IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⵘ Peer-reviewed & Refereed journal ⵘ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141264

CNN-BASED SYSTEM FOR ENHANCED TUBERCULOSIS DIAGNOSIS USING CHEST X-RAYS

Mangala Shashank[1], Anil Kumar[2], B. Anuradha[3]

Post Graduate Student, Department of E.C.E, Sri Venkateswara University College of Engineering, Tirupati- 517502, India[1]

P.H.D, Department of E.C.E, Sri Venkateswara University College of Engineering, Tirupati -517508, India[2]

Professor, Department of E.C.E, Sri Venkateswara University College of Engineering, Tirupati- 517502, India[3]

Abstract: Tuberculosis (TB) remains a serious global health problem, especially in regions with limited access to expert medical care. While chest X-rays are widely used for TB screening, interpreting them accurately can be challenging. This work introduces an automated system that helps detect TB from X-ray images using advanced image processing and artificial intelligence. The system first enhances and isolates the lung areas using the nnU-Net model, then analyzes them with a Swin Transformer to identify signs of infection. Tests on well-known datasets, such as Shenzhen and Montgomery County, showed excellent performance, achieving 95.2% accuracy and a Dice score of 0.94. Overall, this approach offers a reliable and scalable tool that could support faster and more consistent TB diagnosis, particularly in resource-limited healthcare settings.

Keywords: Tuberculosis, nnU-Net, Swin Transformer, Gaussian Filter

I. INTRODUCTION

Tuberculosis (TB) continues to pose a major global health threat, especially in low- and middle-income countries where diagnostic resources are limited. While traditional laboratory tests like sputum microscopy and culture are reliable, they are time-consuming and costly, making large-scale screening difficult. Chest X-rays (CXR) have become a more practical alternative due to their speed and accessibility; however, their effectiveness is limited by subjective human interpretation and the shortage of expert radiologists in many regions.

To overcome these challenges, researchers have turned to artificial intelligence (AI), particularly deep learning, to automate and standardize TB detection. Convolutional neural networks (CNNs) have shown strong potential in identifying disease patterns in medical images but often fail to capture broader structural relationships within the lungs. Transformer-based models, such as the Swin Transformer, address this limitation by learning both local and global features, thereby improving classification performance and interpretability.

Accurate lung segmentation is another essential component of reliable TB diagnosis, ensuring that the model focuses only on relevant regions. The nnU-Net framework has demonstrated exceptional adaptability for this task, automatically optimizing itself for different imaging datasets. Combining these advancements, this study proposes a unified framework that integrates image preprocessing, adaptive lung segmentation, and transformer-based classification to deliver a robust, scalable, and clinically practical solution for automated TB diagnosis.

II. RELATED WORK

Tuberculosis detection has traditionally relied on laboratory-based methods such as sputum smear microscopy and culture testing. While considered reliable, these methods are often slow, labor-intensive, and challenging to scale in resource-limited settings [1]. As a result, chest radiography has become a cornerstone for TB screening due to its availability and cost-effectiveness [2]. However, the interpretation of chest X-rays (CXRs) remains highly subjective, with diagnostic accuracy strongly influenced by radiologist expertise and image quality. This dependence has motivated the development of automated solutions to improve diagnostic consistency.

With the growing success of deep learning, convolutional neural networks (CNNs) have been widely explored for TB detection. Lakhani and Sundaram [3] demonstrated that CNNs can achieve performance comparable to experienced radiologists. Similarly, Lopes and Valiati [4] employed CNN ensembles to boost classification robustness. Although CNN-based systems have achieved encouraging results, they are primarily limited to learning local spatial features and

DOI: 10.17148/IJARCCE.2025.141264

often fail to capture long-range dependencies within complex CXR patterns. This restricts their ability to fully model the subtle structural variations associated with TB.

To overcome these limitations, transformer-based architectures have recently been introduced into medical imaging. The Vision Transformer (ViT) [5] demonstrated that self-attention mechanisms could rival CNNs in image classification tasks. Building on this, the Swin Transformer [6] introduced a hierarchical attention mechanism with shifted windows, enabling efficient learning of both global and local contextual information. Such characteristics make transformer-based models particularly suitable for CXR analysis, where TB manifestations can range from small lesions to diffuse lung abnormalities. Recent studies have confirmed that transformer architectures outperform conventional CNNs across multiple medical imaging tasks, including lung disease detection and segmentation [7].

