

NOVEL EFFICIENT DEEP LEARNING ARCHITECTURES FOR MEDICAL IMAGE ANALYSIS

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
DATA SCIENCE

Submitted by

PAARTH PRASAD
(24/DSC/30)

Under the supervision of
Prof. Ruchika Malhotra

Delhi Technological University, Delhi



SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

JUNE, 2026

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, PAARTH PRASAD, Roll No's – 24/DSC/30 students of M.Tech (DATA SCIENCE), hereby declare that the Dissertation titled **"NOVEL EFFICIENT DEEP LEARNING ARCHITECTURES FOR MEDICAL IMAGE ANALYSIS"** which is submitted by me to the SOFTWARE ENGINEERING, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

PAARTH PRASAD

Date:

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**NOVEL EFFICIENT DEEP LEARNING ARCHITECTURES FOR MEDICAL IMAGE ANALYSIS**" which is submitted by PAARTH PRASAD, Roll No – 24/DSC/30, SOFTWARE ENGINEERING, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge, this work has not been submitted in part or in full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

Prof. RUCHIKA MALHOTRA
SUPERVISOR

ABSTRACT

Medical image analysis is now a significant field of research in the health care domain, because the requirement of more accurate, automated and efficient computational diagnosis systems. It is recently observed that the deep learning methods achieve remarkable success on medical image classification, segmentation, and disease detection, however, existing deep learning models are extremely complex and have high memory and runtime demand that can't be easily used in clinic practice. Beside, the anatomical consistency on medical image segmentation, particularly on 3D medical image applications, is still a challenge. In this thesis, the study develops new efficient deep learning architectures for medical image analysis, based on the investigation of the issues of computational efficiency, representation learning, and anatomical consistent segmentation. Two approaches are introduced: first approach propose a sparse attention based contrastive learning framework on medical image representation learning. Sparse attention helps reducing computational overhead and enhances feature extraction and learning from medical image. Contrastive learning is used to learn discriminative representation. Second approach proposed a topology constrained quantized nnU-Net framework for efficient and anatomically consistent 3D dental image segmentation. By adding topological constraints, anatomy consistency on segmentation is preserved while quantization on the model makes it computationally efficient and reduces model complexity and usage of hardware memory to be easily deployed in real application environments with less loss in segmentation performance. Experimental results show that the proposed models have comparable accuracy but much better computational efficiency and robust performance in terms of learning ability and structural accuracy in dental images segmentation applications. Therefore, these methods have important potential applications in automatic diagnosis system, clinical intelligent support systems, auxiliary system for radiologist, efficient healthcare systems, and so on. The potential future directions are multimodal medical imaging applications, federated learning, explainable ai, and lightweight deployment-oriented models for real time applications.

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I wish to express our sincerest gratitude to Prof. RUCHIKA MALHOTRA for his continuous guidance and mentorship that he provided me during the project. He showed me the path to achieve our targets by explaining all the tasks to be done and explained to me the importance of this project as well as its industrial relevance. He was always ready to help me and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi

PAARTH PRASAD

Date:

LIST OF PUBLICATIONS

Publications arising from the research work presented in this thesis are listed below in chronological order.

Conference Papers

C1. Submission ID : 961

Title: “SparseContrast: Dynamic Sparse Attention for Efficient and Accurate Contrastive Learning in Medical Imaging”

Conference: 4th International Conference on Ambient Intelligence in Health Care ICAIHC-2026

Organized by: National Institute of Technology, Raipur

Date: 18-19 September 2026

Authors: Paarth Prasad & Ruchika Malhotra

Status: [Communicating]

C2. Submission ID : 1138

Title: “Topology-Constrained Quantized nnUNet for Efficient and Anatomically Accurate 3D Tooth Segmentation”

Conference: 4th International Conference on Ambient Intelligence in Health Care ICAIHC-2026

Organized by: National Institute of Technology, Raipur

Date: 18-19 September 2026

Authors: Paarth Prasad & Ruchika Malhotra

Status: [Communicating]

CONTENTS

CANDIDATE'S DECLARATION	i
CERTIFICATE	ii
ABSTRACT	iii
ACKNOWLEDGEMENT	v
LIST OF FIGURES	vi
LIST OF TABLES	vii
Chapter 1: INTRODUCTION	1
1.1 Background of Medical Image Analysis	1
1.2 Deep Learning in Healthcare	2
1.3 Research Problem	3
1.4 Objectives	3
Chapter 2: LITERATURE REVIEW	5
2.1 Medical Imaging	5
2.2 Deep Learning Models	5
2.3 Attention Mechanisms	6
2.4 Contrastive Learning	6
2.5 Medical Image Segmentation	7
2.6 Research Gap	7

Chapter 3: SPARSE ATTENTION-BASED CONTRASTIVE FRAMEWORK	9
3.1 Introduction	9
3.2 Proposed Framework	10
3.3 Sparse Attention Mechanism	11
3.4 Contrastive Learning	13
3.5 Dataset and Preprocessing	13
3.6 Experimental Setup	14
3.7 Results and Discussion	14
Chapter 4: TOPOLOGY-CONSTRAINED QUANTIZED SEGMENTATION FRAMEWORK	16
4.1 Introduction	16
4.2 Proposed nnU-Net Framework	17
4.3 Topology Constraints	17
4.4 Quantization Techniques	19
4.5 Dataset and Preprocessing	20
4.6 Experimental Setup	21
Chapter 5: RESULTS AND DISCUSSION	22
Chapter 6: CONCLUSION AND FUTURE SCOPE	29
REFERENCES	30
Appendix	35

LIST OF FIGURES

Figure 3.1 Proposed Sparse Attention-Based Contrastive Learning Framework	11
Figure 3.2 Visualization of Sparse Attention Maps	12
Figure 4.1 Overall Training Workflow with Topological Constraint	17
Figure 4.2 Distribution of Topological Loss in Segmentation Masks	18
Figure 4.3 Relationship Between Quantization Error and Topological Loss	20
Figure 5.1 Training Time and Memory Consumption Comparison	23
Figure 5.2 Performance Under Different Labeled Data Settings	24
Figure 5.2 Performance Under Different Labeled Data Settings	25

LIST OF TABLES

Table 5.1 Disease Classification Performance and Computational Cost	26
Table 5.2 Ablation Study on Sparsity Mechanisms	26
Table 5.3 Performance Comparison Across Segmentation and Efficiency Metrics	27
Table 5.4 Ablation Study on Topological Loss Components	27
Table 5.5 Inference Speed Comparison	28

Chapter 1: INTRODUCTION

1.1 Background of Medical Image Analysis

Medical image analysis is becoming a core part of modern clinical practice, assisting practitioners in making informed decisions regarding diagnosis, treatment design, and disease management. The vast amounts of useful information extracted from medical images created by medical imaging modalities including X-ray, computed tomography (CT) scan, magnetic resonance imaging (MRI), ultrasound and cone beam CT (CBCT) images need to be processed correctly [1]-[2]. Manual review of medical images by radiologists and clinicians was once a traditional approach in the interpretation of medical images. However, it is both slow and susceptible to inter-observer variability, often resulting in a discrepancy of conclusions.

Artificial intelligence and deep learning are widely acknowledged as revolutionizing automated medical image analysis and a large body of literature exploring the performance of AI in medical image classification, detection of diseases and medical image segmentation has been published over the last decade. Many approaches for image analysis now utilize deep learning methods such as convolutional neural networks (CNNs), transformers and attention based architectures to perform these tasks [3, 14, 25]. These networks learn discriminative features directly from the raw medical image data in an end-to-end manner without the need for handcrafted feature descriptors commonly used in classical algorithms.

Medical image segmentation is a popular field of healthcare application within deep learning due to its significant implications for applications such as dental imaging, cancer diagnosis, surgery planning and treatment monitoring. Segmentation allows the precise delineation of organs, tumors, lesions or other clinically important regions of the diagnostic images [36]. One major drawback of many existing deep learning models is their enormous computational and memory overhead which prevents the adoption of such approaches in many resource limited situations [6], [15]. Furthermore many models lack a strict guarantee for topological or anatomical consistency in volumetric medical imaging data [39].

To mitigate these shortcomings several efficiency oriented methods were explored including sparse attention, contrastive self-supervised learning, topology-aware training, and quantization [12], [27], [37], [46]. Sparse attention limits attention computation to relevant tokens thereby minimizing computational overhead while still retaining meaningful context, contrastive learning learns more representative features by clustering semantically similar items and separating dissimilar items in the embedding space, topology-constrained methods apply constraints that ensure structural coherence and quantization reduces the size of network parameters drastically. In this thesis, we explore new efficient deep learning approaches for medical image analysis, including sparse attention-based contrastive learning to efficient medical image representation learning, and a topology-constrained quantized nnU-Net framework for efficient and anatomically accurate dental image segmentation.

1.2 Deep Learning in Healthcare

Deep learning techniques, as explained above, have revolutionised medical technology, enabling high-throughput, automated analysis of complicated streams of medical data. CNNs, Transformers and Attention-based architectures have achieved significant successes in tasks including classification, segmentation and diagnosis of medical images [3],[14].

Unlike many other domain engineering approaches, these networks learn end-to-end features directly from image data that is relevant to the given task. Recently, interest has grown around Self-supervised learning and Contrastive learning approaches, which allow representation learning with fewer labeled data points [27], [31]. Moreover, Segmentation architectures such as nnU-Net have been presented which achieve high accuracy in medical image segmentation [36].

However, state-of-the-art approaches are usually computationally very expensive, which requires high-performance GPU resources during both training and inference [37]. This lack of affordable deployability is undesirable, especially when used in a clinical or resource constrained environment.

