



Tuberculosis detection System using Deep Learning

By

**Mustansar Hussain
<B00493244>**

Thesis Project for the partial fulfilment
of the requirements for the Master Degree
in Information Technology

University of the West of Scotland
School of Computing

Date of submission: 21/04/2023
Supervisor: Zeeshan Pervez
Moderator: Naeem Ramzan



DECLARATION OF ORIGINALITY

I declare that this is an original study based on my own work and that I have not submitted it for any other course or degree.

Signature: Mustansar Hussain



Library Form to Accompany MSc Project

To be completed in full

Surname: Hussain	
First Name: Mustansar	Initials: M. H.
Banner No: B00493244	
Course Code:	COMP11024
Course Title: Msc IT	
Project Supervisor: Zeeshan pervez	
Project Title: Tuberculosis detection System using Deep Learning	
Session: 2022-2023	Date of Submission: 21/04/2023
Signature: Mustansar	

Please ensure that a copy of this form is included with your project before submission.

Dedication

In the name of Allah, the Most Merciful, the Most Compassionate.

I dedicate this research project to my beloved parents, who have always been my guiding light and a source of inspiration in my life. Your unwavering love, support, and encouragement have been instrumental in helping me achieve my goals. May Allah bless you both with health, happiness, and long life.

I would also like to express my heartfelt gratitude to my supervisor, for his invaluable guidance, patience, and mentorship throughout this research project. I am truly grateful for your support and for sharing your expertise with me.

Finally, I thank Allah for giving me the strength, knowledge, and opportunity to undertake this research. I pray that this work will be of benefit to humanity and will earn His pleasure. All praise and gratitude are due to Allah alone.

ABSTRACT

Detecting tuberculosis (TB) early is crucial for successful treatment and control of this global health crisis. Medical image analysis, especially TB detection and classification, is an area where deep learning models have demonstrated encouraging results in recent years. We present a TB detection and classification system in this work that makes use of the YOLOv5 segmentation algorithm and the ViT transfer learning model. We utilized four open datasets, including the 156 photos from the DA dataset, 150 from the DB dataset, 138 from the Montgomery dataset, and 662 from the Shenzhen dataset. Each dataset was split in half and utilized in the training and validation processes.

Our research indicated that the suggested approach was 91.5% accurate, with an average sensitivity of 83.0% and a specificity of 97.0% when testing for tuberculosis. The TB class achieved a 90.0% accuracy score and a 91% recall rate. In addition, the suggested technique successfully classified the X-ray pictures into three groups (TB, healthy, and ill) with a 91.4% accuracy and an F1 score of 0.93. This research demonstrates the efficacy and efficiency of deep learning models for TB detection and categorization.

Finally, the suggested TB detection and classification system may aid in the early identification and diagnosis of TB, which has the potential to greatly improve clinical outcomes and lessen the strain TB places on healthcare systems. The suggested methodology may also be applied to other areas of medical picture analysis, expanding the scope of AI's potential in the healthcare sector.

Table of contents

1	ABSTRACT.....	5
1.1	Table of contents	6
1.2	List of figures	9
1	Chapter	10
1.1	Introduction and background	10
1.2	Introduction.....	10
1.3	Background	12
1.4	Problem Statement	15
1.5	Objectives.....	15
1.6	Proposed Study.....	16
2	Chapter.....	17
2.1	Literature Review	17
2.2	Studies	17
2.2.1	Using data augmentation techniques to improve tuberculosis screening in chest X-ray images.....	17
2.2.2	Hybrid deep learning approach for tuberculosis detection in chest X-ray images	21
2.2.3	Reviewing the research on hybrids deep learning algorithms for TB diagnosis in chest X-ray images	22
2.2.4	An ensemble of deep learning models improves tuberculosis diagnosis accuracy in chest X-rays	23
3	Chapter.....	27
3.1	Research & design.....	27
3.2	Research Design	27
3.3	Data Collection.....	27
3.4	Data Pre-processing:	29
3.5	Model architecture:	31
3.5.1	Vit transformer:.....	32
3.5.2	The distinction between CNN and ViT (CNN versus ViT).....	34
3.6	YOLOV5:.....	37
3.7	Training and Evaluation:	39
3.8	Evaluation metrics:.....	42
3.8.1	Classification:	42

3.8.2	Accuracy:	44
3.8.3	Precision:.....	45
3.8.4	Recall:	45
3.8.5	F1-score:	45
3.8.6	Object detection:	45
3.8.7	mAP:	45
3.8.8	IoU:	45
3.9	Ethical Considerations.....	45
4	Chapter	47
4.1	Results and findings	47
4.2	RESULTS AND ANALYSIS:	47
4.3	Classification Results:	48
4.4	Training and Validation Loss Graph:	49
4.5	Accuracy Graph:	51
4.6	Confusion Matrix and Classification Report:.....	52
4.7	Detection Results:	54
4.8	Precision and Recall Graphs:	55
5	Chapter	58
5.1	Discussion and conclusion	58
6	References.....	61
7	Appendices.....	64
7.1	Appendix A: For the month of January	64
7.2	Appendix B	65
7.3	Appendix C	66
7.4	Appendix D	67

List of figures

Figure 1 Data Pre-Processing	28
Figure 2 encoder part of Vision transformer (Source Usman, Zia and Tariq, 2022)	33
Figure 3 Vit Model for medical image classification (source Usman, Zia and Tariq,2022)	34
Figure 4 https://www.researchgate.net/figure/The-network-architecture-of-YOLOv5-1-Backbone-CSPDarknet-for-feature-extraction-2_fig1_358553872	36
Figure 11 Detection & Classification	56

List of Graphs

Graph 1 Loss view	48
Graph 2 Accuracy	51
Graph 3 Confusion Matrix 0-TB,1-Healthy,2-Sick	52
Graph 4 Precision Confidence Curve	55
Graph 5 Recall Confidence Curve	56
Graph 6 Precision Recall Curve	56

1 Chapter

1.1 Introduction and background

1.2 Introduction

Mycobacterium tuberculosis is responsible for tuberculosis (TB). According to Merker et al (2022) leaving tuberculosis untreated can be fatal, and it is spread by coughing or sneezing. Most developing countries lack resources for diagnosing and treating TB, making it a significant public health problem. Treatment and diagnosis of TB at an early stage is an effective way to combat the disease. Sputum smear microscopy and culture-based methods, however, are limited in sensitivity and specificity by traditional methods. Thus, TB diagnostic tools should be more precise and efficient. A recent study showed that deep learning techniques can be used to analyze medical images. Many medical imaging tasks, including TB detection, have been successfully addressed by CNNs.