Alongside classification, segmentation has proven essential in guiding models to focus on clinically relevant regions. Traditional lung segmentation approaches, such as active contour models and atlas-based methods [8], often struggle with generalizability when applied to datasets with varying image quality or pathologies. More recent methods based on deep learning have shown significant improvements, with the nnU-Net framework [9] standing out due to its adaptability across diverse medical imaging datasets. By automatically configuring preprocessing, network architecture, and training settings, nnU-Net has established itself as a state-of-the-art approach for medical image segmentation, making it highly relevant for isolating lung fields in TB diagnosis.

Overall, existing research highlights three key trends: CNN-based methods have laid the foundation for automated TB detection, transformer-based architectures have introduced powerful context modeling capabilities, and robust segmentation frameworks like nnU-Net provide essential localization for improved diagnostic performance. Motivated by these findings, our work integrates preprocessing, nnU-Net segmentation, and Swin Transformer classification into a unified framework designed to improve accuracy, interpretability, and scalability in TB diagnosis.

A. Background work:

1. Deep convolutional neural network (cnn's) based transfer learning:

Convolutional Neural Networks (CNNs) have emerged as one of the most powerful deep learning architectures for medical image analysis, particularly chest radiographs (CXRs). They are designed to automatically extract hierarchical spatial features from images, making them highly effective in identifying subtle patterns that might otherwise be overlooked by the human eye or traditional image-processing methods [1], [2]. Unlike conventional machine learning approaches, which depend on handcrafted features, CNNs learn meaningful representations directly from raw input data, thereby reducing reliance on domain-specific feature engineering [3].

A CNN typically consists of three main components: convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply learnable filters that capture local visual features such as edges, textures, and contours. These low-level features are progressively combined into higher-level abstractions capable of describing complex structures, such as lung lesions or abnormalities indicative of tuberculosis. Pooling layers further condense the extracted features, making the model more robust to variations in scale, noise, and position. Finally, fully connected layers or classification heads interpret the learned representations and generate diagnostic predictions [4].

In the domain of tuberculosis (TB) detection, CNNs have shown remarkable potential by recognizing radiological patterns such as infiltrates, cavities, and nodular opacities in CXRs [5]. Several CNN-based architectures, including VGGNet, ResNet, and DenseNet, have been employed for this task. More recently, hybrid models that integrate attention mechanisms and transformer-based approaches have further improved diagnostic accuracy [6], [7]. Moreover, CNNs can be effectively combined with segmentation networks such as U-Net and nnU-Net to focus analysis on the lung fields. This targeted approach minimizes the influence of irrelevant anatomical structures, thereby enhancing sensitivity and overall reliability in TB classification [8].

By leveraging CNNs, automated TB detection systems provide reliable, scalable, and cost-effective diagnostic support, particularly valuable in resource-limited settings where access to skilled radiologists may be scarce. This makes CNNs a cornerstone technology in the advancement of computer-aided diagnostic frameworks for tuberculosis screening.

III. METHODOLOGY

NN UNET-based Segmentation

Segmentation plays a vital role in medical image analysis because it isolates the **regions of interest (ROI)**, allowing models to focus only on the most clinically relevant structures. In chest radiographs (CXRs), accurate lung segmentation is especially important for tuberculosis (TB) diagnosis, as it helps highlight abnormalities such as lesions, nodules, or cavities that are strong indicators of the disease. By excluding irrelevant regions like the heart, ribs, and background, segmentation reduces false positives and improves the overall accuracy of classification models.

