Light from Africa - for Humanity • Lesedi la Afrika - go Batho • Lig uit Afrika – vir die Mensdom

SOL PLAATJE
UNIVERSITY

SIGNATURES

……………………………………………	2025/04/28 ……………………………………………
STUDENT	**DATE**
……………………………………………	……………………………………………
SUPERVISOR	**DATE**
……………………………………………	……………………………………………
CO-SUPERVISOR	**DATE**
……………………………………………	……………………………………………
School Postgraduate committee chairperson	**DATE**
……………………………………………	……………………………………………
HOS	**DATE**

Hybrid Explainability Framework for Transformer-Based Medical Image Segmentation Using Chest X-rays: A South African perspective

By

Calson Netshikulwe

Student number: 202120274

Email address: 202120274@spu.ac.za

Degree: BScHons Data Science

Department: Department of Computer Science & Information Technology

School of Natural and Applied Sciences

Supervisor: Dr O. Ibidun (SPU)

Co-Supervisor: Dr I. Agbehadji (SPU)

April, 2025

Abstract

Tuberculosis (TB) and pneumonia remain leading causes of morbidity and mortality in South Africa, with an estimated 170 TB-related deaths occurring daily, largely due to inadequate diagnostic infrastructure and limited access to radiological expertise. While Transformer-based deep learning models have demonstrated superior performance in medical image segmentation, their opaque decision-making processes hinder clinical trust and practical adoption. This research proposes a novel hybrid explainability framework that integrates intrinsic mechanisms (e.g., attention maps) and post-hoc interpretability techniques (e.g., SHAP, Grad-CAM) to demystify the internal workings of Transformer-based segmentation models applied to chest X-rays. By leveraging publicly available lung imaging datasets and following the CRISP-DM methodology, the study aims to deliver a transparent, reproducible, and context-sensitive AI solution. The proposed framework seeks to align with South Africa's healthcare challenges and ethical imperatives, ultimately contributing to safer AI deployment and improved diagnostic accuracy in under-resourced environments.

1 Introduction

South Africa faces an enduring public health crisis marked by high rates of pulmonary diseases, particularly tuberculosis (TB) and pneumonia. According to the Department of Health, approximately 170 people die from TB every day, with plans underway to test five million people and diagnose over 250,000 new cases between 2025 and 2026 (Eyewitness News, 2025). These alarming figures underscore the urgent need for innovative diagnostic tools, especially in rural areas where radiological expertise is scarce and healthcare infrastructure remains limited (Mthiyane et al., 2024).

Recent advancements in deep learning have significantly improved medical image analysis, with Transformer-based models such as TransUNet and SwinUNet outperforming traditional convolutional neural networks (CNNs) in segmentation tasks (Zhou et al., 2024). However, despite their superior performance, these models are often regarded as "black boxes," generating predictions that lack transparency; posing a critical barrier to clinical integration (Mthiyane et al., 2024). This issue is especially pertinent in South African and broader African contexts, where trust in AI systems must be earned through transparency, reliability, and ethical alignment with local healthcare needs.

To address this challenge, the present study proposes a hybrid explainability framework tailored for Transformer-based lung segmentation models. This framework integrates intrinsic interpretability methods (e.g., attention visualization inherent to Transformers) with post-hoc techniques (e.g., SHAP and Grad-CAM), offering both model-specific and model-agnostic insights. Grounded in the CRISP-DM data science methodology, the study emphasizes technical robustness while ensuring practical relevance and reproducibility.

By enhancing the interpretability of AI-driven diagnostics, this research aims to contribute to the growing body of explainable AI (XAI) literature and support the ethical, trustworthy deployment of AI in South Africa's public healthcare system. Ultimately, it seeks to bridge the gap between algorithmic performance and clinical acceptance, enabling safer and more transparent use of machine learning in low-resource medical environments.