1.3 Research Problem

While deep learning has achieved remarkable progress in medical imaging, some challenges still stand in the way for a large scale adoption of deep learning in clinical practice. The state-of-the-art methods require considerable amounts of GPU compute, large memory and long training time [6], [15]. The resource limitations bring many practical challenges in clinical applications.

One critical factor is the anatomically coherent nature of the segmentation, current methods usually contain geometric irregularities and poor boundary prediction which has been highlighted with the dental volumetric CBCT [39] [46]. Lack of medical data and expensive annotation reduces the generalization capabilities and performance of deep learning models on unseen images.

Therefore, there is a need to design lightweight, computationally efficient deep learning networks which simultaneously capture a better representation and provide anatomically precise segmentation.

1.4 Objectives

The thesis primarily intends to develop effective deep learning models for medical imaging analysis. The research addresses issues such as computational efficiency, feature learning, and anatomically appropriate segmentation for medical image analysis in the healthcare context.

The objectives of this research include:

- To study deep learning techniques used in medical image analysis.
- To develop a sparse attention-based contrastive learning framework for efficient medical image representation learning.
- To design a topology-constrained quantized nnU-Net framework for efficient dental image segmentation.
- To improve computational efficiency and reduce model complexity in medical imaging tasks.

- To evaluate the performance of the proposed approaches using medical imaging datasets and standard evaluation metrics.

Chapter 2: LITERATURE REVIEW

2.1 Medical Imaging

Medical imaging provides critical information about the anatomy and pathology of patients. The existing set of modern medical imaging technologies that incorporate X-ray, CT, MR, Ultrasound and cone beam CT is being extensively employed in medicine for the diagnosis and treatment of medical conditions [1], [2].

Imaging provides anatomical and pathological information. As a result, the adoption of digital imaging acquisition techniques has given rise to a vast number of medical images acquired every day. Analysis of such imagery by human users is extremely labor intensive and needs highly skilled professionals; therefore a fully automated image analysis pipeline is a necessity for the contemporary clinical setup [3]. AI and deep learning based methods are gaining rapid popularity in enhancing diagnostic accuracy and minimizing physician work.

2.2 Deep Learning Models

Deep learning has enabled several major breakthroughs in computer vision and medical imaging, and has substantially advanced state-of-the-art in previously difficult problem areas. CNNs are one of the most prevalent architecture types used in this area due to their inductive bias of being able to learn image feature representations hierarchically in space [14], [24].

Furthermore, Transformer models have become more popular due to their impressive capability of modeling long-range relationships and context in an image [15], [25]. Attention modules further boost network capability by allowing the network to learn to dynamically weight the relevance of different regions of an image, to favor areas that are more useful for making a diagnosis [17]. This methodology is employed to detect various diseases, chest x-ray analysis, organ segmentation, and dental studies [21], [40].

Self-supervised and contrastive learning paradigms have gained significant research interest as they allow effective feature extraction that is transferable to new problems from unannotated or

minimally annotated data [27], [31]. These are highly desirable in medical imaging due to the difficulty and high cost of acquiring labeled data by medical experts.

2.3 Attention Mechanisms

Attention mechanisms have quickly become a building block in deep learning architectures today, acting as mechanisms that allow networks to focus on informative features during the extraction and representation stages. Through their modulations on feature importance, attention modules focus networks on clinically important regions and prevent them from picking up on noise within irrelevant regions of background [17]. The ability to concentrate on important regions through attention has resulted in excellent results in medical image classification, segmentation, and anomaly detection tasks.

One of the major downfalls of the standard dense attention is the quadratic scaling with respect to the sequence length, which becomes impossible to accommodate for both memory and computation given the huge volumes associated with high-resolution medical imaging [6], [15]. In order to counter these problems, several sparse attention architectures have been presented which restrict attention computation to only carefully chosen tokens pairs and preserves global context at a lower cost [12]. These transformer architectures become much more feasible to implement for the massive amounts of healthcare imaging, allowing for application on environments with limited compute.

The use of sparse attention within medical imaging tasks to boost efficiency and performance has only been widely explored recently [13].

2.4 Contrastive Learning

Contrastive learning is a self-supervised paradigm that trains models to embed semantically similar samples in proximity while separating dissimilar ones, yielding highly discriminative and transferable representations [27]. These methods are often particularly applicable in medical imaging where labeled examples are not often plentiful for a supervised approach. Many works have already presented contrastive learning methods on chest x-ray, ct- scans, and other medical contrastive learning is a paradigm in self-supervised learning that train the model to embed

semantically-related samples close to each other while pushing dissimilar samples farther apart, thereby extracting a useful and portable representation [27]. Such methods become meaningful in training medical imaging tasks where fully supervised models are sometimes not available due to lack of labeled examples.

Numerous studies investigate contrastive learning strategies applied to chest x-rays, CT scans and other medical images [7], [8], [10]. In general, they enable medical imaging models to learn transferable representations that are resistant to medical changes. In combination with transformer architectures and attention mechanisms, it could further enhance representation power of learned features [31], [32]. Using contrastive objectives along with sparse attention architectures seems to be a promising avenue for learning representations efficiently without any drop in representational quality for clinical imaging data.

2.5 Medical Image Segmentation

Of the various applications of deep learning to medicine, image segmentation is among the most significant due to its immediate clinical applicability. The algorithms themselves distinguish anatomical structures such as tissues, pathological lesions, and organs within the captured image [36].

While self-configurable and adaptable architectures such as nnU-Net have performed very well on image segmentation applications on medical images [36], common shortcomings with segmentation networks are often topological incoherencies of the prediction, as well as the computational load placed on devices with the use of 3D volumes such as dental CBCT [39] and [40]. Topological learning addresses issues of continuity by maintaining it during training [46] while quantization has been applied to decrease model size, memory, and the computational requirements needed in an end-use scenario [37]. These new developments show promise as the optimization of topology and computational cost has been correlated with clinical performance [44] and [45].

2.6 Research Gap

The aforementioned progresses notwithstanding, the application of deep learning to medical images continues to face several outstanding technical and practical issues that are not yet satisfactorily solved. In fact, a great number of current architectures demand hardware resources that are not commonly available in either clinical or edge settings [6], [37], and to date a consistent volumetric segmentation result with structural and topological preservation has not been achieved yet [39], [46].

From the review, there are not many existing works that combine, in an integrated manner, the notion of sparse attention and contrastive learning, such that to obtain a medical image representation. Concomitantly, there is an analogous lack of specific lightweight segmentation architectures that can perform topologically-informed medical image segmentation without high inference cost. Both technical problems are addressed in this thesis through the introduction of two complementary efficient architectures: a sparse-attention driven framework that exploits contrastive learning for medical image representation, and a topology-aware quantized medical image segmentation system.

Chapter 3: SPARSE ATTENTION-BASED CONTRASTIVE FRAMEWORK

3.1 Introduction

In the field of medical image analysis, deep learning models are expected to be effective and robust to learn meaningful representations from complex medical data sets. Self-supervised approaches learn valuable feature representations by exploring the structure of unlabeled data, among which contrastive learning has proved to be a powerful strategy that learned transferable representations with constant improvement in both general computer vision and medical imaging benchmarks [27, 31]. By attracting representations of data augmentation views from the same sample and repelling those from negative samples, an embedding objective is used. Due to the high cost of acquisition and the scarcity of labeled medical image datasets, contrastive methods are particularly well-suited for healthcare.

Transformer networks, due to their ability of modeling long-range relationships through multi-head self-attention [15, 25], have become more and more popular in medical image analysis. However, due to the quadratic complexity in sequence length, full-attention transformer networks consume large memory and have long inference times in relation to image resolution [6], which hinders the application to large-scale medical image tasks.

In order to address the efficiency and representation learning issues simultaneously, this paper proposes a sparsely attended contrastive learning framework for medical image analysis. The model connects a sparsified self-attention module and contrastive training, so as to reduce the computation without damaging the quality of learned representations. Since only diagnostically important regions and their context are attended, sparse attention avoids redundant computation compared to dense attention [12]. This framework achieves both efficiency improvement and feature quality augmentation, which facilitates medical image analysis to better provide reliable performance while achieving fast processing. We conduct experiments on medical image datasets to evaluate the proposed framework.

3.2 Proposed Framework

In order to realize medical image learning efficiently with high-fidelity, we utilize sparse self-attention coupled with a contrastive objective. In this setup, we increase the discriminative ability of the learnt features while significantly reducing the high computation load involved with the dense attention mechanisms present in many architectures. The process is formulated into 5 stages; input image preprocessing, feature extraction, sparse attention calculation, contrastive learning and finally optimization. Input images are first pre-processed and normalized to enter the deep learning network pipeline. Positive pairs required by the contrastive objective are generated through stochastically different augmentation pipelines providing semantically similar views of the image and features are extracted through a deep neural network.

We integrate sparse attention into the encoder to limit the attention computation to highly relevant spatial locations, avoiding unnecessary computation over uninformative regions [6], [12]. This selective attention computation method allows us to significantly reduce GPU memory and float operations while preserving important information for diagnostics. After obtaining the feature representation of the input image, we apply the contrastive objective to boost the quality of the representations. The two augmented views of the same image would be attracted in the latent space by the contrastive loss while the embeddings of different images repel each other to ensure class separation [27]. This objective learns representation which is robust and discriminative.