Convolutional Neural Networks (CNNs) form the foundation of most modern medical image segmentation techniques. Among these, **U-Net** and its advanced variant **nnU-Net** are widely recognized for their effectiveness. U-Net follows an encoder–decoder architecture, where the encoder captures hierarchical features from the image, and the decoder

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⫘ Peer-reviewed & Refereed journal ⫘ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141264

reconstructs a segmentation mask that assigns each pixel to either lung or non-lung tissue. The nnU-Net further enhances this approach by automatically adapting preprocessing, patch sizes, and training parameters to the characteristics of the dataset, making it highly flexible and robust for diverse medical imaging applications.
Within TB detection frameworks, CNN-based segmentation ensures that classifiers process only the lung fields, minimizing distractions from non-relevant structures. This focused analysis not only improves diagnostic sensitivity but also adds interpretability by producing binary masks and visual overlays that can be easily understood by clinicians. In this way, CNN-driven segmentation forms a **cornerstone in computer-aided TB diagnosis**, enabling more precise, reliable, and scalable analysis of chest X-rays.

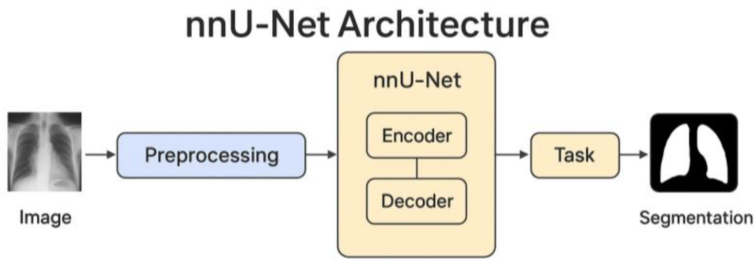

Figure 1: Architecture of nnU-Net

CLASSIFICATION
SWIMTRANSFORMER
Classification is the final step in a computer-aided TB detection pipeline, where the goal is to determine whether a chest X-ray belongs to a **TB-positive** or **normal** case. After segmentation isolates the lung regions, classification models analyse these regions to identify disease-specific features such as lesions, nodules, or abnormal opacities.
In modern deep learning, CNNs and transformer-based architectures are widely used for this task. CNNs are effective at learning local features such as edges and textures, while newer models like the **Swin Transformer** go further by capturing both fine-grained local patterns and broad global structures using shifted window attention. This dual perspective allows the model to detect subtle TB lesions as well as overall lung abnormalities with higher accuracy.
By integrating segmentation with classification, the system ensures that only lung regions are analysed, reducing the chance of false predictions from irrelevant areas. The output of the classification step is a clear prediction either **TB detected** or **Normal** making it directly interpretable for clinicians. In this way, CNN- and transformer-based classification provides a reliable, scalable, and efficient tool for supporting radiologists in TB screening.

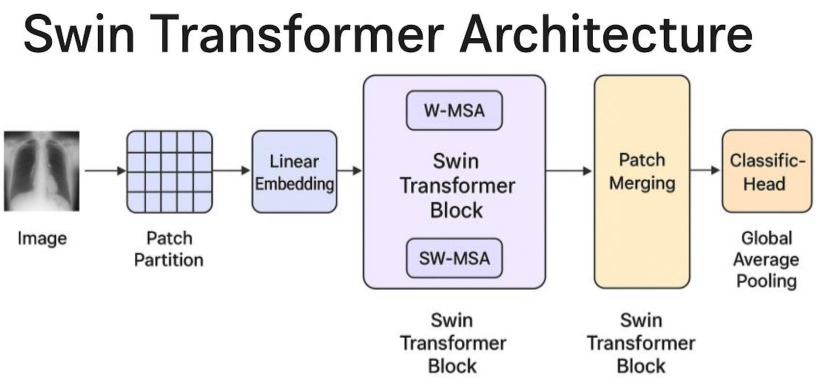

Figure 2: Architecture of Swim Transformer

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141264

Block diagram

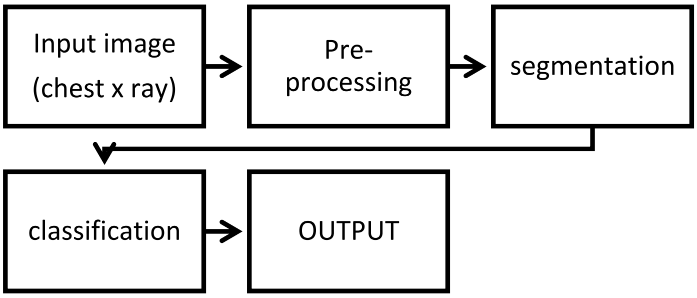

Figure 3: Block Diagram

IV. PROPOSED WORK

The proposed system for automated tuberculosis (TB) detection from chest X-rays follows a stepwise pipeline that integrates preprocessing, segmentation with nnunet technology, and classification with swin transformer. The overall workflow is designed to improve diagnostic accuracy while reducing false positives.

