transactions on Medical Robotics and Bionics*, 5(3):496–506, 2023.

[44] T. Li, W. Jin, R. D. Fu, and C. F. He. Daytime sea fog monitoring using multimodal self-supervised learning with band attention mechanism. *NEURAL COMPUTING APPLICATIONS*, 34(23):21205–21222, 2022.

[45] W. F. Li, B. Li, S. L. Niu, Z. R. Wang, B. H. Liu, and T. Z. Niu. Selecting informative data for defect segmentation from imbalanced datasets via active learning. *ADVANCED ENGINEERING INFORMATICS*, 56, 2023.

[46] X. M. Li, M. Y. Jia, M. T. Islam, L. Q. Yu, and L. Xing. Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 39(12):4023–4033, 2020.

[47] Y. Li, S. Y. Liu, X. J. Wang, and P. G. Jing. Self-supervised deep partial adversarial network for micro-video multimodal classification. *INFORMATION SCIENCES*, 630:356–369, 2023.

[48] Y. F. Li, X. Wang, S. H. Qi, C. K. Huang, O. E. Jiang, Q. Liao, J. Guan, and J. J. Zhang. Self-supervised learning-based weight adaptive hashing for fast cross-modal retrieval. *SIGNAL IMAGE AND VIDEO PROCESSING*, 15(4):673–680, 2021.

[49] Z. Li, M. Shahbazi, N. Patel, E. O'Sullivan, H. Zhang, K. Vyas, P. Chalasani, A. Deguet, P. L. Gehlbach, I. Iordachita, G. Z. Yang, and R. H. Taylor. Hybrid robot-assisted frameworks for endomicroscopy scanning in retinal surgeries. *IEEE Transactions on Medical Robotics and Bionics*, 2(2):176–187, 2020.

[50] J. K. Lin, Q. L. Cai, and M. Y. Lin. Multi-label classification of fundus images with graph convolutional network and self-supervised learning. *IEEE SIGNAL PROCESSING LETTERS*, 28:454–458, 2021.

[51] CY Liu, CY Shi, TS Wang, HX Zhang, L Jing, XY Jin, J Xu, and HY Wang. Bio-inspired multimodal 3d endoscope for image-guided and robotic surgery. *OPTICS EXPRESS*, 29(1):145–157, 2021.

[52] H. Liu, J. D. Han, Y. J. Fu, Y. Y. Li, K. Chen, and H. Xiong. Unified route representation learning for multi-modal transportation recommendation with spatiotemporal pre-training. *VLDB JOURNAL*, 32(2):325–342, 2023.

[53] B. P. L. Lo, K. W. S. Au, and P. W. Y. Chiu. Guest editorial special section on the hamlyn symposium 2022—medtech reimagined. *IEEE Transactions on Medical Robotics and Bionics*, 6(1):2–3, 2024.

[54] B. Lu, H. K. Chu, K. Huang, and J. Lai. Surgical suture thread detection and 3-d reconstruction using a model-free approach in a calibrated stereo visual system. *IEEE/ASME Transactions on Mechatronics*, 25(2):792–803, 2020.

[55] F Luongo, R Hakim, JH Nguyen, A Anandkumar, and AJ Hung. Deep learning-based computer vision to recognize and classify suturing gestures in robot-assisted surgery. *SURGERY*, 169(5):1240–1244, 2021.

[56] Mohamed E M K Abdelaziz, Jinshi Zhao, Bruno Gil Rosa, Hyun-Taek Lee, Daniel Simon, Khushi Vyas, Bing Li, Hanifa Koguna, Yue Li, Ali Anil Demircali, Huseyin Uvet, Gulsum Gencoglan, Arzu Akcay, Mohamed Elriedy, James Kinross, Ranan Dasgupta, Zoltan Takats, Eric Yeatman, Guang-Zhong Yang, and Burak Temelkuran. A p p l i e d s c i e

n c e s a n d e n g i n e e r i n g fiberbots: Robotic fibers for high-precision minimally invasive surgery. Report, 2024 2024.

[57] F. Mahmood, R. Chen, and N. J. Durr. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE Transactions on Medical Imaging*, 37(12):2572–2581, 2018.

[58] N Marahrens, B Scaglioni, D Jones, R Prasad, CS Biyani, and P Valdastri. Towards autonomous robotic minimally invasive ultrasound scanning and vessel reconstruction on non-planar surfaces. *FRONTIERS IN ROBOTICS AND AI*, 9, 2022.