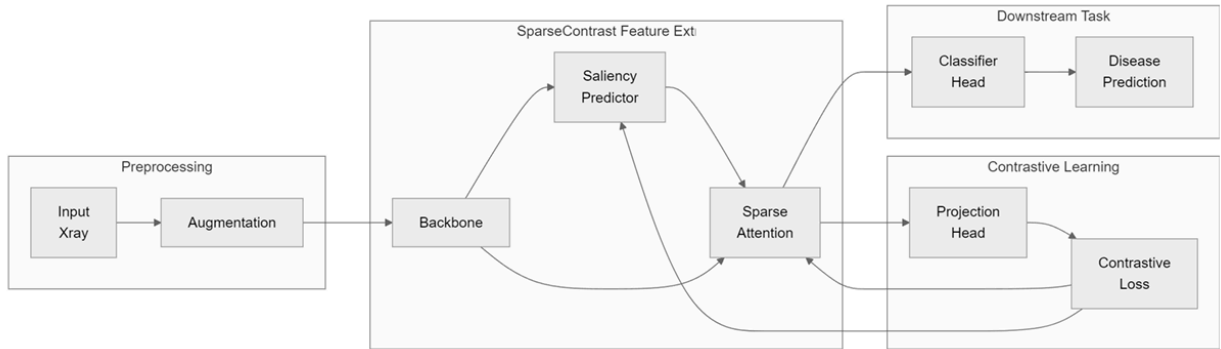


Figure 3.1 Proposed Sparse Attention-Based Contrastive Learning Framework

Figure 3.1 shows the architecture of the proposed SparseContrast framework. Our framework exploits sparse attention and contrastive learning for effective learning of medical image representations.

3.3 Sparse Attention Mechanism

Attention mechanisms have become a cornerstone of modern deep architectures, providing models with the ability to selectively highlight features of a taskrelevant nature within the vast high-dimensional input spaces [17]. In their dense form, attention has a quadratic dependency with the number of input tokens; this prohibits its use with large resolution medical volumes and imaging cohorts [15].

Sparse attention mechanisms are proposed within this framework. Sparse attention reduces computational cost by limiting the attention graph to a subset of semantically important interactions between tokens (identified through some learned or heuristic selection mechanism [6]). This sparse computation maintains enough contextual support to be diagnostically useful, while also dramatically reduces FLOPs and memory usage.

In our approach, the sparse attention mechanism prioritizes computation within areas of interest relevant to clinical diagnoses, while suppressing contributions from background and low contrast areas. The superfluous computations are omitted, which allows the framework to scale well with very large, diverse medical image datasets. In addition to memory requirements, sparse attention decreases both training convergence speed and inference speed, which further promotes usage within critical medical scenarios [12].

In conclusion, this sparse attention is efficient, providing an advantage over dense attention in terms of saving computation resources while having representational power comparable to the dense baselines. This framework leverages sparse attention to efficiently develop sophisticated features at no additional performance cost.

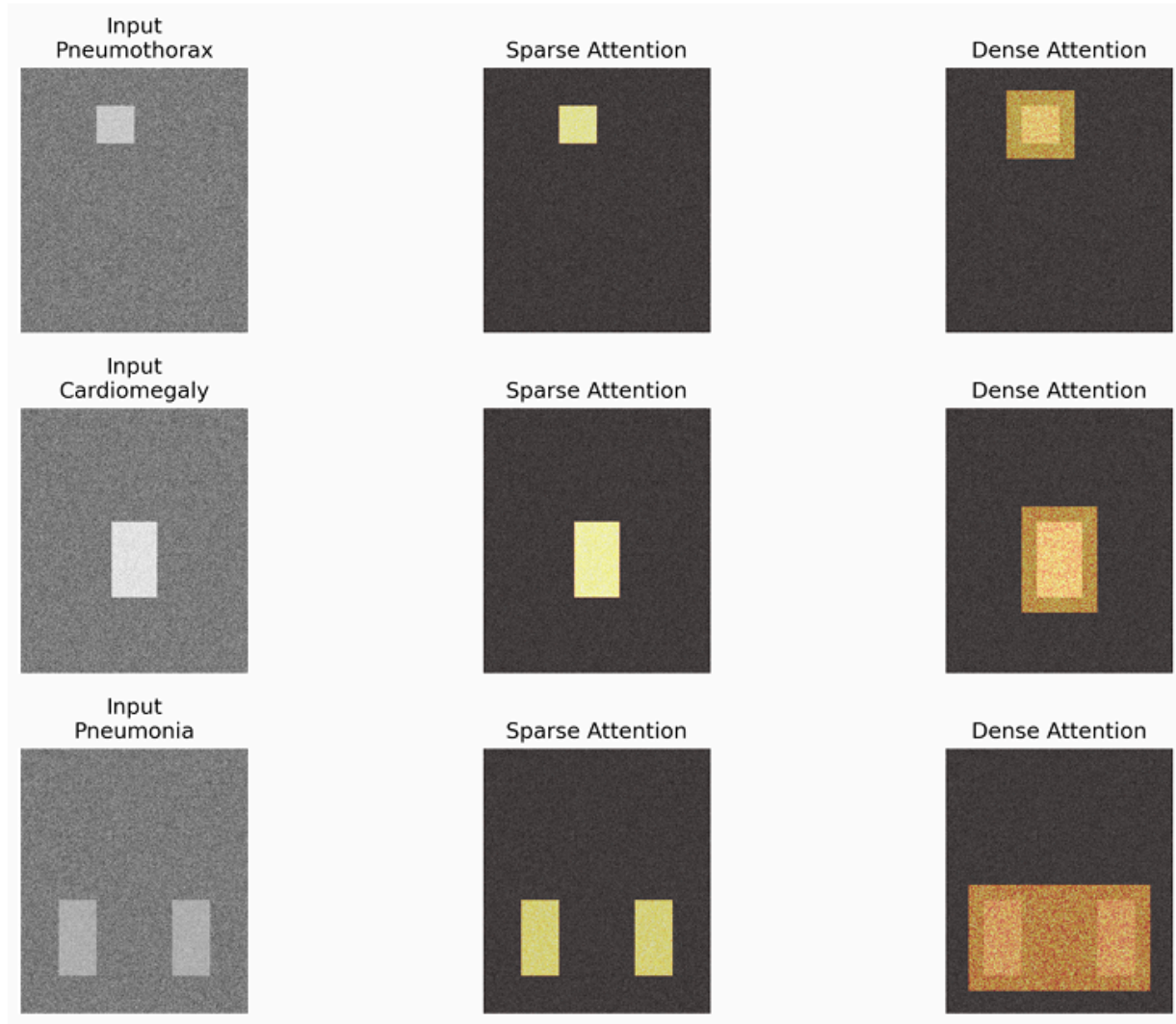


Figure 3.2 Visualization of Sparse Attention Maps

Figure 3.2 visualize the sparsely attended region produced by each thoracic disease. The visualization show that the framework can pinpoint clinically important diagnostically relevant regions.

3.4 Contrastive Learning

Contrastive learning is a self-supervised framework which learns by maximizing the similarity of positive pairs versus the dissimilarity of negative pairs in the same embedding space [27]. This is a promising avenue for achieving effective representations without a dense amount of

supervision as it is often prohibitively expensive to gather annotated data for the medical imaging domain.

Contrastive learning is a self-supervised paradigm that learns representations by comparing similarities of positive pairs to dissimilarities of negative pairs within a common embedding space [27]. Due to the perennial lack and high price of medical data, learning powerful representations by way of a contrastive approach can overcome reliance on dense annotation. The introduced model utilizes the approach of contrastive learning to learn useful feature embeddings from medical images. Augmentation based views of same image instance define positive pairs, whereas view based negative pairs belong to different images. The contrastive objective is utilized to learn embeddings that pull positives closer while pushing negatives apart in the feature space. Learning such embeddings promotes representations which are invariant to innocuous imaging variability yet responsive to variations relevant to disease, enabling enhanced classification or anomaly localization[31].

When such objective is jointly trained with a sparse attention, further computational benefits can be achieved by elimination of needless calculation during feature extraction. A number of works indicate the ability of contrastive learning to improve medical images representations and generalizability of models [7][8][10]. Hence, the introduced model uses sparse attention together with contrastive learning in learning useful feature representations for medical images.

3.5 Dataset and Preprocessing

Empirical evaluation is carried out on several public medical imaging benchmarks, primarily chest radiograph collections and examines the framework under varying settings of representation learning. Due to the differing modalities, resolution, contrast, and coverage of body parts between the various datasets, extensive preprocessing is performed to ensure a stable and well-behaved model.

All images were first spatially resampled to a single resolution before entering the network. All samples were intensity normalized in order to minimize the effect of inter-scanner variance and guarantee a constant input distribution. Several random augmentation operations were applied:

horizontal flipping, random shifts in angle of rotation and translation as well as random cropping. Photometric jitter (random changes to brightness, contrast and saturation) was also used to reduce sensitivity to differences in exposure.

These preprocessing and augmentation steps, therefore, ensured the effective learning and generalization capability of the network in varied conditions and patient populations. Pairs of these augmented images are then presented to the contrastive learning module to train the representation space.

3.6 Experimental Setup

This framework is built with widely available deep learning libraries and is tested with GPU accelerated computation environment. This framework is validated over well-known medical imaging benchmarks using different splits for training, validation and hold out test. The model is trained with Adam optimizer [26].

Hyper parameters of this model, such as batch size, learning rate schedule, number of epochs and contrastive temperature, were determined through a process of systematic ablation with the help of validation set. The task of representation learning at the training phase is done by following a contrastive loss function.

Model performance is measured by standard accuracy metrics, such as accuracy, precision, recall and evaluation of embeddings. GPU memory usage, latency of inference and computational burden of training are measured through empirical examination of the architecture.

3.7 Results and Discussion

Experiments show that the proposed method learns discriminative medical images representations and drastically reduces the computational burden as compared to the dense-attention baselines. Integrating sparse attention makes the model efficiently operate by omitting meaningless cross-token computations and also by reducing the peak GPU memory required.

The contrastive training framework reinforces the discriminative structure in the learned embedding space and enhances the robustness of the model towards imaging noise and distribution shifts. Evaluating with several well-established baselines demonstrates that our method achieves equal or superior representational power compared to other methods, but requires much less computational power.