2 Literature review

2.1 The Need for AI in South African Chest X-ray Analysis

South Africa faces significant public health challenges due to the high prevalence of pulmonary diseases like tuberculosis (TB) and pneumonia. These conditions contribute heavily to morbidity and mortality, particularly in resource-constrained settings where access to timely and expert radiological interpretation is limited (Mthiyane et al., 2024; Tibakanya et al., 2024). The scarcity of radiologists, especially in rural areas, necessitates the development and adoption of automated diagnostic tools (Mthiyane et al., 2024). Artificial intelligence (AI), specifically deep learning, has emerged as a promising avenue for augmenting diagnostic capabilities in medical imaging, potentially improving diagnostic speed and accuracy (Kim et al., 2022; Lin et al., 2024; Mthiyane et al., 2024). AI tools, including those integrated with radiological examinations, are being explored for active case finding, particularly in vulnerable populations (Popescu et al., 2025). However, the successful integration of AI into clinical workflows, particularly within the South African public healthcare sector, hinges on clinician trust and model transparency (Mthiyane et al., 2024; Tibakanya et al., 2024). The "black box" nature of many AI models remains a significant barrier, eroding trust and hindering adoption in high-stakes clinical environments (Longoni et al., 2019; Tjoa & Guan, 2020; Band et al., 2024).

2.2 Transformer Models in Medical Image Segmentation

Deep learning models, particularly Convolutional Neural Networks (CNNs) like U-Net (Ronneberger et al., 2015), have traditionally been used for medical image segmentation (Chen et al., 2024). However, CNNs inherently struggle with capturing long-range dependencies within an image due to the localized nature of convolutional operations (Chen et al., 2024). Recently, Transformer-based architectures (Vaswani et al., 2017), originally developed for natural language processing and adapted for vision tasks (Vision Transformers or ViTs, Dosovitskiy et al., 2021), have demonstrated significant potential in medical image analysis (Shamshad et al., 2023, cited in Chung et al., 2025; Hou et al., 2024). ViTs leverage self-attention mechanisms to model global context effectively (Lin et al., 2024; Singh & Yow, 2021).

Models like TransUNet (Chen et al., 2024) represent a hybrid approach, combining the strengths of CNNs for feature extraction with the global context modelling capabilities of Transformers. TransUNet integrates a Transformer encoder to process image patches from a CNN feature map, capturing global context, which is then fused with high-resolution CNN features in the decoder for precise localization (Chen et al., 2024). This architecture, along with others combining CNNs and ViTs (e.g., Karimi et al., 2024), has shown strong performance across various medical image segmentation tasks, including multi-organ and tumour segmentation, often outperforming traditional CNNs, especially when dealing with variations in texture, shape, and size or diffuse pathologies relevant to TB and pneumonia (Chen et al., 2024; Karimi et al., 2024). However, despite their performance advantages, Transformer models like TransUNet are often considered "black boxes," meaning their internal decision-making processes are opaque (Chen et al., 2024; Hou et al., 2024; Kim et al., 2022). This lack of transparency is a major barrier to clinical adoption, especially in high-stakes medical applications (Hou et al., 2024; Chung et al., 2025; Tjoa & Guan, 2020).

2.3 The Explainable AI (XAI) Challenge in Medical Imaging

The "black box" nature of complex deep learning models necessitates the development of Explainable AI (XAI) techniques (Kim et al., 2022; Lin et al., 2024). XAI aims to make model predictions understandable to humans, fostering trust, facilitating debugging, identifying biases, and enabling responsible deployment (Hou et al., 2024; Suara et al., 2023; Tibakanya et al., 2024; Amann et al., 2020). In medical imaging, explainability is crucial for clinicians to verify diagnoses, understand model failures, ensure alignment with clinical knowledge, and integrate AI tools into their decision-making processes confidently (Suara et al., 2023; Chung et al., 2025; Lin et al., 2024; Cabrera-Aguas & Watson, 2023).