[59] D. Meli and P. Fiorini. Unsupervised identification of surgical robotic actions from small non-homogeneous datasets. *IEEE Robotics and Automation Letters*, 6(4):8205–8212, 2021.

[60] YY Meng, YG You, PX Geng, ZC Song, HP Wang, YD Qin, and IEEE. *Development of an Intra-Operative Active Navigation System for Robot-Assisted Surgery.* 2021.

[61] R. Moccia, C. Iacono, B. Siciliano, and F. Ficuciello. Vision-based dynamic virtual fixtures for tools collision avoidance in robotic surgery. *IEEE Robotics and Automation Letters*, 5(2):1650–1655, 2020.

[62] H. Mohammad-Rahimi, S. R. Motamadian, M. Nadimi, S. Hassanzadeh-Samani, M. A. S. Minabi, E. Mahmoudinia, V. Y. Lee, and M. H. Rohban. Deep learning for the classification of cervical maturation degree and pubertal growth spurts: A pilot study. *KOREAN JOURNAL OF ORTHODONTICS*, 52(2):112–122, 2022.

[63] J. H. Moon, H. Lee, W. Shin, Y. H. Kim, and E. Choi. Multi-modal understanding and generation for medical images and text via vision-language pre-training. *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, 26(12):6070–6080, 2022.

[64] ND Nguyen, T Nguyen, S Nahavandi, A Bhatti, G Guest, and IEEE. *Manipulating Soft Tissues by Deep Reinforcement Learning for Autonomous Robotic Surgery.* 2019.

[65] T. D. Nguyen, D. T. Le, J. Bum, S. Kim, S. J. Song, and H. Choo. Self-fi: Self-supervised learning for disease diagnosis in fundus images. *BIOENGINEERING-BASEL*, 10(9), 2023.

[66] K. Nozawa, S. Maki, T. Furuya, S. Okimatsu, T. Inoue, A. Yunde, M. Miura, Y. Shiratani, Y. Shiga, K. Inage, Y. Eguchi, S. Ohtori, and S. Orita. Magnetic resonance image segmentation of the compressed spinal cord in patients with degenerative cervical myelopathy using convolutional neural networks. *INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGY AND SURGERY*, 18(1):45–54, 2023.

[67] Kutsev Bengisu Ozyoruk, Sermet Can, Berkan Darbaz, Kayhan Başak, Derya Demir, Guliz Irem Gokceler, Gurdeniz Serin, Uguray Payam Hacisalihoglu, Emirhan Kurtuluş, Ming Y. Lu, Tiffany Y. Chen, Drew F. K. Williamson, Funda Yılmaz, Faisal Mahmood, and Mehmet Turan. A deep-learning model for transforming the style of tissue images from cryosectioned to formalin-fixed and paraffin-embedded. *Nature Biomedical Engineering*, 6(12):1407–1419, 2022.

[68] S. A. Pedram, C. Shin, P. W. Ferguson, J. Ma, E. P. Dutson, and J. Rosen. Autonomous suturing framework and quantification using a cable-driven surgical robot. *IEEE Transactions on Robotics*, 37(2):404–417, 2021.

[69] Xuejun Qian, Jing Pei, Hui Zheng, Xinxin Xie, Lin Yan, Hao Zhang, Chunguang Han, Xiang Gao, Hanqi Zhang, Weiwei Zheng, Qiang Sun, Lu Lu, and K. Kirk Shung. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning. *Nature Biomedical Engineering*, 5(6):522–532, 2021.

[70] Y. Q. Qiao, J. H. Lü, T. Wang, K. X. Liu, B. C. Zhang, and H. Snoussi. A multi-head attention self-supervised representation model for industrial sensors anomaly detection. *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, 2023.

[71] Y. Qin, M. Allan, J. W. Burdick, and M. Azizian. Autonomous hierarchical surgical state estimation during robot-assisted surgery through deep neural networks. *IEEE Robotics and Automation Letters*, 6(4):6220–6227, 2021.

[72] L. H. Qu, S. L. Liu, M. N. Wang, S. M. Li, S. Q. Yin, and Z. J. Song. Trans2fuse: Empowering image fusion through self-supervised learning and multi-modal transformations via transformer networks. *EXPERT SYSTEMS WITH APPLICATIONS*, 236, 2024.