The proposed design of sparse attention directly translates to less inference latency, reduced energy consumption, hence boosting its viability in low-bandwidth or edge computing situations within a clinical setting. In addition to that, the learned embeddings can be well transferred to different modalities and imaging conditions.

Chapter 4: TOPOLOGY-CONSTRAINED QUANTIZED SEGMENTATION FRAMEWORK

4.1 Introduction

Medical image segmentation is one of the most clinically relevant applications in medical imaging as it precisely outlines anatomical structures such as organs, tissues, pathological abnormalities, and pathological lesions. Accurate segmentation results enable a variety of clinically important processes such as oncological staging, pre-operative planning, treatment tracking over time, and assessment of dental condition [36]. Over recent years, the performance of deep learning segmentation networks has dramatically advanced the state-of-the-art in segmentation by automatically learning complex spatial and morphological features from annotated training images.

Among the commonly used deep learning architectures in medical image segmentation, nnU-Net stands out and gained popularity among the biomedical image segmentation communities since its self-configuring capabilities coupled with impressive segmentation results [36]. However, models of this level of sophistication still suffers from the inherent shortcomings in computational efficiency, memory demands and the generation of anatomically consistent output especially for the 3D dental image segmentation [39], [40].

One of the main limitations of the current approaches in the CBCT dental segmentation relies on the complicated morphology of dental structures and their neighboring position in anatomy, which is usually separated by a thin layer of cortical bone [42]. Existing deep learning models fail to segment accurately in the boundary or have the issue with topographically inconsistent output which goes against the intrinsic topological characteristic of the teeth [44], [45]. The other drawback that makes current top performance segmentation models un-deployable at the clinical bed side or at the point of care centers lies on the high memory footprint and computation power required by these models [37].

Herein, we present the topology-constrained quantized nnU-Net model where topological constraints are integrated to preserve the integrity of the anatomy while quantization technique is adopted to enhance computational efficiency.

4.2 Proposed nnU-Net Framework

We use nnU-Net [36] as our architecture: a robust self-configuring segmentation model which adaptively learns preprocessing, network architecture, and training recipe according to the statistical characteristics of a target dataset. Besides, we embed topological losses and adopt training using quantization-aware approaches in this architecture in order to obtain greater anatomical accuracy and greater inference speed.

In the encoder-decoder backbone strided convolutions learn features of different resolutions, which then are passed to skip connected decoders and output the masks of the dental structures. For the purpose of model footprint reduction, we adopt quantization-aware training, which transforms weights into a low-bit width format.

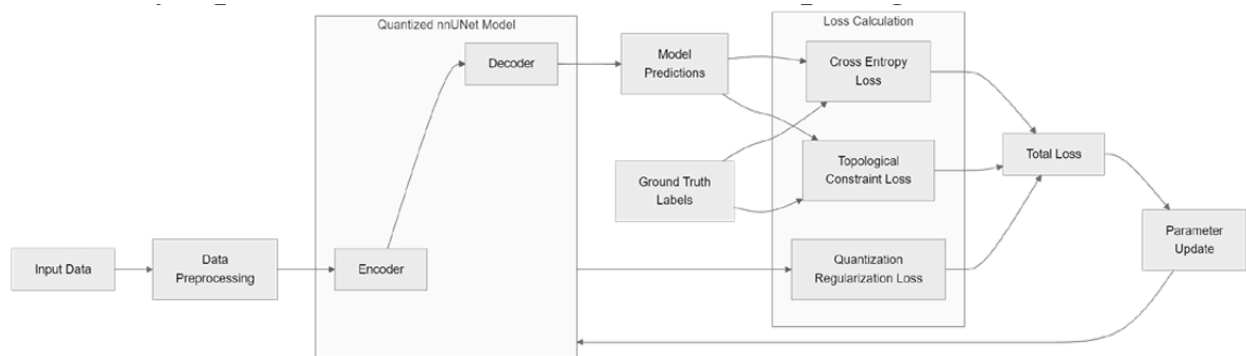


Figure 4.1 Overall Training Workflow with Topological Constraint

Figure 4.1 shows the entire training procedure of the proposed topology-constrained quantized nn-U-Net model for dental image segmentation.

4.3 Topology Constraints

Our architecture is built upon nnU-Net which is a very strong self-configuring segmentation model which adaptively learns preprocessing, network architecture and training recipes based on

the statistical characteristics of a target dataset [36]. To achieve better anatomical precision and faster inference we add topological loss terms and apply quantization-aware training to this framework.

The process is first started by pre-processing CBCT dental images and standardizing them, and is then fed to segmentation network. In the encoder-decoder backbone, strided convolutions were used for the multi-resolution features and decoders with skip connection processing to produce the masks of the dental structure. Quantization-aware training technique is used for bit width reducing of weights in order to reduce the size of model.

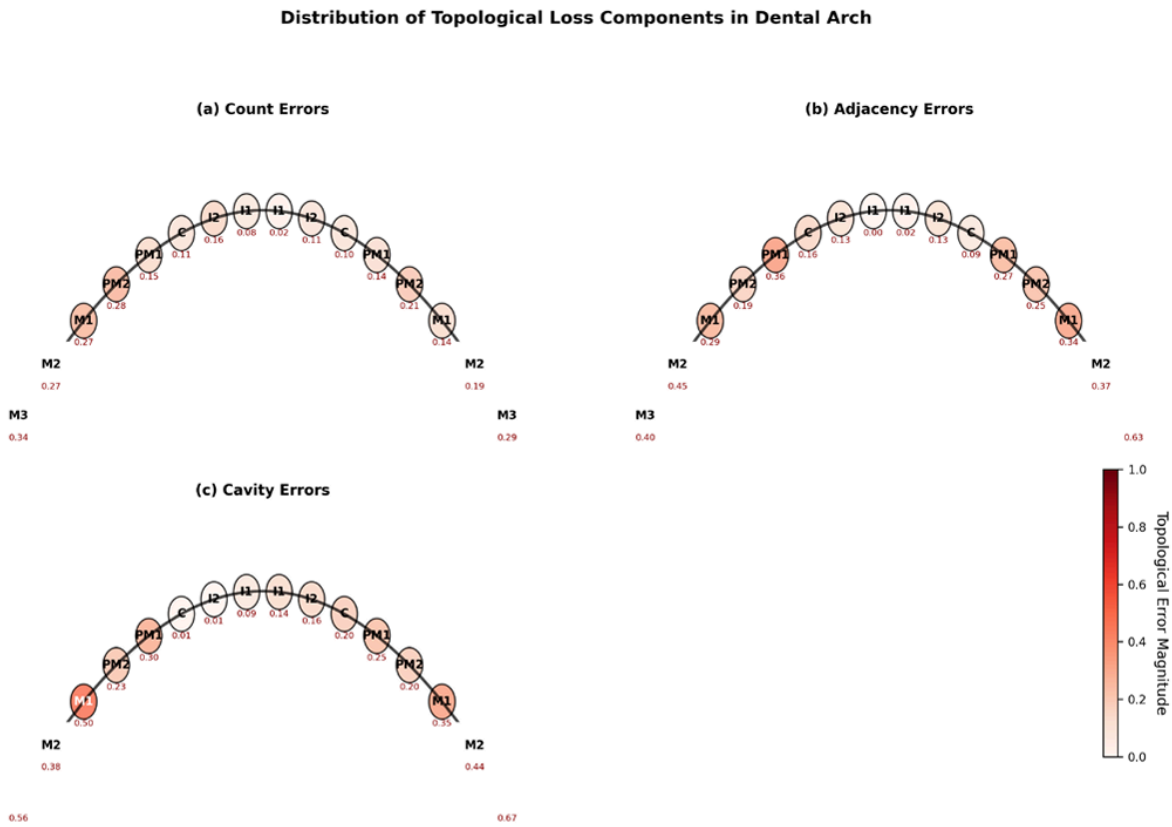


Figure 4.2 Distribution of Topological Loss in Segmentation Masks

Figure 4.2 presents the topological loss distribution in the process of tooth segmentation. It shows the significance of topological preservation in teaching the anatomical context of 3D models.

4.4 Quantization Techniques

Many contemporary deep segmentation networks require computationally demanding levels of memory and resources, presenting limitations on deployment within clinical environments due to hardware constraints in clinical workstations. Quantization allows this by converting network weights and activations from 32-bit floating point representations to fixed low-bit integers [37], thereby reducing memory storage requirements, and enabling computations to be carried out using computationally cheaper integer arithmetic operations.

The proposed system implements quantization-aware training across the entire network. This allows for the entire network to be compressed without significant losses to accuracy through the addition of learnable quantization parameters to the system. This compression resulted in a model with significantly reduced memory requirements, while still producing acceptable segmentation accuracy.

Memory reduction has a general application in low computational power situations such as intraoral scanners or mobile medical imaging devices. As well as increased memory capabilities, integer computation results in increased throughput of inference and reduced energy consumption per pass, and thus is perfectly suitable for the deployment of the model into an actual clinical workflow.

This system will tackle the efficiency/accuracy trade off, achieve a usable segmentation accuracy, using minimal computation.