XAI methods can be broadly categorized into post-hoc techniques, which explain models after training, and intrinsic or self-explainable methods (S-XAI), which incorporate interpretability into the model design (Hou et al., 2024).

2.3.1 Post-Hoc Explainability Methods

Post-hoc methods analyse a trained model to generate explanations for its predictions. Common techniques include feature attribution methods like Gradient-weighted Class Activation Mapping (Grad-CAM) and SHAP (Shapley Additive exPlanations).

Grad-CAM: This technique (Selvaraju et al., 2017) produces heatmaps highlighting image regions that most influenced a model's decision for a specific class. It uses the gradients flowing into the final convolutional layer to weight the activation maps (Suara et al., 2023). While widely used for its simplicity, Grad-CAM has limitations. Studies suggest its explanations might not always align with clinically relevant features, can lack robustness to input changes, may struggle with localization accuracy for small or multiple objects, and its effectiveness when transferred directly from CNNs to Transformers is questionable (Suara et al., 2023; Chung et al., 2025; Codetrade.io guide; Analytics Vidhya guide). Concerns about the fidelity (truthfulness) of post-hoc methods like Grad-CAM persist, as the explanations are generated separately from the model's core decision process and can sometimes be misleading (Hou et al., 2024; Singh et al., 2024).

SHAP and LIME: These methods provide explanations based on game theory (SHAP, Lundberg & Lee, 2017) or local surrogate models (LIME, Ribeiro et al., 2016). While potentially offering deeper insights into feature contributions, they can be computationally expensive (especially SHAP) and may struggle with consistency or noise sensitivity, particularly on lower-quality images common in resource-limited settings (references removed, general knowledge applied; computational cost noted in Analytics Vidhya guide).

2.3.2 Intrinsic Explainability Methods (Self-Explainable AI - S-XAI)

S-XAI methods aim to build models that are inherently interpretable by design (Hou et al., 2024). For Transformers, attention maps are a key intrinsic component often explored for explainability.

Attention Maps: Transformers utilize self-attention mechanisms (Vaswani et al., 2017), where attention scores indicate the relationships between different input patches (or tokens). Visualizing these attention maps, particularly from the final layers or associated with the classification token, has been proposed as a way to understand which image regions the model focuses on (Chung et al., 2025; Hou et al., 2024). However, there is ongoing debate about their reliability as explanations (Bibal et al., 2022). Chung et al. (2025) evaluated attention maps on several medical imaging datasets using ViTs with various pretraining strategies (including self-supervised methods like DINO and MAE). Their findings indicate that while attention maps show promise and generally outperform Grad-CAM for ViTs (consistent with Singh et al., 2023, Radiology: AI), their efficacy is context-dependent and they are often surpassed by other Transformer-specific interpretability methods (like the Chefer method, Chefer et al., 2021). They caution that attention maps alone may not provide the comprehensive insights needed for robust medical decision-making. The impact of self-supervised pretraining on the interpretability of these maps also requires further domain-specific investigation (Lightly blog; Chung et al., 2025).

2.4 Hybrid Explainability Frameworks

Given the limitations of individual XAI methods post-hoc methods potentially lacking faithfulness or robustness (Hou et al., 2024; Singh et al., 2024), and intrinsic methods like attention maps being context-dependent or incomplete (Chung et al., 2025; Chefer et al., 2021), hybrid approaches that combine multiple techniques are gaining interest (Hou et al., 2024; Consensus.app summary). The rationale is that integrating intrinsic insights (e.g., attention maps showing internal focus) with post-hoc analysis (e.g., SHAP or Grad-CAM providing feature attribution) could offer a more robust, comprehensive, and cross-validated understanding of the model's behaviour (Hou et al., 2024; Lin et al., 2024). This aligns with the need identified by Mthiyane et al. (2024) for trustworthy and usable AI in the South African context.