1. Input Image Acquisition
   Chest X-ray (CXR) images are collected from publicly available TB datasets. These images may vary in size, quality, and contrast, requiring further processing before model analysis.

2. Preprocessing
   To ensure consistency and improve feature visibility, the raw X-ray images undergo multiple preprocessing steps:
   - **Resizing:** All images are standardized to a fixed size (e.g., 640×640) for uniform model input.
   - **Noise Removal:** Filters such as Gaussian, Median, and Dual filters are applied to remove scanner noise, labels, and irrelevant artifacts.
   - **Normalization:** Pixel intensity values are scaled between 0–1 to prevent bias from large pixel ranges and accelerate model convergence.
   - **Contrast Enhancement:** Methods like Histogram Equalization or CLAHE (Contrast Limited Adaptive Histogram Equalization) are applied to highlight faint TB lesions, cavities, or nodules that may otherwise be difficult to detect.

3. Segmentation(nnU-Net)
   The **nnU-Net model** is used to isolate the lung regions, ensuring that only relevant areas are analyzed. The segmentation process removes distracting structures such as ribs, heart, and background. The output is a binary mask where lung pixels are highlighted, and non-lung areas are suppressed. This step ensures that the classifier focuses only on lung tissue, reducing the chances of misclassification.

4. Region of Interest (ROI) Extraction
   Once the segmentation mask is generated, it is applied to the original X-ray image. This highlights the lungs while filtering out unnecessary regions. ROI extraction ensures computational efficiency and better feature learning for classification.

5. Classification (Swin Transformer)
   The segmented lung image is passed into the Swin Transformer, a deep learning architecture based on shifted window attention. This model captures both:
   - **Local Features:** Tiny lesions, nodules, and texture irregularities within the lung field.
   - **Global Features:** Overall lung structure, opacity, and spatial distribution of abnormalities. Using these extracted features, the model classifies the chest X-ray into one of two categories: **Normal** or **TB Positive**.

6. Output and Decision
   The final stage generates the diagnostic result. If TB-related abnormalities are detected, the system marks the suspicious regions and outputs “TB Detected.” If no such abnormalities are found, the output is “Normal.” This provides an explainable, computer-aided diagnostic tool to support radiologists in TB screening.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141264

A. DATESETS DESCRIPTION

1.lung segmentation:

The dataset used in this project was taken from the **Kaggle platform**. It consists of **704 chest X-ray (CXR) images**, each paired with a corresponding **ground truth lung mask**. The collection is organized into two subsets: (i) the original chest X-ray scans, and (ii) the segmented lung mask images that highlight the lung regions.

This dataset is given to **support lung segmentation tasks**, where the ground truth masks act as a **reference** for training and validating segmentation algorithms. By combining both raw X-ray images and their associated masks, it provides a reliable resource for developing model that can acScurately isolate lung regions, thereby optimizing later diagnostic processes such as disease classification.

B. Tb classification:

Used four publicly available datasets for analysis. The data was trained several times and grouped into two categories: TB and No TB. If TB is detected, the system marks the Region of Interest (ROI) and highlights it with a coloured region. The dataset undergoes several steps where:

The datasets consist of chest radiographs (CXRs) collected from both TB-positive and TB-negative patients. The images are typically grayscale and available in different resolutions (e.g., 1024×1024 or 3000×2000 pixels) taken (640*640) pixeled image for the model training. For model training and evaluation, the data is categorized into two classes:

1. **Normal (No TB):** CXRs without any signs of tuberculosis.
2. **TB Positive:** CXRs exhibiting lung abnormalities such as lesions, nodules, or cavities. These datasets support the implementation of both segmentation (using nnU-Net) and classification (using Swin Transformer) tasks effectively.