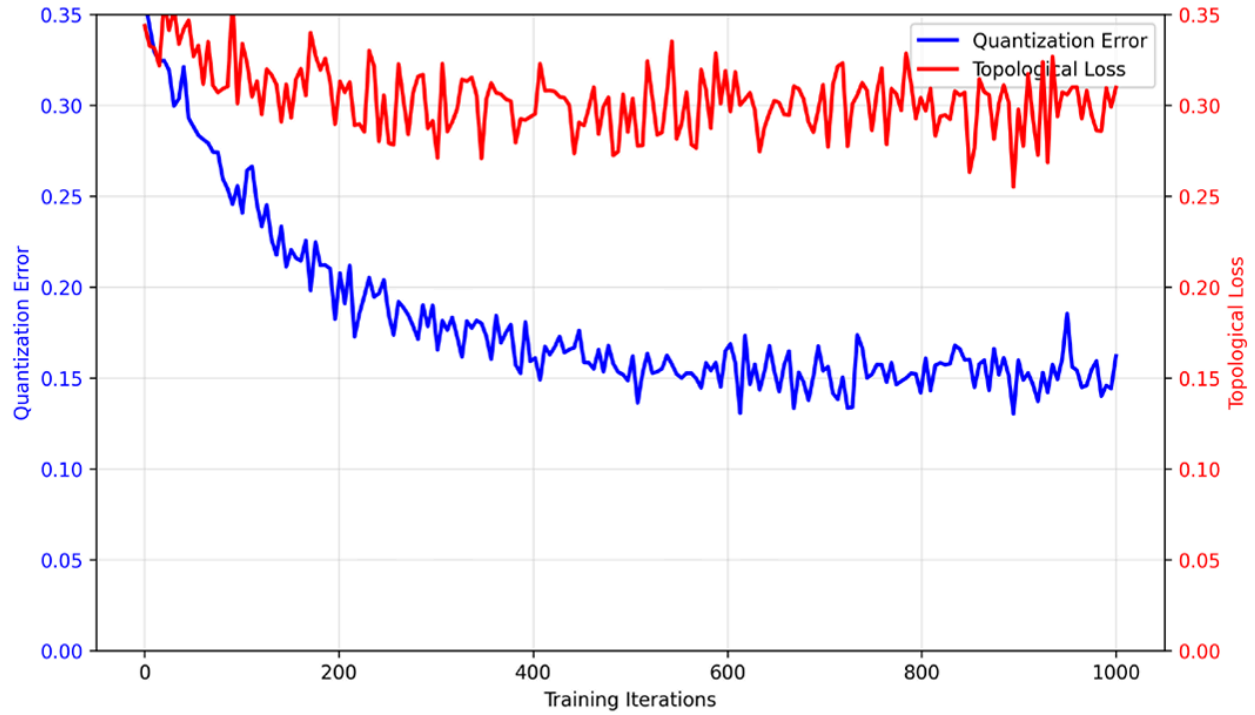


Figure 4.3 Relationship Between Quantization Error and Topological Loss

Quantization error against topological loss during training is shown in the Figure 4.3. Examination of the graph clearly presents the relationship between efficiency and the segmentation quality.

4.5 Dataset and Preprocessing

Performance is measured over a test set of 3D CBCT dental volumes and provide an objective and clinically driven way of validating the performance of the method. Unlike panoramic views CBCT scans have better spatial resolution, they are also acquired using iso-voxels and as a result is considered the most suitable imaging modality for detailed dental imaging applications, such as implant placement and orthodontic treatment [40].

To ensure the standardisation of images, and minimise variation between scans before feeding into a network a set of identical preprocessing steps were carried out on each image; the scans were resampled to have a standard voxel spacing and their intensities normalized before being

fed into the segmentation network. Additionally denoising and contrast normalisation was also applied to minimize scan acquisition specific artifacts from interfering with model performance.

When the model is trained it can also be augmented through rotation, reflection and scaling of the volumes which helps to create more variations within the training data and minimize over-fitting; this is common practice for such data augmentation methods as it will help to create a more generalized model that will account for variability within the patient anatomy and scan parameters (e.g. Different scan protocols and different scanner calibration).

4.6 Experimental Setup

The implementation leverages commonly-used deep learning libraries and uses multi-GPU distributed training since volumetric segmentation has very high memory usage and compute intensity. For an unbiased estimation of accuracy, experiments are carried out on predefined train and held out test split of the dental CBCT data.

The segmentation network is trained with appropriate loss function and training scheme. Quantization aware training approach is applied throughout the optimization process, so that the model has enough time to adapt itself to the quantization error and retain the segmentation accuracy in training process.

We report the segmentation accuracy in terms of Dice similarity coefficient, precision, recall and Intersection over union to show boundary and volume agreement in different perspectives. Also we profile efficiency including peak GPU memory, compressed model size in MB, and per-volume inference latency for different hardware configurations.

The experiments are designed to investigate the segmentation accuracy as well as the deployment efficiency of the topology-constrained quantized segmentation approach.

Chapter 5: RESULTS AND DISCUSSION

The proposed frameworks were thoroughly evaluated on several benchmark medical imaging datasets regarding their representational ability, computational efficiency and deployability. Experiments were performed on the proposed SparseContrast framework and the topology-constrained quantized nnU-Net under different imaging modalities and data configurations.

SparseContrast achieved strong representation learning capabilities while significantly decreasing computational complexity arising from dense attention mechanisms of typical transformer models. Sparseness of attention removed redundant cross-token computations resulting in concrete improvements in training throughput and inference memory. Contrasted training objectives supported sparse attention by boosting the representational richness by utilizing discriminative structures learnt from unlabeled medical images. In comparison to dense attention models, SparseContrast reached the same or higher performance on downstream classification tasks while significantly reducing GFLOPs and peak memory usage.

Qualitative analysis on attention maps demonstrated that the mechanism successfully attended to regions of diagnostic relevance while ignoring irrelevant background. Such observation resulted in scalability improvements on large medical image cohorts and latency reduction per image. Furthermore, learnt representation has higher cross-domain transferability and less impact from different scanner, acquisition protocol or patient population variation.

Topology-constrained quantized nnU-Net generated clinically viable 3D dental CBCT segmentations while exhibiting noticeable improvement in the structural coherence compared to non-topological baselines. Topological regularization enforced more connected and coherent masks which reduced the generation of disconnected dental fragments or non-anatomical components. The algorithm reliably preserved boundaries of key dental anatomical structures, resulting in reduction of disconnected regions and boundary discontinuities. The models are quantized aware trained for 4 reduction in size and significant speedup on all evaluated hardware

platforms with minimal compromise on Dice scores. The lightweight model is well-suited for deployment in the clinical environment, where limited GPU memory is a concern, on workstations or even on edge devices for clinical analysis, implant planning, orthodontic assessment, and CAD tasks.

Systematic benchmarking demonstrates that both models are capable of reaching or exceeding representative state-of-the-art counterparts on their respective tasks while at the same time providing efficiency gains. SparseContrast contributes an efficient model for self-supervised representation learning. Topology-constrained quantized nnU-Net push the state-of-the-art in the efficiency-accuracy trade-off for 3D dental segmentation. Both proposed methods offer significant contributions towards building AI models in medical imaging which are efficient, deployable and clinically valuable by incorporating explicit efficiency optimization into the architecture design stage.

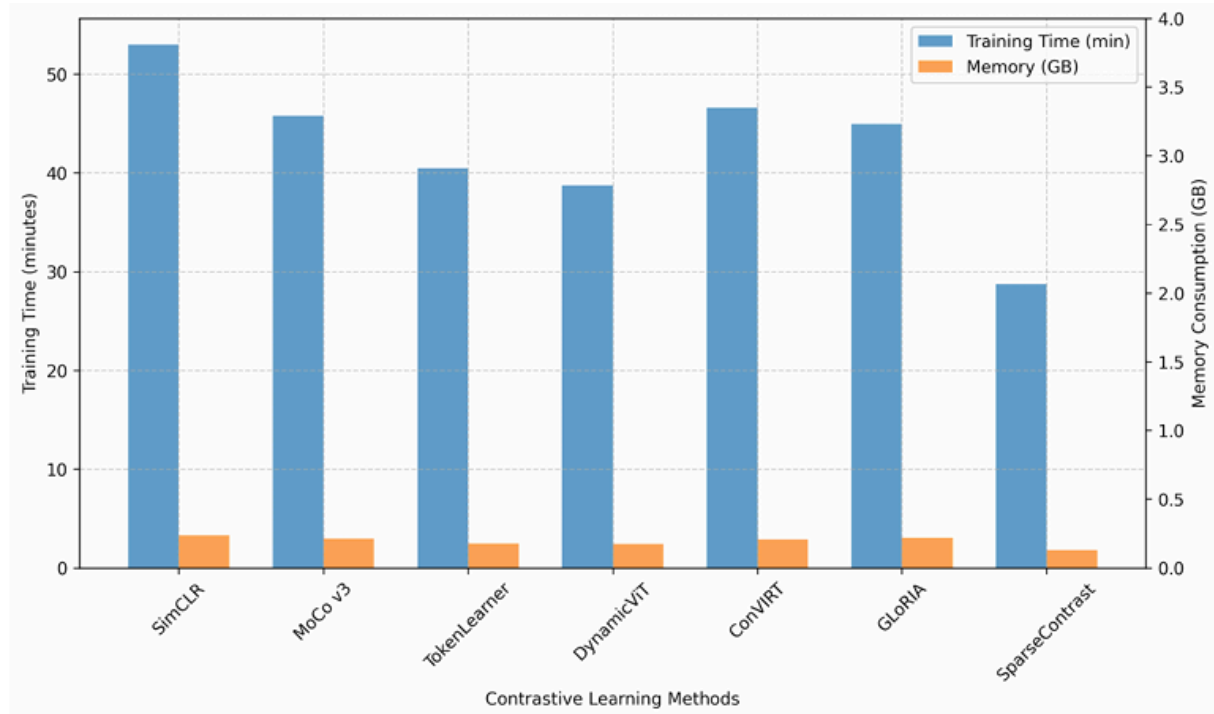


Figure 5.1 Training Time and Memory Consumption Comparison

Figure 5.1 Comparing the training time and memory usage of our proposed framework and existing contrastive learning approaches. Our proposed method has better computational efficiency.