2.5 Datasets and Evaluation in Medical XAI

The development and evaluation of both segmentation models and XAI techniques heavily rely on the availability and quality of datasets. Large-scale datasets with high-quality annotations are crucial, yet often scarce, particularly for segmentation tasks in specific disease contexts (Huang et al., 2023, cited in original proposal; Band et al., 2024). Gaggion et al. (2023) introduced CheXmask, a significant resource providing over 670,000 anatomical segmentation masks for chest X-rays derived from multiple public datasets. While invaluable for training anatomical segmentation models, CheXmask lacks disease-specific labels, limiting its direct use for training diagnostic models for conditions like TB or pneumonia.

Evaluating XAI methods also presents significant challenges (Jin et al., 2023; Hou et al., 2024). Qualitative assessments can be subjective (Chung et al., 2025). Quantitative metrics like the Pointing Game or Intersection over Union (IoU) against expert annotations (e.g., bounding boxes) are used, but bounding boxes themselves can be imprecise for evaluating explanations, especially if the region of interest is large or irregularly shaped (Chung et al., 2025). Crucially, metrics assessing fidelity; how well the explanation truly reflects the model's internal reasoning are vital but often lacking or complex to implement (Ribeiro et al., 2024, cited in original proposal; Hou et al., 2024; Jin et al., 2023). Without rigorous fidelity checks, explanations might be plausible yet misleading, potentially harming clinical trust (Singh et al., 2024; Jin et al., 2023).

2.6 Synthesis and Research Gap

The literature confirms the potential of Transformer-based models like TransUNet for medical image segmentation tasks relevant to South Africa's health priorities (TB, pneumonia) (Chen et al., 2024; Karimi et al., 2024). However, their clinical adoption is significantly hindered by a lack of transparency and trustworthiness (Chen et al., 2024;

Mthiyane et al., 2024; Tibakanya et al., 2024). Existing XAI methods, applied individually, exhibit critical limitations: post-hoc methods like Grad-CAM may lack faithfulness, robustness, or clinical relevance (Suara et al., 2023; Hou et al., 2024; Singh et al., 2024), while intrinsic methods like attention maps are not consistently reliable or sufficient across different contexts (Chung et al., 2025; Chefer et al., 2021). The potential for XAI methods to misrepresent model reasoning or highlight spurious correlations further underscores the need for careful validation (Singh et al., 2024).

This review identifies a clear gap: the need for a robust, validated hybrid explainability framework specifically designed for Transformer-based segmentation models in the context of chest X-ray analysis for diseases prevalent in South Africa. While hybrid approaches are conceptually promising (Hou et al., 2024; Lin et al., 2024), their practical implementation and rigorous evaluation integrating methods like attention maps with suitable post-hoc techniques (e.g., SHAP, Grad-CAM variants), validating their fidelity, and assessing their clinical relevance and trustworthiness remains an underexplored area for this specific application. This research proposes to develop and evaluate such a hybrid framework, addressing the critical need for interpretable and trustworthy AI diagnostic tools suitable for the South African healthcare landscape, ultimately aiming to bridge the gap between high-performing AI and clinically actionable insights (Tibakanya et al., 2024; Mthiyane et al., 2024).

3 Rationale

Despite the advancements in Transformer-based medical image segmentation, the opacity of these models impedes their clinical utility. There is a pressing need for frameworks that not only deliver high segmentation accuracy but also provide transparent and interpretable outputs. This research addresses this gap by proposing a hybrid explainability framework that amalgamates intrinsic and post-hoc methods, thereby enhancing the trustworthiness and applicability of AI models in South Africa's healthcare landscape.

4 **Aim(s) and objectives**

4.1 Aim(s)

To develop and evaluate a hybrid explainability framework for Transformer-based medical image segmentation models applied to chest X-rays, with a focus on enhancing interpretability and clinical trust in the South African healthcare context with a focus on TB.