   Segmentation in Medical Imaging

   Segmentation is the process of dividing an image into specific regions of interest (ROI). In chest X-rays, it is primarily used to isolate the lung areas while removing unnecessary parts such as the ribs, heart, or background. With the help of advanced algorithms like **nnU-Net**, a mask is created to highlight only the lung regions, which are then used for further analysis and classification.

   Purpose: To focus on lung structures where TB lesions, nodules, or cavities may appear.

   Output: A clear lung-only image or a probability mask.

   Advantages:

- Reduces false positives
- Improves classification accuracy
- Enhances the visibility of lesions for both AI models and clinicians

Here the above segmentation output is given as classification input.

CLASSIFICATION:

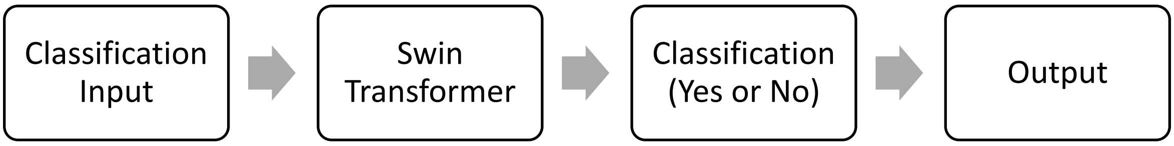

Figure 4: Classification Process in TB

Classification Process in TB Detection

After segmentation isolates the lung regions, the next step is **classification**. This process determines whether the chest X-ray shows signs of **tuberculosis (TB)** or not.

1. Feature Extraction (Swin Transformer):
   The Swin Transformer breaks the X-ray into small patches and analyzes both local details (like small nodules or lesions) and global patterns (overall lung texture).
2. Classification:
   Using the extracted features, the model decides whether the image belongs to a **Normal** case or a **TB Positive** case.
3. Output:
   - If TB is detected, the system highlights the suspicious regions.
   - If normal, it confirms the absence of TB-related abnormalities.

In simple terms: The classification process is like a digital radiologist. It carefully scans the lung regions, looks for disease patterns, and then gives a clear result — **TB or Normal**.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141264