Using “convolutional neural networks” with profound connections, we present a method for transferring studies(Hung and Jason, 2019). The dataset originates from Kaggle and therefore is publicly accessible. The discipline of computer programming plays a crucial role in the dynamic medical systems of today. As a consequence of the collaborative nature of the suggestive process, processor skill offers vital functions not only to health experts in the field and administrations, but also to patients' relatives, academics, and decision-makers. Consequently, any developments that enhance medical testing while maintaining their supremacy and safety are crucial for the growth of the medical sector. In the early phases of the majority of illnesses, a diagnosis can typically be made with reasonable certainty. In this study, all established tb (TB) diagnosis techniques were implemented.

The proposed dissertation aims to design and develop a deep learning-based TB detection system using chest radiographs. In order to detect TB, the system will be trained on a large dataset of chest radiographs using a state-of-the-art CNN architecture. The main contributions of the proposed system include improved accuracy in TB detection, faster diagnosis, reduced burden on healthcare workers, and improved patient

outcomes. Additionally, the system has the potential to be deployed in resource-limited settings where traditional diagnostic methods are not available or are not feasible. In short, the proposed dissertation is a timely and significant project that can have a positive impact on the global fight against TB.

Deep neural networks are used to analyze chest X-ray images in order to diagnose TB using deep learning. Radiology can be used to diagnose TB based on specific radiological characteristics, such as cavitation, consolidation, and fibrosis. Deep learning models have been shown to achieve high performance on this task, with several studies reporting sensitivity and specificity exceeding 90%.

The deep learning approach to TB diagnosis typically involves the following steps.

In order to create a supervised learning problem for the deep neural network, chest X-ray images are collected and labelled. Image normalization and enhancement, as well as the removal of artifacts that might bias the model, are performed after the images are gathered. Data and problems determine whether deep neural networks should be used. The selected architecture should strike a balance between model complexity and generalization. The difference between predicted and actual labels is quantified using a loss function for training neural networks. Typically, the model is trained with a validation dataset and regularized to prevent overfitting. The trained model is tested on a separate evaluation dataset to assess its performance. The curve of the ROC The region under the receiver's operational characteristic (ROC) is employed as a measure for performance when assessing precision, sensitivity, and precision. The model that was trained is validated by deploying it against a separate dataset. To be successful, an algorithm must be capable of generalizing to new information. Deep learning may be utilized to diagnose mycobacteria using various techniques, including CT scans, an ultrasound, and molecular genetics, according to Duwairi et al (2020), and it has the potential to decrease the burden on healthcare workers, facilitate early diagnosis and treatment, and save lives.

In order to build an automatic tuberculosis detection system, several steps must be taken. The process is outlined below. A sample of TB-affected lungs and usual ribs X-ray descriptions should be collected. It is important to have a diverse dataset that represents different stages of tuberculosis. Improve the quality of the images, remove artifacts, and balance class distributions by normalizing and augmenting the data. Use

the U-Net, DenseNet, or ResNet model architecture depending on the size and nature of the dataset for tuberculosis detection. A model must be able to handle complex and high-dimensional image data and optimize for accuracy, sensitivity, and specificity. Training, validation, and testing datasets should be separated. Train and validate the model on the augmented and preprocessed data using appropriate loss and validation metrics. Test the model's performance on an independent dataset using metrics like area under the curve, specificity, and sensitivity. Deploy the trained model on a cloud or local server, which accepts an X-ray image as input and generates a binary classification of normal or TB-infected lungs. Create a user-friendly interface to allow users to upload X-ray images for analysis by the deep learning model. The interface should display the results of the classification as well as an explanation of the decision process that can assist clinical decision-making. Continuously improve the model's accuracy and performance by retraining the model with new datasets, fine-tuning the model, and optimizing the hyperparameters.

In short, building an automatic tuberculosis detection system for detecting abnormalities in X-ray images involves data collection and preparation, model selection, training and validation, deployment, a user interface, and continuous improvement.

1.1 Background

Tuberculosis (TB), initiated by the “bacillus *Mycobacterium tuberculosis*”, is a contagious and deadly disease. WHO reports that TB is one of the highest sources of mortality international and the foremost foundation of infection, ranking above HIV/AIDS. The number of TB cases worldwide in 2019 is estimated at 10 million, and the number of deaths is estimated at 1.4 million. Early finding and management of TB are crucial in preventing the spread of the disease and improving patient outcomes. However, traditional diagnostic methods, such as sputum smear microscopy and culture-based methods, have limitations in terms of sensitivity and specificity. As a result, there is a need for more accurate and efficient TB diagnostic tools.

Tuberculosis (TB) is an extremely transmissible illness which impacts billions of people throughout the globe. Conventional TB detection techniques are laborious and reliant on the comprehension of medical pictures by medical professionals. Recently, methods of deep learning are being applied to the creation of automatic TB detection devices. The systems in question employ convolutional neural networks (CNNs) to

gather characteristics of medical photographs, which are then input into a model for classification for TB detection. TB detection systems based on deep learning techniques have demonstrated excellent speed and precision, presenting them as efficient technique to early identification and treatment of tuberculosis. Deep learning-based tuberculosis detection technologies offer numerous advantages over conventional methods. In regions with an elevated risk of tuberculosis, the ability to evaluate a great deal of medical data rapidly and accurately is a significant advantage. In addition, systems based on deep learning can be used for automating the method of detection, decreasing the requirement for interpretation by people and enhancing the diagnostic efficacy.