[73] N Raison, K Ahmed, A Aydin, A Mottrie, H Van Der Poel, and P Dasgupta. Multi-institutional validation and assessment of training modalities in robotic surgery (the mars project). *JOURNAL OF UROLOGY*, 195(4):E117–E117, 2016.

[74] M. D. I. Reyzabal, M. Chen, W. Huang, S. Ourselin, and H. Liu. Dafoes: Mixing datasets towards the generalization of vision-state deep-learning force estimation in minimally invasive robotic surgery. *IEEE Robotics and Automation Letters*, 9(3):2527–2534, 2024.

[75] F. Richter, S. Shen, F. Liu, J. Huang, E. K. Funk, R. K. Orosco, and M. C. Yip. Autonomous robotic suction to clear the surgical field for hemostasis using image-based blood flow detection. *IEEE Robotics and Automation Letters*, 6(2):1383–1390, 2021.

[76] JRU Roldan, MY Jung, F Shen, D Milutinovic, and IEEE. *Robot-Assisted Orthopedic Surgery Bone Pose Identification Using Task-Specific Capability Maps*. 2023.

[77] H Saeidi, J D Opfermann, M Kam, S Wei, S Leonard, M H Hsieh, J U Kang, and A Krieger. Autonomous robotic laparoscopic surgery for intestinal anastomosis. Report, 2022 2022.

[78] H Salman, E Ayvali, RA Srivatsan, YF Ma, N Zevallos, R Yasin, L Wang, N Simaan, H Choset, and IEEE. *Trajectory-Optimized Sensing for Active Search of Tissue Abnormalities in Robotic Surgery*. 2018.

[79] A. M. L. Santilli, A. Jamzad, A. Sedghi, M. Kaufmann, K. Logan, J. Wallis, K. Y. M. Ren, N. Janssen, S. Merchant, J. Engel, D. McKay, S. Varma, A. Wang, G. Fichtinger, J. F. Rudan, and P. Mousavi. Domain adaptation and self-supervised learning for surgical margin detection. *INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGY AND SURGERY*, 16(5):861–869, 2021.

[80] A. Saracino, T. J. C. Oude-Vrielink, A. Menciassi, E. Sinibaldi, and G. P. Mylonas. Haptic intracorporeal palpation using a cable-driven parallel robot: A user study. *IEEE Transactions on Biomedical Engineering*, 67(12):3452–3463, 2020.

[81] Y. Sato, Y. Takegami, T. Asamoto, Y. Ono, T. Hidetoshi, R. Goto, A. Kitamura, and S. Honda. Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study. *BMC MUSCULOSKELETAL DISORDERS*, 22(1), 2021.

[82] P. M. Scheikl, E. Tagliabue, B. Gyenes, M. Wagner, D. Dall'Alba, P. Fiorini, and F. Mathis-Ullrich. Sim-to-real transfer for visual reinforcement learning of deformable object manipulation for robot-assisted surgery. *IEEE Robotics and Automation Letters*, 8(2):560–567, 2023.

[83] F. M. Schönleitner, L. Otter, S. K. Ehrlich, and G. Cheng. Calibration-free error-related potential decoding with adaptive subject-independent models: A comparative study. *IEEE Transactions on Medical Robotics and Bionics*, 2(3):399–409, 2020.

[84] Azad Shademan, Ryan S Decker, Justin D Opfermann, Simon Leonard, Axel Krieger, and Peter C W Kim. Supervised autonomous robotic soft tissue surgery. Report.

[85] S. B. Shafiei, A. S. Elsayed, A. A. Hussein, U. Iqbal, and K. A. Guru. Evaluating the mental workload during robot-assisted surgery utilizing network flexibility of human brain. *IEEE Access*, 8:204012–204019, 2020.

[86] A. Soleymani, X. Li, and M. Tavakoli. A domain-adapted machine learning approach for visual evaluation and interpretation of robot-assisted surgery skills. *IEEE Robotics and Automation Letters*, 7(3):8202–8208, 2022.

[87] A Soleymani, XY Li, M Tavakoli, and IEEE. *Deep Neural Skill Assessment and Transfer: Application to Robotic Surgery Training*. 2021.