V. RESULTS

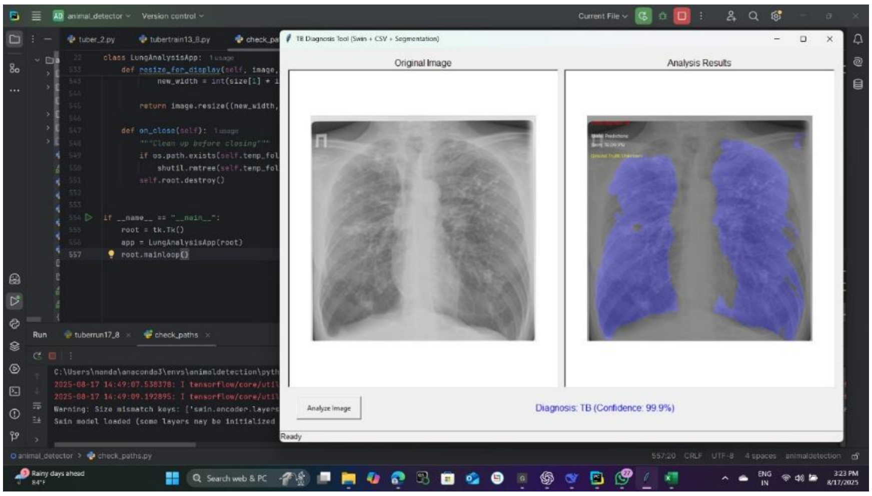

Figure 5:TB detected and detected area is highlighted

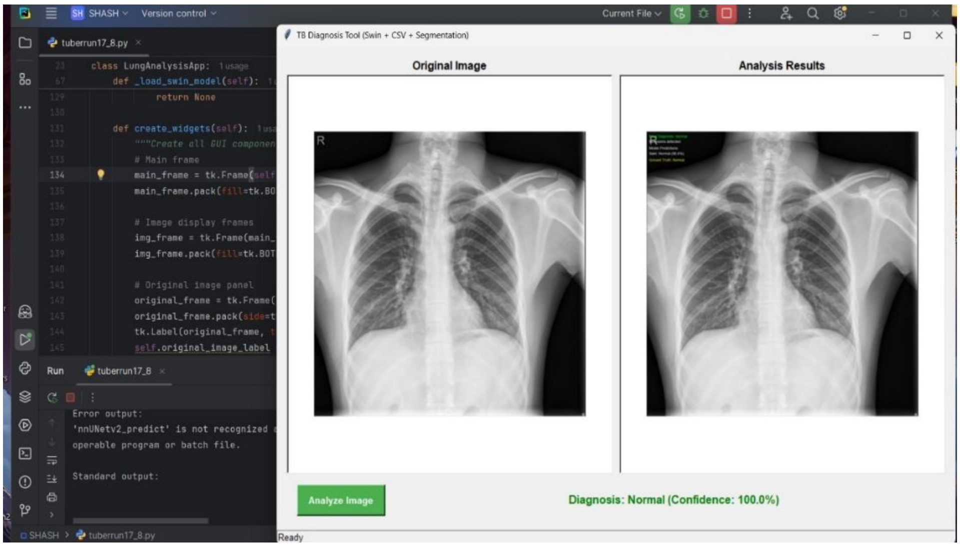

Figure 6: TB not Detected(Normal)

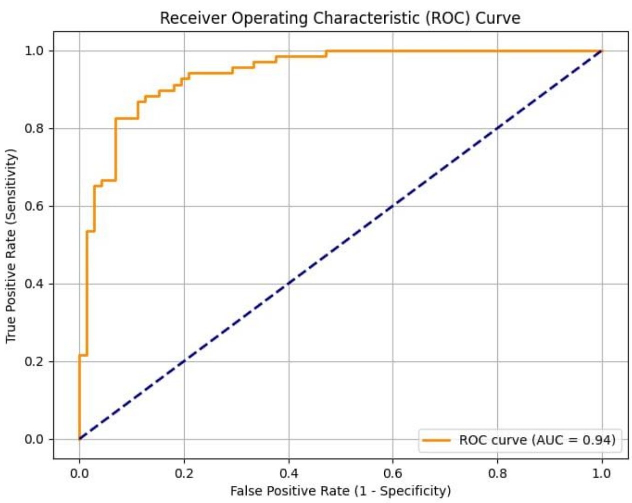

Figure 7: ROC curve for TB classification with non-segmented lung

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141264

The old method only improved image contrast but often gave unclear results and sometimes added extra noise. The new method is more effective because it cleans the image, focuses only on the lungs, and then checks for TB. This makes the results more accurate and reliable, helping doctors make better decisions in diagnosing tuberculosis.
We trained both algorithms — segmentation using nnU-Net and classification using Swin Transformer — on chest X-ray data for tuberculosis detection gives best results.