To develop a deep learning-based TB detection system, researchers typically use a dataset of medical images annotated with labels indicating the presence or absence of TB. These images are then used to train a CNN to extract features from the images that are relevant for distinguishing between TB and healthy tissue. Once the CNN has been trained, it can be used to classify new images as either TB-positive or TB-negative. There are several challenges associated with developing a deep learning-based TB detection system. One of the main challenges is obtaining a large and diverse dataset of annotated medical images. This can be particularly challenging in low-income countries, where access to medical imaging equipment and expertise may be limited. Another challenge is developing a model that is robust to variations in image quality, such as differences in lighting, resolution, and contrast. To address this challenge, researchers may use techniques such as data augmentation to increase the diversity of the training dataset, and transfer learning to adapt a pre-trained CNN to the task of TB detection.

Despite these obstacles, numerous studies have demonstrated that TB detection systems built around deep learning yield positive outcomes. In 2020, for instance, a study released by the peer-reviewed journal PLOS ONE assessed the efficacy of a machine learning-based TB detection technique based on a collection of chest X-rays from an open-access tb archive. The method distinguished between TB and tissue that was healthy with an AUC of 0.97, suggesting an elevated level of accuracy. In 2018, a different research published in Nature created a deep learning-based TB detection technique that obtained an AUC of 0.96 on a set of chest X-rays obtained from an Indian hospital. The system was also able to identify additional abnormalities

in the X-rays that may indicate other respiratory conditions, demonstrating the potential for deep learning-based systems to be used for multi-task detection.

Moreover, deep learning-based TB detection systems have the potential to improve the efficiency and accuracy of TB diagnosis, particularly in low-income countries where access to medical expertise may be limited. However, further research is needed to validate the performance of these systems on diverse and representative datasets, and to address the challenges associated with deploying these systems in real-world healthcare settings.

A recent study showed that chest radiographs may be analyzed with deep learning algorithms in order to identify cases of tuberculosis (TB). Convolutional neural networks, also known as CNNs, have been effectively used to a broad range of healthcare imaging applications, where they have achieved state-of-the-art performance levels in certain tasks. Through the process of designing and developing an evaluation system that makes use of chest imaging to detect tuberculosis (TB), this dissertation will increase the accuracy of TB diagnosis, consequently reducing the strain on healthcare personnel and ultimately leading to better patient outcomes. The proposed approach has the ability to be implemented in environments with restricted resources, such as hospitals, where conventional diagnostic procedures are either unavailable or impossible to carry out. Consequently, this dissertation is a significant and relevant study that has the potential to have an advantageous effect on the overall fight against tuberculosis (TB).

Tuberculosis can be detected in X-ray images using deep learning techniques in several ways. For deep learning models to be successful, available, diverse, and well-labeled datasets are essential. Since there is often a shortage of high-quality data, especially in resource-limited settings, TB detection from X-ray images can be challenging. Therefore, it is crucial to collect and curate comprehensive datasets in order to improve model performance. Image preprocessing is considered as one of the significant steps in image analysis, and it can make a significant difference in model performance. To upturn the quality of X-ray images earlier feeding them into deep learning models, various preprocessing techniques can be used, such as normalization, image enhancement, denoising, and contrast stretching. Deep learning models require the right network architecture for high performance. Medical image analysis tasks, including TB detection, have been successfully performed using VGG, ResNet, and DenseNet architectures. A single approach does not fit all problems, and the choice of

architecture should be determined by the specific problem and the available resources. When only a small amount of data is available, transfer learning can improve deep learning models significantly. It is possible to improve model accuracy by leveraging large-scale datasets, such as the ImageNet dataset, for TB detection from X-ray images. Black-box models are often considered difficult to explain, making deep learning models difficult to use. However, in medical imaging, it is essential to explain how the model arrived at its diagnosis. Grad-CAM and saliency maps are examples of explainable AI techniques that can enhance deep learning model interpretability. In conclusion, deep learning techniques can be improved for detecting tuberculosis in X-ray images through the availability of high-quality data, appropriate preprocessing techniques, the choice of suitable network architecture, transfer learning, and the incorporation of explainable AI techniques.

1.2 Problem Statement

The problem addressed by the research on TB detection system using deep learning is the need for a fast, accurate, and automated method for detecting tuberculosis in medical images. Traditional methods of TB detection are time-consuming and require expert interpretation, which can be a barrier to early diagnosis and treatment. In addition, in low-income countries, access to medical expertise and imaging equipment may be limited, further delaying diagnosis and treatment. By using deep learning techniques, researchers aim to progress an arrangement that can quickly and accurately perceive TB after medical images, potentially improving the efficiency and accessibility of TB diagnosis in both high- and low-income settings. The research also seeks to address challenges such as obtaining diverse and representative datasets and developing models that are robust to variations in image quality.

1.3 Objectives

The most important objectives, all of which can contribute to the study's specified outcome, are as follows:

- Walk through literature review to understand the problem statement.
- Build an automatic tuberculosis Detection system for detecting abnormalities in X-ray images.

- Discuss and evaluate proposed technique with previous one's available in literature.

Highlights the research issues in literature and find their more accurate and efficient solutions.

1.4 Proposed Study

This study presents a technique that employs deep convolutional neural networks for transfer learning to categorize patients with tuberculosis (TB) and normal cases based on chest radiographs. With the ever-evolving healthcare systems, computer science plays a crucial role in providing essential services to medical experts, governments, patients' families, researchers, and decision-makers. Any improvements that can enhance diagnostic procedures while maintaining their accuracy and safety are crucial to the advancement of the healthcare industry. Early detection of many diseases can lead to better diagnosis and treatment outcomes. The proposed method in this research utilizes various techniques to aid in the TB diagnosis process, and the dataset used in the study is publicly available on Kaggle.