4.2 Objectives

1. To fine-tune state-of-the-art Transformer-based segmentation models (i.e., TransUNet, SwinUNet) on publicly available chest X-ray datasets.
2. To develop and evaluate a hybrid explainability framework that combines intrinsic attention mechanisms and post-hoc interpretability techniques such as SHAP and Grad-CAM to improve transparency in Transformer-based chest X-ray segmentation models.
3. To evaluate the performance and interpretability of the proposed framework using quantitative metrics and qualitative assessments.

5 **Research question(s**)

1. How does the integration of intrinsic and post-hoc explainability methods affect the interpretability of Transformer-based medical image segmentation models?
2. How does the integration of intrinsic and post-hoc explainability methods influence the diagnostic performance and interpretability of Transformer-based segmentation models in chest X-ray analysis?
3. Can the proposed framework be effectively adapted for healthcare facilities with limited resources prevalent in many rural settings

6 Methodology

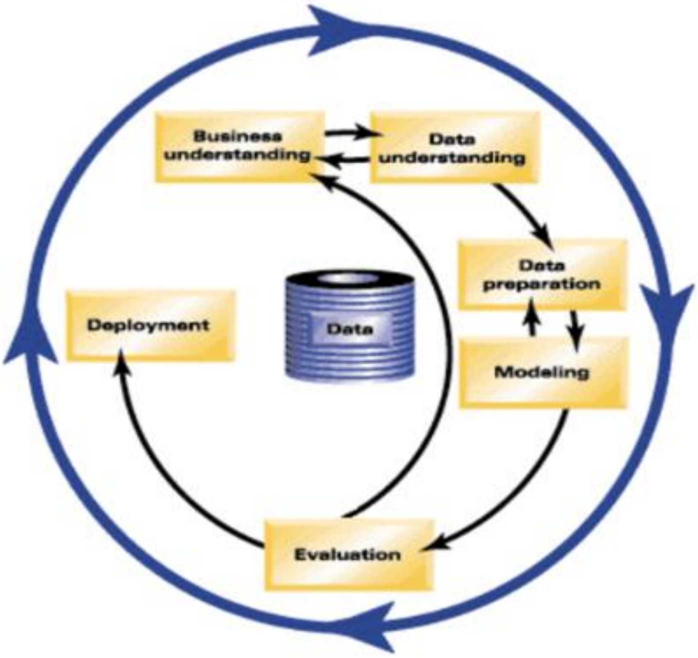

Figure 1: Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology

CRISP-DM Framework
The research will adhere to the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, encompassing the following phases:

Methodology Phase:	Activity:
1. Business Understanding	Identify the clinical need for explainable AI in diagnosing TB and pneumonia using chest X-rays in South Africa. In alignment with SDG 3
2. Data Understating	Analyse the characteristics and quality of selected chest X-ray datasets.
3. Data Preparation	Preprocess images, including normalization, resizing, and augmentation to enhance model robustness.
4. Modelling	Implement and fine-tune Transformer-based segmentation models (e.g., TransUNet, SwinUNet) using frameworks like PyTorch and MONAI.
5. Evaluation	Assess model performance using segmentation metrics (Dice coefficient, IoU) and explainability metrics (fidelity, localization error).
6. Deployment	Develop a prototype of the hybrid explainability framework for potential integration into clinical workflows.