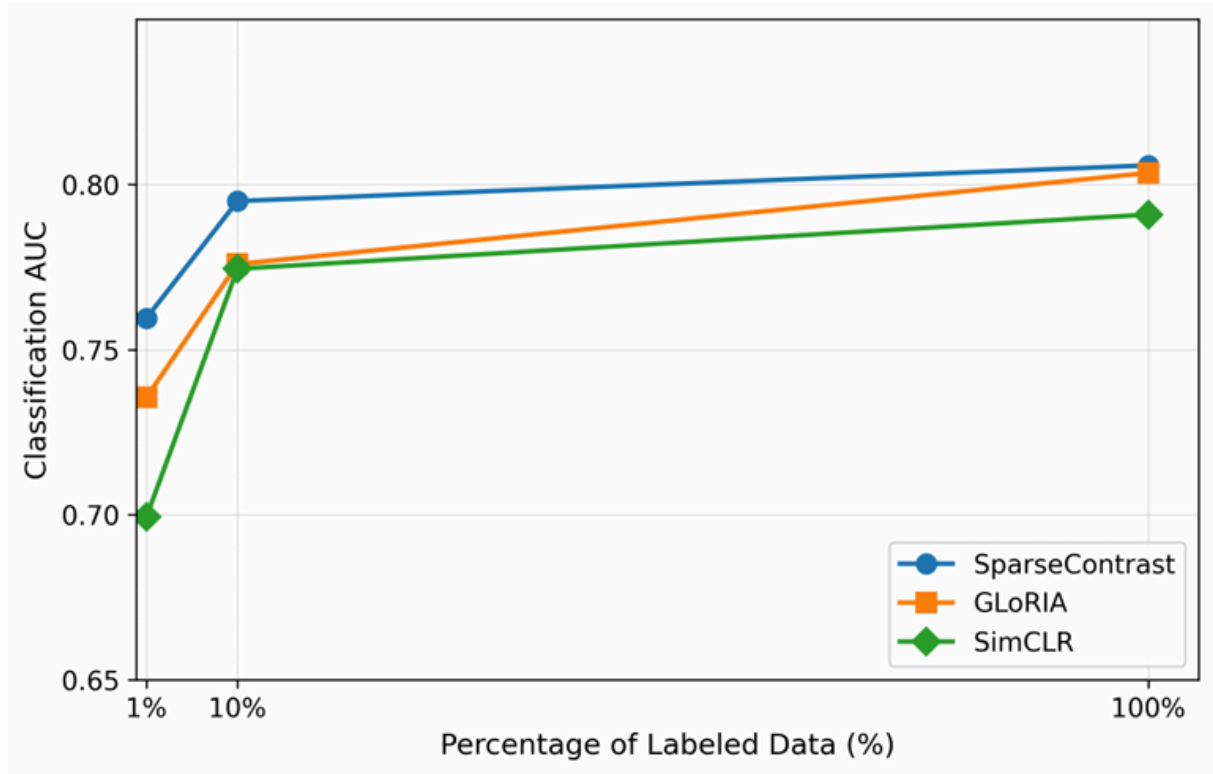


Figure 5.2 Performance Under Different Labeled Data Settings

Figure 5.2 shows the performance of our proposed framework using different numbers of labeled training data. Our system works well using small numbers of label samples.

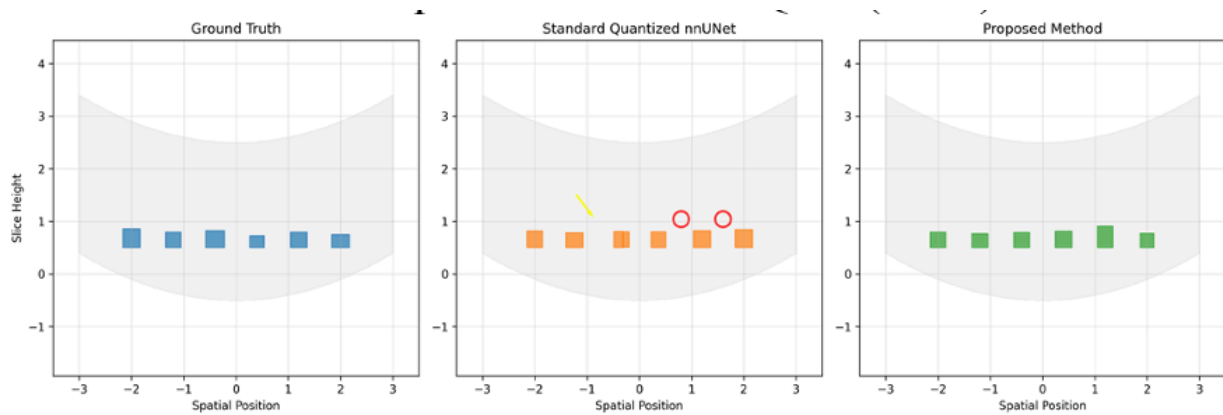


Figure 5.3 Comparison of 3D Tooth Segmentation Results

Figure 5.3 compares segmentation outputs generated by the ground-truth, standard quantized nnU-Net, and the proposed framework. The proposed method demonstrates improved structural consistency.

Method	CheXpert	MIMIC-CXR	NIH-14	GFLOPs	Memory (GB)
SimCLR	0.79	0.784	0.772	72.5	3.2
MoCo v3	0.80	0.788	0.776	70.8	3.1
TokenLearner	0.80	0.781	0.768	58.4	2.6
DynamicViT	0.80	0.792	0.779	53.7	2.4
ConVIRT	0.80	0.790	0.774	65.2	2.9
GLoRIA	0.80	0.793	0.781	68.3	3.0
SparseContrast	0.81	0.797	0.785	41.2	1.9

Table 5.1 Disease Classification Performance and Computational Cost

Table 5.1 summarizes the classification accuracy and efficiency of the proposed SparseContrast framework on chest X-ray datasets.

Variant	$\rho=0.1$	$\rho=0.3$	$\rho=0.5$	Dense
Static	0.782	0.792	0.799	-
Dynamic	0.789	0.812	0.808	0.798

Table 5.2 Ablation Study on Sparsity Mechanisms

Table 5.2 displays the performance of different sparsity mechanisms on the model using the CheXpert dataset. These findings suggest that sparse attention under the proposed framework is consistent.

Method	DSC (%)	IoU (%)	BF1 (%)	TCA (%)	ACS (%)	CER (%)	Size (M B)	Time (s)
Full Precision nnUNet	92.3	86.1	89.7	94.2	91.5	3.1	1024	8.2
Post Training Quant	88.7	80.2	83.4	82.6	78.3	12.8	256	2.1
QAT nnUNet	90.1	82.3	86.2	85.4	83.7	9.5	256	2.3
TopoNet	91.8	85.2	88.9	93.1	90.2	4.3	896	7.5
Proposed	91.5	84.9	88.6	93.8	91.0	3.9	256	2.4

Table 5.3 Performance Comparison Across Segmentation and Efficiency Metrics

Table 5.3 presents the segmentation accuracy, topological consistency, and computational efficiency of the proposed framework.

Configuration	DSC (%)	TCA (%)	ACS (%)	CER (%)
QAT-only	90.1	85.4	83.7	9.5
+ Count Loss	90.3	89.2	85.1	8.7
+ Adjacency Loss	90.8	91.6	89.3	6.4
+ Cavity Loss	90.6	90.8	88.5	5.1

Full Topo Loss	91.5	93.8	91.0	3.9
----------------	------	------	------	-----

Table 5.4 Ablation Study on Topological Loss Components

Table 5.4 illustrates the extent to which different portions of the topological loss impact segmentation accuracy and atlas consistency.

Method	A100 GPU	Xeon CPU	Jetson AGX
Full-Precision	1.2	8.2	14.7
Post-Training	0.4	2.1	3.8
QAT-nnUNet	0.5	2.3	4.1
Proposed	0.5	2.4	4.2

Table 5.5 Inference Speed Comparison

Table 5.5 shows the inference time using our framework versus typical segmentation methods; the results show faster computational speed.

Chapter 6: CONCLUSION AND FUTURE SCOPE

In this thesis we developed two novel deep learning architectures which can be combined to improve medical image analysis tasks by tackling the challenges of computational efficiency, the quality of self-supervised representations and topologically-preserving volumetric segmentation at the same time. Two complementary directions were pursued: a sparse attention-enhanced contrastive learning framework for efficient representation learning and a quantized topology-aware extension of nnU-Net for 3D dental segmentation.

In the first approach (SparseContrast) we show that it is possible to create computationally efficient encoders that produce discriminative, transferable representations for medical images by combining contrastive learning with sparse self-attention mechanisms. We can significantly reduce both FLOPs and memory usage by only calculating the attention between a subset of potentially interacting tokens selected by a learned attention mechanism. This allows us to extract fine-grained contextual information from the images. The contrastive learning objective helps the model to produce an embedding space with high discriminability and allows for cross-modal, cross-task transferability and robustness against anatomical variations.

In the second approach (topology-aware, quantized nnU-Net) we adapt nnU-Net with topological loss components and quantisation aware training to create a small, 3D dental segmentation network which is topology-preserving but extremely efficient. Topology-aware components can be used to impose constraints onto the resulting segmentation mask, thereby forcing it to preserve anatomical connectivity and avoid holes. Quantisation further reduces the model size by about factor 4 and enables correspondingly faster inference and allows easy deployment at the edge or in the point of care environment.