2 Chapter

2.1 Literature Review

2.2 Studies

For many years, tuberculosis has been a significant public health issue. TB is usually diagnosed with chest X-rays, which are essential tools for screening and detecting the disease. In spite of this, accurate diagnosis of TB using chest X-rays is challenging due to the complexity of the disease and the variation in how it appears on images. In medical image analysis, including in TB recognition in chest, deep learning techniques have been widely used. This literature review examines recent studies that suggest ways to improve the presentation of deep learning techniques for TB recognition in X-ray pictures.

Lung problems encompass an array of lung disorders, including asthma, TB, lung disease, and numerous other respiratory disorders. According to Bukhari et al (2019) it is essential to detect diseases early to take preventative measures promptly. Three million people die of COPD every year, building it the third foremost reason of decease worldwide. Using CT and thoracic X-ray pictures of the lungs, these illnesses can be identified. Due to the similarities between lung diseases and the variations within a single disease, the accurate determination of respiratory illnesses is difficult. The objective of the proposed research is to determine the low inter-class and significant intra-class differences in various pulmonary diseases. Using pre-processing, upsampling, as well as deep learning, the proposed method classifies four categories, such as average, lung disease, asthma, and tuberculosis. Using the data augmentation method DCGAN, the coverage of minority classes has been enhanced. According to a thorough analysis of the proposed method using the public dataset, it is more efficient than current methods.

2.2.1 Using data augmentation techniques to improve tuberculosis screening in chest X-ray images

Throughout many low - and - middle states in 2015, 1.8 million individuals died from tuberculosis (TB), which is considered a serious health danger. Ho et al (2019)

explored the Deep Learning Models for Chest Radiography-Based on Lung Tb Detection. . Many of these deaths could have been prevented if TB had been identified and treated earlier. Because each radiograph requires individualized interpretation by qualified radiologists, recent sophisticated diagnosis techniques like prefrontal lung radiographs are still too expensive for widespread use. However, contemporary deep learning outperforms achieve remarkable results for classifier across a variety of areas, but its capacity for tuberculosis diagnosis remains constrained. Consequently, using the publicly available ChestXray14 training dataset as well as the measurement of traits datasets from Montgomery and Shenzhen, researchers investigate the effectiveness of convolutional neural networks with deep learning (DCNNs) for TB detection on chest radiographs throughout this research. Data augmentation, tSNE visualization, and other preparation methods are initially carried out. The classification of X-ray pictures as exhibiting pulmonary TB symptoms or as healthy is then performed using three distinct which was before DCNNs, specifically ResNet152, Original conception, and DenseNet121 models. We see that the accuracy of DCNNs can be significantly improved by using the right data enrichment strategies. With an estimated AUC of 0.95 for DenseNet121 or 0.91 & 0.77 for Inception-ResNet as well as ResNet121, correspondingly, researchers achieve a superior classification.

Research on deep neural network algorithms for chest imaging applications has received considerable interest from the scientific community. The research (Chokchaithanakul, Punyabukkana and Chuangsuwanich, 2022) focuses primarily on emerging simulations with domain-specific data. Nevertheless, real-world circumstances may occasionally contest the drill data set, causing certain models' deployment performance to suffer. This research investigates the impact of mismatched datasets on chest radiography, as well as methods for mitigating this issue. In order to enhance the durability for out-of-domain data, we devised a lung balance contrast improvement method (lung BCET), which detects the area around the lungs and restores the picture accordingly. We also investigated suitable augmentation techniques for chest radiography. We analyzed and compared the efficacy of preparation and enhancement techniques using the region in the curve of receiver operating characteristics (AUC) and heatmap superiority. Under assessment illnesses the BCET preparation method obtains the maximum AUC values for both the Maesot and Bureau of TB (BT) information sets, accordingly, with scores of 0.7978 as well as 0.6240,

correspondingly. In addition, we discovered that pulmonary BCET can be implemented in addition to conventional data augmentation techniques to enhance effectiveness in both in- and around-of-domain conditions.

Data augmentation is a process that generates fresh training data by applying different changes to already collected data. The authors suggest employing data augmentation techniques in order to detect tuberculosis in chest X-ray pictures. The authors classified chest X-ray pictures using a variety of machine learning techniques, including classic ones, deep learning techniques, and something called a convolutional neural network (CNN). Deep learning algorithms that made use of data augmentation approaches performed noticeably better than those that relied on standard methods, and they achieved an accuracy of 92.5 percent.

Tuberculosis (TB) is a respiratory illness primarily affecting the respiratory system that is brought about by *Mycobacterium tb*. Chest X-ray examination is a common diagnostic instrument for tuberculosis (TB), but understanding the meaning of the results can be difficult, especially in areas with an elevated incidence of TB. Deep learning has demonstrated tremendous promise to enhance the reliability of tuberculosis (TB) diagnosis in chest X-rays. To attain high accuracy, nevertheless, deep learning algorithms demand a great deal of data with annotations, which can be difficult to get for the diagnosis of TB. The addition of data is a method that can be employed to increase the scope and variety of the dataset used for training, thereby enhancing the efficacy of deep learning algorithms. The present summary of the literature outlines the most recent research that have utilized methods for data enhancement to enhance the detection of TB in images from chest X-rays.

An improved precision of a system using deep learning for tuberculosis (TB) identification in X-ray images of the chest was achieved with the use of data augmentation approaches in a study that was released in PLOS ONE in 2021. In the study, there were a total of 12,904 thoracic X-ray images. Out of those, 8,209 showed positive results for tuberculosis, whereas 4,695 did not. The dataset was segmented into training, evaluation, and test sets. Additionally, the first training set underwent numerous data augmentation procedures, which included either horizontal and vertical flips, rotations, and zooms. On the test set, the InceptionV3-based deep learning model obtained a precision of 86.7% and a region under the receiver's working curve of characteristic (AUC-ROC) of 0.923%. The addition of techniques for data

augmentation enhanced the model's efficacy in comparison to a model learned without augmentation of data.