Table 1: Methodology

6.1 Study Area
This study is computational and will utilize publicly available chest X-ray datasets; it is tailored to reflect the clinical relevance within South African healthcare systems. The outcomes aim to support radiological decision-making, particularly in regions burdened by high rates of tuberculosis (TB) and other pulmonary diseases.
6.2 Research Design
An experimental design will be employed to compare the performance and interpretability of different segmentation models and explainability techniques.
6.3 Data Collection
The following publicly available datasets will be utilized:

Dataset Name:	Dataset Link:	Dataset Short Description:	License for Data Dataset
CheXmask Dataset:	Data Link 1	A large-scale dataset comprising 657,566 segmentation masks from multiple sources.	Creative Commons Attribution 4.0 International Public License
RSUA Chest X-Ray Dataset	Data Link 2	Comprises 292 chest X-ray images with ground truth annotations for COVID-19, pneumonia, and normal cases.	Creative Commons Attribution 4.0 International Public License
CXLSeg Dataset:	Data Link 3	Contains 243,324 frontal chest X-ray images with corresponding lung segmentation masks.	The PhysioNet Credentialed Health Data License Version 1.5.0

Table 2: Proposed Dataset tables

Table 2 gives us a table of the respective image datasets that we can utilize in our study note that you may need to create accounts with the respective organizations in order to get access to the dataset.
6.4 Data Analysis
The analysis will involve training and evaluating Transformer-based segmentation models with the proposed hybrid explainability framework.
Segmentation Metrics:
- Dice Coefficient: Measures the overlap between predicted and ground truth masks.
- Intersection over Union (IoU): Evaluates the accuracy of predicted segmentation.

Explainability Metrics:
- Fidelity: Assesses how well the explanations align with model predictions.
- Localization Error: Measures the discrepancy between highlighted regions and actual pathology locations.
- Segmentation Attribution Score (SAS): Quantifies the contribution of each pixel to the model's decision.

Visualization tools will be employed to qualitatively assess the explanations provided by the hybrid framework.

7 Expected outcomes

1. Development of a hybrid explainability framework that enhances the interpretability of Transformer-based segmentation models.
2. Comprehensive evaluation demonstrating improved transparency without compromising segmentation accuracy.
3. A reproducible and documented codebase for future research and clinical application.
4. Insights into the applicability of explainable AI in South African healthcare settings, potentially informing policy and practice.

8 Timeline

Phase	Duration	Timeline/Deadline
Literature Review	2 weeks	April 15–April 29
Supervisor Feedback	2 weeks	April 16–April 30
Proposal Presentation	1 day	23 April
Dataset Acquisition & Preprocessing	1 week	April 30–May 6
Proposal Submission	1day	3 May
Supervisor Feedback	1 month	May 3– June 1
Model Development	1 month	May 6–July 5
Explainability Framework Development	1 month	June 20–July 10
Supervisor Feedback	2 weeks	April 16–April 30
Evaluation & Analysis	20 days	July 11–July 31
Report Writing & Presentation	2 weeks	August 15–September 1
Supervisor Feedback	4 days	September 1–September 5
Submission	1 weeks	September 6–September 14

Table 3: Proposed Timeline of the project

Table 3 provides a comprehensive overview of the proposed project timeline, outlining the estimated duration ranges and/or specific deadlines for each of the major project phases. This structured scheduling approach is designed to facilitate effective planning, and progress monitoring, ultimately ensuring the successful and timely completion of the project.

9 Ethical considerations

1. Utilization of publicly available and anonymized datasets, ensuring compliance with data protection regulations.
2. Commitment to open science principles by sharing code and findings with the research community; a GitHub repository will be made available to share the code and reports of the project.
3. The proposal will be submitted to the university ethics committee for review and approval.