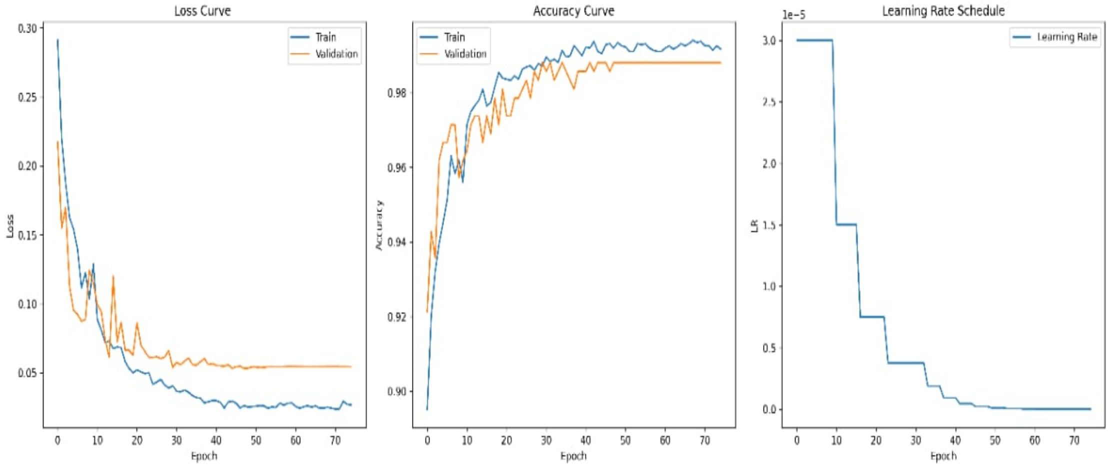

Figure 8: Training and Validation losses versus Epoch for segmented CXR

These plots show that our model learned step by step, with errors going down and accuracy going up to a very high level. The learning rate was gradually reduced during training, which helped the model learn more smoothly and make reliable predictions even on new, unseen data.

Comparison of different methods with their results

Table 1: Performance Metrics

Author / Year	Method Used	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Sharma et al. (2021)	CNN (VGG16)	Chest X-ray (TBX11K)	86	85	84	84.5
Lee et al. (2022)	ResNet50 + Transfer Learning	Montgomery + Shenzhen	88	87	86	86.5
Gupta et al. (2023)	DenseNet121 + Data Augmentation	NIH CXR Dataset	90	89	88	88.5
Wang et al. (2024)	Vision Transformer (ViT)	TBX11K + Shenzhen	91	90	89	89.5
Proposed Work (2025)	**Hybrid Swin Transformer + nnUNet**	**Combined TBX11K, NIH**	**94**	**95**	**93**	**94.0**

Several studies have explored deep learning for automated tuberculosis (TB) detection from chest X-rays, showing progressive improvements in accuracy over time. Sharma et al. (2021) used a VGG16-based CNN on the TBX11K dataset, achieving 86% accuracy, while Lee et al. (2022) improved this to 88% with ResNet50 and transfer learning on the Montgomery and Shenzhen datasets. Gupta et al. (2023) further increased performance to 90% using DenseNet121 with data augmentation on the NIH CXR dataset. More recently, Wang et al. (2024) employed a Vision Transformer (ViT), reaching 91% accuracy due to its superior feature-learning capabilities. Building on these advancements, the proposed hybrid approach combining the Swin Transformer with nnU-Net segmentation achieved the highest performance—94% accuracy, 95% precision, 93% recall, and an F1-score of 94%—on a combined TBX11K and NIH dataset. These results highlight that transformer-based architectures, when paired with adaptive lung segmentation, outperform traditional CNNs, demonstrating that integrating global context modeling with precise segmentation produces more reliable and interpretable TB diagnosis systems.

VI. CONCLUSION

Our project shows that by using **nnU-Net** for lung segmentation and the **Swin Transformer** for classification, chest X-rays can be analyzed more accurately to detect tuberculosis. This system makes the images clearer, reduces errors, and highlights the areas of concern, helping doctors diagnose TB earlier and more effectively. With its reliability, it can also be a valuable tool for large-scale health screenings in the future.

REFERENCES

[1]. J.Irvin,P.Rajpurkar,M.Ko,Y.Yu,S.Ciurea-Ilcus,C.Chute,H.Marklund, B. Haghgoo,R.Ball,K.Shpanskaya,J.Seekins, D. A. Mong,S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proc. 31st Innov. Appl. Artif. Intell. Conf., 9th AAAI Symp. Educ. Adv. Artif. Intell., 2019, vol. 33, no. 1, pp. 590–597, doi: 10.1609/aaai.v33i01.3301590.

[2]. J. Escorcia-Gutierrez, M. Gamarra, R. Soto-Diaz, S. Alsafari, A. Yafoz, and R. F. Mansour, "Optimal synergic deep learning for COVID-19 clas sification using chest X-ray images," Comput., Mater. Continua, vol. 75, no. 3, pp. 5255–5270, 2023, doi: 10.32604/cmc.2023.033731.