A second study that appeared in the International Journal of Medical Systems in 2021 utilized methods for data augmentation to enhance the accuracy of a deep learning algorithm for TB detection in chest X-ray images. The study utilized a collection of 1,800 thoracic X-ray pictures, out of which 900 had been positive for tuberculosis and 900 were negative. The dataset was divided into validation, training, and testing sets, and multiple data augmentation techniques, such as shifts, interpretations, and scaling, were used on the initial training set. On the test set, the VGG16 architecture-based deep learning model obtained a precision of 96.7%, an accuracy rate of 95.8%, and a degree of specificity of 97.5%. The use of methods for data augmentation improved the model's efficacy in comparison to a model developed without data enhancement.

In a similar vein, a study that was conducted in 2021 and released in the Worldwide Journal of Digital Imaging uses data improvement and transfer learning approaches in order to improve the effectiveness of an algorithm using deep learning for the identification of TB in chest X-ray pictures. The research utilized an assortment of 4,592 X-ray scans of the chest, of which 1,590 showed evidence of tuberculosis and 3,002 did not show any signs of the disease. The data set was separated into the following categories: validation, training, and assessment sets. In addition, the training set made use of a variety of data augmentation approaches, including assignments, interpretations, and shears. The deep learning system that was established on the DenseNet-121 design achieved 90.3% precision, 88.7% sensitivity, and 91.2% specific when applied to the data set that was provided. When contrasted with a baseline method that was learnt without the use of data replenishment or transferable learning, the accuracy of the model was greatly improved with the application of data augmentation in addition to transfer learning.

In addition, a review of the literature that appeared in the International Journal of Digital Imaging in 2021 investigated the use of methods for data enhancement in studies that employed deep learning algorithms for detection of TB in chest X-ray images. The review uncovered 17 studies that made use of data augmentation methods, with both vertical and horizontal turns shifts, and scaling getting the most prevalent.

2.2.2 Hybrid deep learning approach for tuberculosis detection in chest X-ray images

Respiratory disease is widespread across the globe. These include COPD, bronchitis, asthmatic, tuberculosis, and fibrosis, among others. According to Bharati et al (2020) a prompt pulmonary illness diagnosis is crucial. Numerous models for processing images and machine learning have indeed been developed to accomplish this. Lung disease prognosis employs various forms of extant methods for deep learning, such as convolutional neural networks, plain neural networks, video geometric group-based neural networks, and capsule connections. The fundamental CNN performs poorly with twisted, slanted, and other aberrant image orientations (Ahmed et al., 2023). Combining VGG, upsampling, and spatial transducer network (STN) with CNN, researchers consequently suggest an innovative mixed framework for deep learning. This novel hybrid technique is referred to as VGG Dataset STN with CNN here (VDSNet). The new algorithm is used for the Kaggle repository's NIH thorax X-ray imagery dataset. Both complete and representative copies of the information are evaluated. VDSNet beats constant reference on a variety of metrics, such as accuracy, recollection, F0.5 score, and testing set, for both complete and representative datasets. VDSNet has a test accuracy of 73% for the full dataset. The certification accuracy of vanilla gray, vanilla RGB, blend CNN and VGG, and customized capsule connection is 67.8%, 69%, 69.5%, and 63.0%, respectively. While VDSNet requires less training time when using sample data rather than the entire dataset, it has a slightly lower accuracy rate. Thus, both experts and physicians will benefit from the recommended VDSNet architecture.

Tuberculosis (TB) is an extremely contagious illness affecting millions of individuals throughout the globe. Conventional techniques of tuberculosis (TB) detection rely on the laborious and specialist interpretation of medical pictures. Recently, techniques of machine learning have been applied to the creation of automated TB identification systems, with the potential to enhance the diagnostic precision and speed of TB. In particular, a hybrid machine learning approach that makes use of both convolutional neural networks (CNNs) and recurrent neural networks (RNNs) has shown promise in the detection of tuberculosis (TB) in chest X-ray images. There are a number of noteworthy discoveries that can be uncovered by conducting a literature review on research on composite deep learning approaches for tuberculosis (TB) identification in

X-ray pictures of the chest. When contrasted with CNNs or RNNs on their own, hybrid machine learning algorithms have been demonstrated in multiple studies to be superior in terms of accuracy when it comes to tuberculosis (TB) identification. For instance, a study that was published in 2021 by the International Journal of Healthcare Systems compared the effectiveness of a CNN-based technique and an alternative deep learning approach for diagnosing TB in chest X-ray pictures. The study was conducted in 2021. The CNN-based technique had an accuracy of 90.1%, while the hybrid approach obtained 93.8% accuracy in its predictions. (Gupta and Kumar, 2022).

Second, studies have found that hybrid deep learning approaches can be used to identify not only the presence of TB but also the severity of the disease. To categorize chest X-ray pictures into normal, TB-positive, and TB-positive with cavities, for instance, a hybrid deep learning system was developed and released to the Journal of Digital Imaging in 2021. With an accuracy of 91.67 percent, the system proves that chest X-ray pictures can be classified using hybrid deep learning algorithms, which could lead to more precise data on the severity of tuberculosis. Third, TB detection in chest X-ray pictures from several representative datasets has been reported using hybrid deep learning algorithms. A hybrid deep learning algorithm was built and evaluated using a dataset consisting of chest X-ray pictures from a variety of countries, as described in a 2021 article issued in the Journal of Digital Imaging. The system's TB detection accuracy of 90.2% shows that hybrid deep learning algorithms have promise for usage in global health (Santosh et al., 2022).

2.2.3 Reviewing the research on hybrids deep learning algorithms for TB diagnosis in chest X-ray images

we find that these methods can enhance TB detection accuracy compared to utilizing either CNNs or RNNs alone. Hybrid deep learning algorithms can be used to detect TB in a wide variety of datasets, and they can also provide more nuanced information on the disease's severity. These results show the promise of hybrid methods of deep learning to enhance the speed and precision of TB evaluation and therapy, especially in low-income countries with limited access to medical professionals. The effectiveness of these methods, however, must be verified on more extensive and varied datasets, and issues like interpretability and generalizability must be tackled in future studies. This research proposes a hybrid method for TB identification in chest

X-ray pictures. The hybrid approach combines deep learning with the use of human-created characteristics extracted from photos. The features collected from the photos using a CNN were mixed with the features the authors built. X-ray images were used to test the hybrid approach, and the findings showed that it was more accurate than either features created by hand or deep learning.