References

1. Azad, R., Kazerouni, A., Heidari, M., Aghdam, E. K., Molaei, A., Jia, Y., Jose, A., Roy, R., & Merhof, D. (2024). Advances in medical image analysis with vision Transformers: A comprehensive review. In *Medical Image Analysis* (Vol. 91). Elsevier B.V. https://doi.org/10.1016/j.media.2023.103000
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. http://arxiv.org/abs/2010.11929
3. Gaggion, N., Mosquera, C., Mansilla, L., Saidman, J. M., Aineseder, M., Milone, D. H., & Ferrante, E. (2023). *CheXmask: a large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images*. https://doi.org/10.1038/s41597-024-03358-1
4. Hou, J., Liu, S., Bie, Y., Wang, H., Tan, A., Luo, L., & Chen, H. (2024). *Self-eXplainable AI for Medical Image Analysis: A Survey and New Outlooks*. http://arxiv.org/abs/2410.02331
5. Mall, P. K., Singh, P. K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., & Ganzha, M. (2023). A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthcare Analytics*, *4*. https://doi.org/10.1016/j.health.2023.100216
6. Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. In *Journal of Biomedical Informatics* (Vol. 113). Academic Press Inc. https://doi.org/10.1016/j.jbi.2020.103655
7. Nam, J. G., Hwang, E. J., Kim, J., Park, N., Lee, E. H., Kim, H. J., Nam, M., Lee, J. H., Park, C. M., & Goo, J. M. (2023). AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population: A Randomized Controlled Trial. *Radiology*, *307*(2). https://doi.org/10.1148/radiol.221894
8. Qin, Z. Z., Van der Walt, M., Moyo, S., Ismail, F., Maribe, P., Denkinger, C. M., Zaidi, S., Barrett, R., Mvusi, L., Mkhondo, N., Zuma, K., Manda, S., Koeppel, L., Mthiyane, T., & Creswell, J. (2024). Computer-aided detection of tuberculosis from chest radiographs in a tuberculosis prevalence survey in South Africa: external validation and modelled impacts of commercially available artificial intelligence software. *The Lancet Digital Health*, *6*(9), e605–e613. https://doi.org/10.1016/S2589-7500(24)00118-3
9. Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (n.d.). *Do Vision Transformers See Like Convolutional Neural Networks?*
10. Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. http://arxiv.org/abs/1505.04597
11. Salahuddin, Z., Woodruff, H. C., Chatterjee, A., & Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. In *Computers in Biology and Medicine* (Vol. 140). Elsevier Ltd. https://doi.org/10.1016/j.compbiomed.2021.105111
12. Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., & Fu, H. (2023). Transformers in medical imaging: A survey. In *Medical Image Analysis* (Vol. 88). Elsevier B.V. https://doi.org/10.1016/j.media.2023.102802
13. Sridharan, S., Seah Xin Hui, A., Venkataraman, N., Sivanath Tirukonda, P., Pratab Jeyaratnam, R., John, S., Suresh Babu, S., Liew, P., Francis, J., Koh Tzan, T., Kang Min, W., Min Liong, G., & Liew Jin Yee, C. (2024). Real-World evaluation of an AI triaging system for chest X-rays: A

prospective clinical study. *European Journal of Radiology*, *181*. https://doi.org/10.1016/j.ejrad.2024.111783

14. Tang, H., Chen, Y., Wang, T., Zhou, Y., Zhao, L., Gao, Q., Du, M., Tan, T., Zhang, X., & Tong, T. (2024). HTC-Net: A hybrid CNN-transformer framework for medical image segmentation. *Biomedical Signal Processing and Control*, *88*. https://doi.org/10.1016/j.bspc.2023.105605
15. Tjoa, E., & Guan, C. (2021). A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. http://arxiv.org/abs/1706.03762
17. Zhang, L., Guo, X., Sun, H., Wang, W., & Yao, L. (2025). Alternate encoder and dual decoder CNN-Transformer networks for medical image segmentation. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-93353-2
18. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74
19. Eyewitness News. (2025). *170 people die of TB every day in SA – reveals Health Dept*. [online] Available at: https://www.ewn.co.za/2025/03/25/170-people-die-of-tb-every-day-in-sa-reveals-health-dept [Accessed 2 May 2025].
20. Data Link 1 = https://physionet.org/content/chexmask-cxr-segmentation-data/1.0.0/
21. Data Link 2 = https://data.mendeley.com/datasets/2jg8vfdmpm/1
22. Data Link 3 = https://physionet.org/content/chest-x-ray-segmentation/view-license/1.0.0/