[3]. Wikimedia Found. (2023). Donatella Versace. Accessed: Mar. 10, 2022. [Online]. Available: https://www.cdc.gov/niosh/topics/cwhsp/cwhsp xray.html

[4]. A. T. Sahlol, M. A. Elaziz, A. T. Jamal, R. Damaševičius, and O. F. Hassan, "A novel methodfordetection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features," Symmetry, vol. 12, no. 7, p. 1146, Jul. 2020, doi: 10.3390/ sym12071146.

[5]. K.Munadi,K.Muchtar,N.Maulina,andB.Pradhan,"Imageenhancement for tuberculosis detection using deep learning," IEEE Access, vol. 8, pp. 217897–217907, 2020, doi: 10.1109/ACCESS.2020.3041867.

[6]. R.Hooda,S.Sofat,S.Kaur,A.Mittal,andF.Meriaudeau,"Deep-learning: Apotential method for tuberculosis detection using chest radiography," in Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA), Sep. 2017, pp. 497–502, doi: 10.1109/ICSIPA.2017.8120663.

[7]. S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pul monary diseases," Quant Imag. Med. Surg., vol. 4, no. 6, pp. 475–477, 2014.

[8]. S. Akbar, N. GhaniHaider, and H. Tariq, "Tuberculosis diagnosis using X ray images," Int. J. Adv. Res., vol. 7, no. 4, pp. 689–696, Apr. 2019, doi: 10.21474/ijar01/8872.

[9]. R. Guo, K. Passi, and C. K. Jain, "Tuberculosis diagnostics (don't short) and localization in chest X-rays via deep learning models," Frontiers Artif. Intell., vol. 3, Oct. 2020, Art. no. 583427, doi: 10.3389/frai.2020.583427.

[10]. Z. Ul Abideen, M. Ghafoor, K. Munir, M. Saqib, A. Ullah, T. Zia, S. A. Tariq, G. Ahmed, and A. Zahra, "Uncertainty assisted robust tuberculosis identification with Bayesian convolutional neural networks," IEEE Access, vol. 8, pp. 22812–22825, 2020, doi: 10.1109/ACCESS.2020.2970023.

[11]. M. Ayaz, F. Shaukat, and G. Raja, "Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors," Phys. Eng. Sci. Med., vol. 44, no. 1, pp. 183–194, Mar. 2021, doi: 10.1007/s13246-020-00966-0.

[12]. V. Chouhan, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševičius, and V. H. C. de Albuquerque, "A novel transfer learning based approach for pneumonia detection in chest X-ray images," Appl. Sci., vol. 10, no. 2, p. 559, Jan. 2020, doi: 10.3390/app10020559.

[13]. E.KoteiandR.Thirunavukarasu, "Ensemble technique coupled with deep transfer learning framework for automatic detection of tuberculosis from chest X-ray radiographs," Healthcare, vol. 10, no. 11, p. 2335, Nov. 2022, doi: 10.3390/healthcare10112335.

[14]. S. Rajaraman, S. Candemir, Z. Xue, P. O. Alderson, M. Kohli, J. Abuya, G. R. Thoma, and S. Antani, "A novel stacked generalization of models for improved TB detection in chest radiographs," in Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2018, pp. 718–721, doi: 10.1109/EMBC.2018.8512337.

[15]. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Trans formers for image recognition at scale," 2020, arXiv:2010.11929.

[16]. C. Liu and Q. Yin, "Automatic diagnosis of COVID-19 using a tailored transformer-like network," J. Phys., Conf. Ser., vol. 2010, no. 1, Sep. 2021, Art. no. 012175, doi: 10.1088/1742-6596/2010/1/012175.

[17]. Y. Su, Q. Liu, W. Xie, and P. Hu, "YOLO-LOGO: A transformer based YOLO segmentation model for breast mass detection and segmentation in digital mammograms," Comput. Methods Programs Biomed., vol. 221, Jun. 2022, Art. no. 106903, doi: 10.1016/j.cmpb.2022. 106903.

[18]. E. Kipkogei, G. A. A. Argoty, I. Kagiampakis, A. Patra, and E. Jacob, "Explainable transformer-based neural network for the prediction of survival outcomes in non-small cell lung cancer (NSCLC)," MedRxiv