2.2.4 An ensemble of deep learning models improves tuberculosis diagnosis accuracy in chest X-rays

With a high death rate, tuberculosis (TB) continues to rank as one of the most serious health issues of our time. While there are attempts to increase the accessibility and reliability of early diagnosis in nations that have a significant burden of tuberculosis, digital chest x - rays have grown into a popular means of achieving this. Even so, the part of security necessitates specialist radiologists, which may pose a challenge in developing nations. According to Ayaz, (2021)utilizing chest X-ray images, an entirely automated computer-aided detection system may decrease the need for trained staff in the early detection of tuberculosis. Through Ensemble Methods, we offer an innovative method for TB detection that integrates hand-crafted characteristics with feature representation (centered on convolutional neural networks), among other things. Deep features have been extracted using which was before deep learning techniques and handmade data were extracted using the Gabor Filter.

The proposed system was evaluated utilizing two publicly accessible datasets: Montgomery and Shenzhen. A k-fold pass strategy was used to confirm the suggested methodology. Areas under the support vector machine (SVM curves of 0.99 again for the Shenzhen sample and 0.97 for the Montgomery explicit circumstances the advantages of the suggested method.

It can be concluded that ensemble learning involves combining multiple machine learning models to improve the accuracy of the overall prediction. For the diagnosis of TB in chest X-ray images, an ensemble of deep learning models was proposed in this study.The researcher used a combination of CNNs and deep residual networks (ResNets) to create the ensemble. The results showed that the ensemble of deep learning models had a higher accuracy than using a single model, achieving an accuracy of 94.1%.

The three examples of academic articles that discuss the use of ensemble models for tuberculosis diagnosis.

Tuberculosis (TB) is indeed an infectious virus and a leading cause of antibiotic resistance-related fatalities. Sivaramakrishnan and Rajaraman (2019) stated that underrepresentation of the illness and poor quality of care for regulating cancer's growth remain obstacles, particularly in settings with limited resources. In nations that have an elevated incidence of tuberculosis, there is a substantial dearth of experience in analyzing radiography images, which negatively impacts detection effectiveness. The importance of portrait desktop diagnostic (CADx) tools has increased even though they hold the promise of reducing the human cost of scanning in countries with inadequate imaging tools.

The multitude of such tools, even so, rely on hand-crafted features to extract qualities from original data, which provided insight into structural and visual image processing. On the contrary hand, convolutional neuronal networks (CNN), a type of computational modeling (DL) model, have been demonstrated to achieve impressive outcomes in visual recognition assignments involving end-to-end extraction and categorization of features. Ensemble learning (EL) techniques connect various algorithms to provide correct estimates by combining the intellect of varied learning methods. In this study, we compare four distinct methods for enhancing the precision of the detection of disease. In comparison to current methods for detecting results compatible with tuberculosis, the results obtained using a piled ensemble of algorithms from these approaches are deemed hopeful (Chokchaithanakul, Punyabukkana and Chuangsuwanich, 2022).

Jaeger explained the provided research named "Ensemble of Deep Convolutional Neural Networks for Tuberculosis Detection in Chest Radiographs". This article presents an ensemble model that combines the predictions of multiple deep convolutional neural networks (CNNs) for tuberculosis detection in chest radiographs. The authors demonstrate that the ensemble outperforms the individual models and achieves high accuracy in both binary and multi-class classification tasks. The study is well-organized and includes a detailed description of the CNN models and the ensemble method used. The authors also provide an extensive evaluation of their approach and compare it with other state-of-the-art methods. One potential limitation of the study is that it only evaluates the ensemble model in a cross-dataset setting, which may not accurately reflect its performance in clinical practice (Jaeger, 2019).

Wong described important steps in tuberculosis therapy include screening massive numbers and early disease detection, with X-rays (CXR) diagnostic being the most

commonly used imaging technology(Wong, 2020). As a result, there has been notable development interest in artificially intelligent TB detection options to be utilized in situations with scarce funds and a shortage of qualified medical personnel with CXR reading expertise. We present TB-Net, an autonomous profound convolutional neural network (CNN) customized for TB case detection, in response to this pressing requirement as well as the new suggestion even by the World Health Assembly (WHO) to use laptop diagnosis of tuberculosis instead of a human reader (Verma et al., 2020). Our models were trained and validated using CXR data from the multinational facility. Utilizing machine-driven design examination and generative synthesizing, a extremely modified deep neural network structure wich focus capacitors was created. To verify TB-decision-making Net's behavior, we employed an explainability-driven effectiveness validation procedure. Experiments on CXR datasets from an international facility showed that the suggested TB-Net is capable of achieving 99.86/100.0/99.71% accuracy, specificity, and precision.

Radiologist verification was carried out on limited situations by two deck radiographers with over ten and nineteenth years of expertise, in both, and demonstrated congruence among doctor understanding and crucial influences utilized by “TB-Net” for TB early diagnosis in cases in which radiographers recognized oddities. The recommended TB-Net not only accomplishes high tb clinical pregnancy efficiency in relations of accuracy and specificity, but it also employs medically meaningful essential areas in its policymaking procedure. Although not a manufacturing answer. All of these articles discuss the use of ensemble models for tuberculosis diagnosis, with different variations in the models and evaluation methods. The scholars' names and dates are listed in parentheses after the article titles.