[88] A. Stone, C. Kalahiki, L. Li, N. Hubig, F. Iuricich, and H. Dunn. Evaluation of breast tumor morphologies from african american and caucasian patients. *COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL*, 21:3459–3465, 2023.

[89] B Triana, J Cha, A Shademan, A Krieger, JU Kang, and PCW Kim. *Multispectral tissue analysis and classification towards enabling automated robotic surgery*, volume 8935. 2014.

[90] M. Unberath, J. N. Zaech, C. Gao, B. Bier, F. Goldmann, S. C. Lee, J. Fotouhi, R. Taylor, M. Armand, and N. Navab. Enabling machine learning in x-ray-based procedures via realistic simulation of image formation. *INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGY AND SURGERY*, 14(9):1517–1528, 2019.

[91] E. Upschulte, S. Harmeling, K. Amunts, and T. Dickscheid. Contour proposal networks for biomedical instance segmentation. *MEDICAL IMAGE ANALYSIS*, 77, 2022.

[92] NS van den Berg, T Buckle, GH KleinJan, HG van der Poel, and FWB van Leeuwen. Multispectral fluorescence imaging during robot-assisted laparoscopic sentinel node biopsy: A first step towards a fluorescence-based anatomic roadmap. *EUROPEAN UROLOGY*, 72(1):110–117, 2017.

[93] G. T. Wang, X. D. Luo, R. Gu, S. J. Yang, Y. J. Qu, S. W. Zhai, Q. F. Zhao, K. Li, and S. T. Zhang. Pymic: A deep learning toolkit for annotation-efficient medical image segmentation. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE*, 231, 2023.

[94] W. Wang and Y. S. Wang. Deep learning-based modified yolact algorithm on magnetic resonance imaging images for screening common and difficult samples of breast cancer. *DIAGNOSTICS*, 13(9), 2023.

[95] Y. Wang, D. Ni, and Z. Y. Huang. A momentum contrastive learning framework for low-data wafer defect classification in semiconductor manufacturing. *APPLIED SCIENCES-BASEL*, 13(10), 2023.

[96] Z. H. Wang, A. Mariani, A. Menciassi, E. De Momi, and A. M. Fey. Uncertainty-aware self-supervised learning for cross-domain technical skill assessment in robot-assisted surgery. *IEEE TRANSACTIONS ON MEDICAL ROBOTICS AND BIONICS*, 5(2):301–311, 2023.

[97] J White and A Sharma. Development and assessment of a transoral robotic surgery curriculum to train otolaryngology residents. *ORL-JOURNAL FOR OTO-RHINO-LARYNGOLOGY HEAD AND NECK SURGERY*, 80(2):69–76, 2018.

[98] A. J. Witten, N. Patel, and A. Cohen-Gadol. Image segmentation of operative neuroanatomy into tissue categories using a machine learning construct and its role in neurosurgical training. *OPERATIVE NEUROSURGERY*, 23(4):279–286, 2022.

[99] Steve Wozniak and Gina Smith. *iWoz: Computer Geek to Cult Icon*. W. W. Norton & Company, New York, NY, 2006.

[100] JY Wu, P Kazanzides, and M Unberath. Leveraging vision and kinematics data to improve realism of biomechanic soft tissue simulation for robotic surgery. *INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGY AND SURGERY*, 15(5):811–818, 2020.

[101] Zhenqin Wu, Alexandro E. Trevino, Eric Wu, Kyle Swanson, Honesty J. Kim, H. Blaize D'Angio, Ryan Preska, Gregory W. Charville, Piero D. Dalerba, Ann Marie Egloff, Ravindra Uppaluri, Umamaheswar Duvvuri, Aaron T. Mayer, and James Zou. Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nature Biomedical Engineering*, 6(12):1435–1448, 2022.

[102] P. Yadlapalli and D. Bhavana. Segmentation and pre-processing of interstitial lung disease using deep learning model. *SCALABLE COMPUTING-PRACTICE AND EXPERIENCE*, 23(4):403–420, 2022.

[103] Y. Yan, Y. Shu, S. Chen, J. H. Xue, C. H. Shen, and H. Z. Wang. Spl-net: Spatial-semantic patch learning network for facial attribute recognition with limited labeled data. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 131(8):2097–2121, 2023.