Through empirical validation on several tasks we show that both approaches can achieve task performance that competes with that of the much more computationally intensive full-precision dense-attention baselines. This shows the importance of designing systems with efficiency and clinical utility as upfront constraints rather than as an afterthought. The described architectures are applicable to numerous clinical problems including automatic screening, improvement of the radiologists' workflows, diagnosis and therapy of dental diseases, computer aided detection and

advanced clinical decision support. By deploying lightweight, clinically usable deep learning systems we can relieve both radiologists and the other clinical staff from tedious work, and, importantly, conduct screening tasks at large scales.

REFERENCES

- [1] K. Shung, M. Smith, and B. Tsui, Principles of Medical Imaging. 2012.
- [2] H. Kasban, M. El-Bendary, et al., “A comparative study of medical imaging techniques,” 2015.
- [3] K. Suzuki, “Overview of deep learning in medical imaging,” Radiological Physics and Technology, 2017.
- [4] J. A. B. Rubak, K. Naveed, S. Jain, L. Esterle, A. Iosifidis, and R. Pauwels, “Impact of labeling inaccuracy and image noise on tooth segmentation in panoramic radiographs using federated, centralized and local learning,” Dentomaxillofacial Radiology, p. twag001, 2026.
- [5] B. Khalil, M. Baraka, S. Haghighat, S. Jain, N. Manila, R. Ramani, A. Tichy, E. Tolstaya, F. Schwendicke, and R. Pauwels, “Synthetic imaging in dentistry: A narrative review of deep learning techniques and applications,” Journal of Dentistry, p. 106274, 2025.
- [6] Z. Chen, Z. Qu, Y. Quan, L. Liu, Y. Ding, and Y. Xie, “Dynamic n:M fine-grained structured sparse attention mechanism,” in Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023.
- [7] Y. Vu, R. Wang, N. Balachandar, et al., “MedAug: Contrastive learning leveraging patient metadata improves representations for chest X-ray interpretation,” in Proceedings of Machine Learning Research, 2021.
- [8] A. Jaiswal, T. Li, C. Zander, Y. Han, et al., “SCLAP: Supervised contrastive learning for cardiopulmonary disease classification and localization in chest X-rays using patient metadata,” in IEEE International Conference on Data Mining, 2021.
- [9] S. Rizvi, R. Tang, X. Jiang, X. Ma, et al., “Local contrastive learning for medical image recognition,” in AMIA Annual Symposium Proceedings, 2024.
- [10] Y. Han, C. Chen, A. Tewfik, Y. Ding, et al., “Pneumonia detection on chest X-ray using radiomic features and contrastive learning,” in 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021.
- [11] J. Li et al., “Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19,” Pattern Recognition, 2021.
- [12] J. Yuan, H. Gao, D. Dai, J. Luo, L. Zhao, et al., “Native sparse attention: Hardware-aligned and natively trainable sparse attention,” in Annual Conference of the Association for Computational Linguistics, 2025.

- [13] Y. Zhang et al., “A medical image segmentation method based on adaptive graph sparse algorithm under contrastive learning framework,” *Displays*, 2025.
- [14] C. Öksüz, O. Urhan, and M. Güllü, “An integrated convolutional neural network with attention guidance for improved performance of medical image classification,” *Neural Computing and Applications*, 2024.
- [15] H. Ren, H. Dai, Z. Dai, M. Yang, et al., “Combiner: Full attention transformer with sparse computation cost,” in *Advances in Neural Information Processing Systems*, 2021.
- [16] G. Shen, J. Zhao, Q. Chen, J. Leng, C. Li, et al., “SALO: An efficient spatial accelerator enabling hybrid sparse attention mechanisms for long sequences,” in *Proceedings of the 59th Annual Design Automation Conference*, 2022.
- [17] X. Li et al., “Deep learning attention mechanism in medical image analysis: Basics and beyonds,” 2023.
- [18] C. Parmar, J. Barry, A. Hosny, et al., “Data analysis strategies in medical imaging,” *Clinical Cancer Research*, 2018.
- [19] J. Li et al., “A systematic collection of medical image datasets for deep learning,” *ACM Computing Surveys*, 2023.
- [20] I. Dimitrovski, D. Koccev, S. Loskovska, and S. Džeroski, “Hierarchical annotation of medical images,” *Pattern Recognition*, 2011.
- [21] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, et al., “CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.
- [22] A. Johnson, T. Pollard, N. Greenbaum, et al., “MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs,” *arXiv preprint arXiv:1901.07042*, 2019.
- [23] E. Yanar, F. Kutan, K. Ayturan, U. Kutbay, O. Algin, et al., “A comparative analysis of the Mamba, Transformer, and CNN architectures for multi-label chest X-ray anomaly detection in the NIH ChestX-Ray14 dataset,” *Diagnostics*, 2025.
- [24] I. Baltruschat, H. Nickisch, M. Grass, T. Knopp, et al., “Comparison of deep learning approaches for multi-label chest X-ray classification,” *Scientific Reports*, 2019.
- [25] A. Dosovitskiy et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.

- [26] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.
- [27] T. Chen, S. Kornblith, M. Norouzi, et al., “A simple framework for contrastive learning of visual representations,” in International Conference on Machine Learning, 2020.
- [28] X. Chen, S. Xie, and K. He, “An empirical study of training self-supervised vision transformers,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.
- [29] M. Ryoo, A. Piergiovanni, A. Arnab, et al., “TokenLearner: Adaptive space-time tokenization for videos,” in Advances in Neural Information Processing Systems, 2021.
- [30] Y. Rao, W. Zhao, B. Liu, J. Lu, J. Zhou, et al., “DynamicViT: Efficient vision transformers with dynamic token sparsification,” in Advances in Neural Information Processing Systems, 2021.
- [31] Y. Zhang, H. Jiang, Y. Miura, et al., “Contrastive learning of medical visual representations from paired images and text,” in Proceedings of Machine Learning Research, 2022.
- [32] S. Huang, L. Shen, M. Lungren, et al., “GLORIA: A multimodal global-local representation learning framework for label-efficient medical image recognition,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.
- [33] N. A. Pritam, S. Jain, et al., “SkinGenBench: Generative model and preprocessing effects for synthetic dermoscopic augmentation in melanoma diagnosis,” arXiv preprint arXiv:2512.17585, 2025.
- [34] S. Pedersen, S. Jain, M. Chavez, V. Ladehoff, B. N. de Freitas, and R. Pauwels, “Pano-GAN: A deep generative model for panoramic dental radiographs,” Journal of Imaging, vol. 11, no. 2, p. 41, 2025.
- [35] S. Jain, B. N. de Freitas, A. Basse-OConnor, A. Iosifidis, and R. Pauwels, “PanoDiff-SR: Synthesizing dental panoramic radiographs using diffusion and super-resolution,” arXiv preprint arXiv:2507.09227, 2025.
- [36] F. Isensee, P. Jaeger, S. Kohl, J. Petersen, et al., “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,” Nature Methods, 2021.
- [37] B. Jacob, S. Kligys, B. Chen, M. Zhu, et al., “Quantization and training of neural networks for efficient integer-arithmetic-only inference,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [38] H. Awari, N. Subramani, A. Janagaraj, et al., “Three-dimensional dental image segmentation and classification using deep learning with tunicate swarm algorithm,” Expert Systems, 2024.
- [39] S. Bohlender, I. Oksuz, et al., “A survey on shape-constraint deep learning for medical image segmentation,” IEEE Reviews in Biomedical Engineering, 2021.

- [40] N. V. Nistelrooij, L. Krämer, S. Kempers, et al., “ToothSeg: Robust tooth instance segmentation and numbering in CBCT using deep learning and self-correction,” *IEEE Journal of Biomedical and Health Informatics*, 2026.
- [41] P. Wang, H. Gu, and Y. Sun, “Tooth segmentation on multimodal images using adapted Segment Anything Model,” *Scientific Reports*, 2025.
- [42] Z. Lambert, C. L. Guyader, et al., “A geometrically constrained deep network for CT image segmentation,” in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021.
- [43] Y. Singh, C. Farrelly, Q. Hathaway, T. Leiner, et al., “Topological data analysis in medical imaging: Current state of the art,” *Insights into Imaging*, 2023.
- [44] G. Zheng, X. Cui, A. Song, and M. Lin, “GFACNet: 3D dental segmentation from intraoral scans integrating geometric features and anatomical constraints,” *Electronic Research Archive*, 2025.
- [45] S. Xi, Z. Liu, J. Chang, H. Wu, et al., “3D dental model segmentation with geometrical boundary preserving,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.
- [46] A. Demir, E. Massaad, and B. Kiziltan, “Topology-aware focal loss for 3D image segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023.
- [47] J. Huang, H. Yan, J. Li, H. Stewart, et al., “Combining anatomical constraints and deep learning for 3-D CBCT dental image multi-label segmentation,” in *2021 IEEE 37th International Conference on Data Engineering*, 2021.
- [48] P. Wang, “Latent anomaly detection: Masked VQ-GAN for unsupervised segmentation in medical CBCT,” *arXiv preprint arXiv:2506.14209*, 2025.
- [49] A. Ben-Hamadou, O. Smaoui, A. Rekik, S. Pujades, et al., “3DTeethSeg’22: 3D teeth scan segmentation and labeling challenge,” *arXiv preprint arXiv:2305.18277*, 2023.
- [50] Y. He, D. Yang, H. Roth, C. Zhao, et al., “DiNTS: Differentiable neural network topology search for 3D medical image segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [51] S. Pedersen, S. Jain, M. Chavez, V. Ladehoff, B. N. de Freitas, and R. Pauwels, “Pano-GAN: A deep generative model for panoramic dental radiographs,” *Journal of Imaging*, vol. 11, no. 2, p. 41, 2025.