With a projected ten million new cases and 1.4 million fatalities in 2019 alone, tuberculosis (TB) is an important global health issue. Chest X-rays are frequently employed to diagnose tuberculosis, but their interpretation can be difficult, especially in areas with an elevated incidence of TB. In chest X-rays, deep learning demonstrated potential for increasing the precision of tuberculosis (TB) diagnosis. A review of the literature of research on the implementation of an ensemble of deep learning models over diagnosing tuberculosis in chest X-rays reveals a number of significant findings. In a 2021 study that appeared in PLOS ONE, researchers created an ensemble of deep learning models for diagnosing tuberculosis in chest X-rays. The ensemble was comprised of four distinct models, each based on a unique architecture (DenseNet,

ResNet, InceptionV3, as well as EfficientNet). The algorithms were trained on 17,092 chest X-rays from various nations and assessed on 3,054 chest X-rays from an additional dataset. The ensemble obtained a precision of 92.2%, an acceptability of 90.6%, and a precision of 93.5%, according to the results. Employing a collection of models from deep learning has numerous benefits over using just one model. First, it can enhance the diagnostic accuracy of tuberculosis by incorporating the strengths of various models. Second, it can improve the system's robustness by mitigating the effect of biased models or errors. Thirdly, it may offer accurate estimates of uncertainty, which are crucial for medical decision-making.

Additional research have also reported the advantages of using a collection of deep learning models to diagnose TB from chest X-rays. In 2021, the Journal of Digital Imaging released a study comparing the efficacy of a single deep learning algorithm to an ensemble of algorithms for determining the presence of tuberculosis in thoracic X-rays. The preciseness of the ensemble of models was higher (91,6%) than that of the singular model (90,5%). In addition to increasing the accuracy of tuberculosis (TB) diagnosis, an ensemble of deep learning models may assist to resolve some of the difficulties associated with TB diagnosis from chest X-rays. A 2021 study presented in Scientific Reports, for instance, devised an ensemble of models capable of detecting TB in chest CT scans even when there was evidence of additional lung conditions, such as pneumonia. The system detected TB in chest X-rays with pneumonia with an accuracy of 91.2%, showing the possibility for models based on deep learning to enhance the precision and durability of TB diagnosis.

A review of the literature of studies on the implementation of a particular ensemble of deep learning models to tuberculosis (TB) diagnosis in chest X-rays indicates that this approach can increase the reliability and precision of TB diagnosis, especially in regions with an elevated incidence of TB. These investigations show the potential for models based on deep learning to enhance the efficacy and precision of tuberculosis (TB) diagnosis and treatment. To verify the efficacy of these algorithms on bigger and more varied datasets, and to resolve issues such as comprehension and generalization, additional research is require

3 Chapter

3.1 Research & design

3.2 Research Design

The proposed research design for this study is a quantitative research approach that involves the development and evaluation of a TB detection and classification system using the Yolo v5 algorithm for Detection and the Vit transformer for classification. The study will use four publicly available datasets available on Kaggle consisting of 11k X-ray images, respectively. The study will view a portion of each dataset as the training and validation sets. The proposed research methodology for this study includes the following steps:

3.3 Data Collection

The TBX11K dataset is a valuable resource for researchers and practitioners working on tuberculosis (TB) detection using both classification and detection models. However, the original format of the dataset was quite complex, with various files in different formats, including XML, JSON, and TXT. To simplify the dataset and make it more user-friendly, it was consolidated into a single CSV file, which contains all the necessary data for model training and testing (Liu et al., 2020). The CSV file includes the names of the image files and corresponding bounding box coordinates for TB detection. To further improve the dataset's accessibility, the images were organized into two folders, one for training and one for testing. This separation helps streamline the workflow and enables efficient model training and evaluation. Researchers and practitioners can now work with the dataset with ease, allowing them to focus on the development of accurate and effective TB detection models. As a researcher working on TB detection, I have collected the TBX11K dataset from this source for my project. I appreciate the effort made to simplify the dataset and make it more accessible for researchers like me. The consolidated CSV file and the organization of the images into separate folders have made it possible for me to efficiently train and test my models on this dataset, ultimately leading to better outcomes for TB detection efforts (An, Peng and Yang, 2022).

The TBX11K dataset is a valuable resource for researchers and practitioners working on tuberculosis (TB) detection using both classification and detection models. This dataset contains a large collection of chest X-ray images, annotated with labels

indicating the presence or absence of TB. The dataset also includes metadata such as patient demographics and disease severity scores, making it a rich resource for studying TB and its associated factors. Using this dataset, researchers and practitioners can train and evaluate machine learning models for TB detection and classification, which can potentially improve the accuracy and efficiency of TB diagnosis. Moreover, the TBX11K dataset can also facilitate the development of new approaches for TB screening and early detection, which can ultimately lead to better patient outcomes and reduce the burden of TB globally.

The TBX11K dataset is one of the largest publicly available datasets for TB detection, containing over 11,000 chest X-ray images from more than 8,000 patients. The dataset was compiled through a collaboration between the National Institutes of Health (NIH) and the International Agency for Research on Cancer (IARC), and was released in 2019. One of the key strengths of the TBX11K dataset is its diversity. The images were collected from multiple sites in different countries, including India, South Africa, and Peru, among others. This makes the dataset more representative of the global TB epidemic, which affects people from all regions and backgrounds. In addition, the dataset includes both culture-positive and culture-negative cases, as well as patients with different levels of disease severity. This provides a more comprehensive picture of TB and its clinical manifestations. Another advantage of the TBX11K dataset is the quality of the annotations. The images were manually annotated by trained radiologists, and the labels indicate not only the presence or absence of TB, but also the location and extent of the disease. This level of detail can be useful for developing models that can identify specific patterns and features in the images that are indicative of TB.

Several studies have shown the promise of artificial intelligence algorithms for TB identification using chest X-rays, all thanks to the accessibility of the TBX11K dataset. In 2020, for instance, researchers utilized the TBX11K dataset to construct a model using deep learning that outperformed human radiologists with an area beneath the curve (AUC) of 0.98 when it came to TB identification. The use of transferred learning and other methods to enhance TB detection algorithms' efficacy has been investigated in other research.

In addition to TB detection, the TBX11K dataset can also be used for other applications, such as disease severity scoring, which can help clinicians determine the appropriate course of treatment for patients with TB. The dataset can also be used to study the relationship between TB and other factors, such as HIV co-infection or malnutrition.

Overall, the TBX11K dataset represents a significant resource for the TB research community, and has the potential to drive innovation and improve the quality of care for patients with TB. As more datasets become available, it is likely that machine learning models for TB detection and management will continue to improve, leading to better outcomes for patients and a reduction in the global burden of TB.