[104] JH Yang, ED Goodman, AJ Dawes, JV Gahagan, MM Esquivel, CA Liebert, C Kin, S Yeung, and BH Gurland. Using ai and computer vision to analyze technical proficiency in robotic surgery. *SURGICAL ENDOSCOPY AND OTHER INTERVENTIONAL TECHNIQUES*, 37(4):3010–3017, 2023.

[105] J. R. Yi, P. X. Wu, M. L. Jiang, Q. Y. Huang, D. J. Hoeppner, and D. N. Metaxas. Attentive neural cell instance segmentation. *MEDICAL IMAGE ANALYSIS*, 55:228–240, 2019.

[106] Michael Yip and Septimiu Salcudean. Leveraging ai for medical image-guided robotics. Report.

[107] S. H. Yoo, H. Geng, T. L. Chiu, S. K. Yu, D. C. Cho, J. Heo, M. S. Choi, I. H. Choi, C. C. Van, N. V. Nhung, B. J. Min, and H. Lee. Deep learning-based decision-tree classifier for covid-19 diagnosis from chest x-ray imaging. *FRONTIERS IN MEDICINE*, 7, 2020.

[108] L. Zaadnoordijk, T. R. Besold, and R. Cusack. Lessons from infant learning for unsupervised machine learning. *NATURE MACHINE INTELLIGENCE*, 4(6):510–520, 2022.

[109] H. Z. Zhang, E. S. L. Ho, and H. P. H. Shum. Cp-agcn: Pytorch-based attention informed graph convolutional network for identifying infants at risk of cerebral palsy ? *SOFTWARE IMPACTS*, 14, 2022.

[110] J. Zhang, Y. Liu, A. P. Liu, Q. G. Xie, R. Ward, Z. J. Wang, and X. Chen. Multimodal image fusion via self-supervised transformer. *IEEE SENSORS JOURNAL*, 23(9):9796–9807, 2023.

[111] S. L. Zhang, F. Chen, J. F. Zhang, A. Q. Liu, and F. Wang. Multi-level self-supervised representation learning via triple-way attention fusion and local similarity optimization. *NEURAL PROCESSING LETTERS*, 55(5):5763–5781, 2023.

[112] W. T. Zhang, J. Peng, S. Zhao, W. L. Wu, J. J. Yang, J. Y. Ye, and S. S. Xu. Deep learning combined with radiomics for the classification of enlarged cervical lymph nodes. *JOURNAL OF CANCER RESEARCH AND CLINICAL ONCOLOGY*, 148(10):2773–2780, 2022.

[113] Y. Y. Zhang, Z. Q. Gong, W. E. Zhou, X. Y. Zhao, X. H. Zheng, and W. Yao. Multi-fidelity surrogate modeling for temperature field prediction using deep convolution neural network. *ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE*, 123, 2023.

[114] M. Zhou, M. Hamad, J. Weiss, A. Eslami, K. Huang, M. Maier, C. P. Lohmann, N. Navab, A. Knoll, and M. A. Nasseri. Towards robotic eye surgery: Marker-free, online hand-eye calibration using optical coherence tomography images. *IEEE Robotics and Automation Letters*, 3(4):3944–3951, 2018.

[115] P. F. Zhou, K. N. Ying, Z. H. Wang, D. Y. Guo, and C. Bai. Self-supervised enhancement for named entity disambiguation via multimodal graph convolution. *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, 2022.

[116] Q. Zhou and H. Zou. A layer-wise fusion network incorporating self-supervised learning for multimodal mr image synthesis. *FRONTIERS IN GENETICS*, 13, 2022.

[117] Tian Zhou, Jackie S. Cha, Glebys Gonzalez, Juan P. Wachs, Ch Sundaram, ru P., and Denny Yu. Multimodal physiological signals for workload prediction in robot-assisted surgery. *J. Hum.-Robot Interact.*, 9(2), 2020.

[118] C. Y. Zhu, Y. K. Wang, H. P. Chen, K. L. Gao, C. Shu, J. C. Wang, L. F. Yan, Y. G. Yang, F. Y. Xie, and J. Liu. A deep learning based framework for diagnosing multiple skin diseases in a clinical environment. *FRONTIERS IN MEDICINE*, 8, 2021.

[119] E. Özelçi, E. Etesami, L. A. Rohde, A. C. Oates, and M. S. Sakar. A robotic surgery platform for automated tissue micromanipulation in zebrafish embryos. *IEEE Robotics and Automation Letters*, 9(1):327–334, 2024.