- [52] S. Jain, B. N. de Freitas, A. Basse-OConnor, A. Iosifidis, and R. Pauwels, “PanoDiff-SR: Synthesizing dental panoramic radiographs using diffusion and super-resolution,” arXiv preprint arXiv:2507.09227, 2025.
- [53] H. Mohammad-Rahimi, S. Jain, K. Naveed, S. Hosseinpour, L. L. Kirkevang, A. Nosrat, and R. Pauwels, “Generative artificial intelligence for computer vision in endodontics: A review of current state and future potential,” *International Endodontic Journal*, 2026.
- [54] J. A. B. Rubak, S. Haghighat, S. Jain, M. Aldesoki, A. Chaurasia, S. S. Ehsani, et al., “Deep learning-based assessment of the relation between the third molar and mandibular canal on panoramic radiographs using local, centralized, and federated learning,” arXiv preprint arXiv:2603.11850, 2026.

Appendix : LIST OF PUBLICATIONS & PROOFS

1. Paarth Prasad & Ruchika Malhotra (2026). *SparseContrast: Dynamic Sparse Attention for Efficient and Accurate Contrastive Learning in Medical Imaging* (ICAIHC 2026). [Scopus Indexed][Communicating]

2. Paarth Prasad & Ruchika Malhotra (2026). *Topology-Constrained Quantized nnUNet for Efficient and Anatomically Accurate 3D Tooth Segmentation* (ICAIHC 2026). [Scopus Indexed][Communicating]

Paper Communicating Proofs: Undertaking & Declaration



Paarth Prasad <pdkn.prasad@gmail.com>

Undertaking Regarding Author-Owned arXiv Preprints – Papers 1138 and 961

1 message

Paarth Prasad <pdkn.prasad@gmail.com>
To: "Dr. Dibakar Saha" <dsaha.mca@nitrr.ac.in>

Sun, Jun 21, 2026 at 12:06 AM

Dear TPC Chairs,

I hope this email finds you well.

I am writing to formally place on record the following information regarding our ICAIHC-2026 submissions:

Paper ID: 1138

Title: Topology-Constrained Quantized nnUNet for Efficient and Anatomically Accurate 3D Tooth Segmentation

Paper ID: 961

Title: SparseContrast: Dynamic Sparse Attention for Efficient and Accurate Contrastive Learning in Medical Imaging

We hereby declare and undertake that both manuscripts were authored entirely by the listed authors and were uploaded by the authors themselves as public preprints on arXiv prior to their submission to ICAIHC-2026.

The corresponding arXiv preprints are:

Paper 1138: <https://arxiv.org/abs/2605.04201>

Paper 961: <https://arxiv.org/abs/2605.00887>

Accordingly, any textual similarity detected between the submitted conference manuscripts and the above arXiv preprints represents overlap with our own author-owned work and not with the work of any third party. To the best of our knowledge, such overlap arises solely because the conference submissions and the arXiv versions describe the same research contributions prepared by the same authors.

We further undertake that:

1. The submitted manuscripts represent our original research work.
2. We have not copied or plagiarized material from other authors without proper attribution.
3. Any overlap identified with the above-mentioned arXiv manuscripts pertains exclusively to author-owned preprints.
4. We remain fully committed to complying with all ICAIHC-2026 policies and requirements.
5. We will make all revisions deemed necessary by the conference and will update the manuscripts accordingly in the CMT portal.

This communication is submitted for the purpose of transparency and to formally document the source of any similarity arising from the publicly available author-owned preprints.

We respectfully request that this declaration be considered during the evaluation of our submissions.

For the committee's reference, plagiarism/similarity reports excluding the corresponding author-owned arXiv preprints are attached with this email.

Thank you for your time and consideration.

Sincerely,

Paarth Prasad
(On behalf of all authors)

Paper IDs: 1138, 961

2 attachments



Paarth Prasad <pdkn.prasad@gmail.com>

Declaration Regarding Papers 1138 and 961

1 message

ICAIHC NIT Raipur <icaihc.ca@nitrr.ac.in>
To: Paarth Prasad <pdkn.prasad@gmail.com>

Mon, Jun 22, 2026 at 11:48 AM

Dear Authors,

Greetings from ICAIHC-2026.

Thank you for your detailed explanation and for providing the declaration regarding Papers ID 1138 and 961, along with the supporting documentation.

We acknowledge the information provided concerning the author-owned arXiv preprints and appreciate your transparency in clarifying the source of the detected similarity.

Your declaration and supporting documents have been noted and will be considered as part of the evaluation process. Both papers are currently under review, and the final decision will be communicated upon completion of the review and assessment process.

We appreciate your cooperation and your commitment to complying with the conference requirements.

Regards,
ICAIHC-2026 Organizing Committee
NIT Raipur

Indexing Proof

4th International Conference on Ambient Intelligence in Health Care ICAIHC-2026

18-19 September 2026

ORGANIZED BY:

Department of Computer Applications
National Institute of Technology, Raipur

[Home](#)
[About](#)
[Speaker](#)
[Committee](#)
[Submission](#)
[Registration](#)
[Previous ICAIHC Conferences](#)
[Contacts](#)

ORGANIZERS

CHIEF PATRON
N V Ramana Rao — Director, NIT Raipur, India

PATRON
Subrata Gupta — Dean (Academic), NIT Raipur
Guruprasad Subashchandra Mishra — Dean (BSC), NIT Raipur

GENERAL CHAIR
Ronaldo Fumiko Hashimoto — University of Sao Paulo, Brazil
Pabitra Mitra — IIT, Kharagpur, India
Mithilesh Atulkar — NIT Raipur, India

PROGRAM CHAIR
Tripti Swamkar — NIT Raipur, India
Priyanka Tripathi — NIT Raipur, India
Debdoot Sheel — IIT Kharagpur, India

ORGANIZERS


Tripti Swamkar
 Organizing Chair
 NIT Raipur, India


Mithilesh Atulkar
 Organizing Secretary
 NIT Raipur, India


Dibakar Saha
 Organizing Secretary
 NIT Raipur, India


Naseem Ahmad
 Organizing Secretary
 NIT Raipur, India

ABOUT

The **4th International Conference** on Ambient Intelligence in Health Care (ICAIHC-2026) aims to explore advanced applications of Artificial Intelligence, Machine Learning, IoT, and Computational Biology in healthcare systems. The conference will serve as a platform for researchers, academicians, practitioners, and industry experts to discuss innovation, challenges, and future trends in intelligent healthcare systems.

CONFERENCE TRACKS / TOPICS

Track-1: Intelligent Computation in Healthcare
Track-2: IoT & Edge Computing for Healthcare
Track-3: Computational Biology and Bioinformatics
Track-4: Robotics & Rehabilitation in Healthcare

KEYNOTE SPEAKERS


Dr. Prasen Kumar Sahoo
 Chang Jung University, Suwon, Incheon, Taiwan, Taiwan


Dr. Rohit Thakri
 Senior AI/ML scientist, Wolfsburg, Germany


Dr. Jaajit S. Suri
 CEO and Chairman, Athawani (TM), California, USA


Dr. Pralay Mitra
 IIT Kharagpur, INDIA


Dr. Deepthi R. Bathula
 IIT Raipur, India

PUBLICATION

Accepted papers will be submitted for inclusion into IEEE Xplore subject to meeting IEEE Xplore's scope and quality requirements.

Indexing: Leading indexing services including ISI Proceedings, ISI Compendex, DBP, SCOPUS, Google Scholar and IEEE link.

PAPER SUBMISSION LINK

All papers must be submitted via [Microsoft CMT link](#).


 Scan the QR code

REVIEWERS REGISTRATION

Register via the [Google Form link](#).


 Scan the QR code

IMPORTANT DATES

Paper Submission Open: **01 Jan 2026**
 Paper Submission Deadline: **30 April 2026 (Closed)**
 Notification of Acceptance: **25 May 2026 Onwards**
 Camera Ready Submission: **01 June 2026 Onwards**
 Registration Opens: **15 June 2026 Onwards**

REGISTRATION FEES

Category (Author)	IEEE Member	Non-IEEE Member
Student/Research Scholar	\$6000 / \$88.28	\$7000 / \$79.66
Academician	\$9000 / \$104.4	\$10000 / \$122.42
Industry	\$9000 / \$102.42	\$10000 / \$113.80
Listener	\$2000 / \$22.76	\$2000 / \$22.76

AUTHOR'S PROFILE

Paarth Prasad is currently pursuing a Master of Technology (M. Tech.) in Data Science from the Department of Software Engineering, Delhi Technological University (DTU), Delhi, India. He is working in the areas of Artificial Intelligence, Deep Learning, Computer Vision and Medical Image Analysis.



His thesis is titled "**NOVEL EFFICIENT DEEP LEARNING ARCHITECTURES FOR MEDICAL IMAGE ANALYSIS**" which aims at proposing a computationally efficient, yet anatomically relevant deep learning framework for medical image representation learning and 3D medical image segmentation.

The proposed work combines sparse attention, contrastive learning, topology-aware optimisation and model quantization to build efficient and lighter networks that would prove beneficial for real-time Healthcare applications.

His broad research interests include Artificial Intelligence, Machine Learning, Deep Learning, Medical Image Analysis, Computer Vision, Medical Image Segmentation, Contrastive Learning, Self-Supervised Learning, Transformer Architectures, Explainable AI, and Trustworthy AI. In his Master's program, he has explored and developed efficient deep learning architectures, sparse attention modules, topology-preserving segmentation approaches, quantized neural networks, and representation learning techniques, all with a focus on achieving excellent predictive performance for medical imaging with computational efficiency.

He has contributed to several research publications in the fields of medical image analysis, artificial intelligence, and deep learning, including: SparseContrast for efficient contrastive representation learning on medical images. A Topology-Constrained Quantized nnU-Net for accurate, robust, and efficiently computable 3D tooth segmentation. His broader interests lie in efficient AI systems, intelligent healthcare, and AI explainability for medical applications.