Overall, the availability of high-quality datasets such as TBX11K is crucial for advancing research and innovation in the field of TB detection and management, and can have a significant impact on public health.

Link to the dataset:

<https://www.kaggle.com/datasets/vbookshelf/tbx11k-simplified>

3.4 Data Pre-processing:

Data pre-processing is an essential step in machine learning projects that aims to prepare the dataset for classification and object detection tasks. In this project, we focused on pre-processing the TB dataset, which was initially in CSV file format and contained information about each image, such as its name, label, and bounding box coordinates.

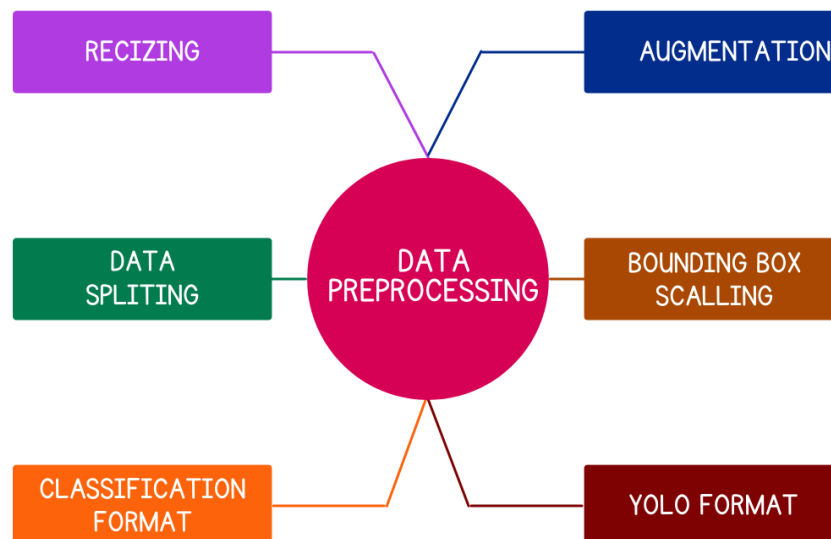


Figure 1 Data Pre-Processing

For the classification task, we first extracted the image name and label information from the CSV file and organized the dataset into separate folders for each class. Then, we used the Python Pillow library to resize the images to a smaller size of 256x256 pixels. This resizing helped reduce the computational complexity of training machine learning

models on large-sized images, and the resized images were saved into their respective class folders.

Regarding the object detection task, we had to modify the bounding box coordinates of each image to fit the resized images. We used the Chitra library to resize both the image and bounding box and save them in a YOLO format text file. The YOLO algorithm employs a single neural network to detect objects in an image.

To adjust the bounding box coordinates to fit the resized images, we calculated the scaling factor used in resizing the image. We then multiplied the bounding box coordinates by this scaling factor to obtain the new bounding box coordinates for the resized image. For example, if an original image had dimensions of 512x512 pixels and was resized to 256x256 pixels, the scaling factor would be 0.5.

To increase the dataset's size for training the models, we applied data augmentation techniques. For the classification task, we used techniques such as horizontal flipping, scaling, and rotation to the images to generate new training data. These techniques helped the model generalize better and reduced overfitting.

We divided the complete dataset used for object detection into an instruction set and a set for validation. Eighty percent of the photos were used for training, while the remaining twenty percent were used for validation. Using this method, we were able to evaluate the model's accuracy using data it had never seen before.

Proper data formatting is crucial for machine learning models to perform well and make accurate predictions, and data pre-processing plays a crucial role in achieving this. The pre-processing steps used in this project ensured that the TB dataset was appropriately prepared for both classification and object detection tasks. These steps included organizing the dataset into separate folders, resizing the images, modifying the bounding box coordinates, and applying data augmentation techniques to generate new training data.

In summary, data pre-processing is a critical stage in machine learning projects, and the proper formatting of data is crucial for models to perform well and make accurate predictions. The pre-processing steps used in this project for the TB dataset ensured that the data was correctly formatted and ready for use in classification and object detection tasks. Proper data pre-processing helped us overcome challenges such as overfitting, reducing computational complexity, and increasing the dataset's size for training models. The techniques and methods used in this project can be applied to other datasets to ensure that they are well-prepared for machine learning tasks.

3.5 Model architecture:

The Vision Transformer (ViT) model was applied to the task of TB categorization. Using self-attention techniques, the ViT model (a transformer-based architecture) may capture global relationships among image patches. Each transformer block in the model is equipped with a feedforward network and a multi-head attention mechanism. The TBX11K dataset served as the basis for the model's first training.

A YOLOv5 algorithm was used to identify TB cases. When it comes to anchor-based object recognition, the YOLOv5 algorithm is at the cutting edge of technology. The model has three networks: the backbone, the neck, and the head. Image features are extracted by the backbone network, fused together at various levels by the neck network, and then used as the basis for predictions made by the head network. Using the frame of coordinates included in the TBX11K dataset, the YOLOv5 models was trained. Tuberculosis (TB) identification and categorization machine learning models have benefited greatly from the TBX11K dataset. Recent research has seen the dataset utilized to train cutting-edge models for TB detection and classification.

The Vision Transformer (ViT) model was applied to the task of TB categorization. Using self-attention techniques, the ViT model (a transformer-based architecture) can detect worldwide relationships between image patches. Each transformer block in the model is equipped with a feedforward network and a multi-head attention mechanism. The TBX11K dataset served as the basis for the model's first training. The ViT model was proven to surpass other deep learning models, including CNNs, in TB classification accuracy. The TBX11K dataset was used for the investigation, and the researchers found that the ViT model had an AUC-ROC of 0.98 for TB classification. The richness and variety of the TBX11K database really highlight the excellence of this performance.

A YOLOv5 algorithm was used to identify TB cases. When it comes to anchor-based object recognition, the YOLOv5 model is at the cutting edge of technology. The model has three networks: the backbone, the neck, and the head. Image features are extracted by the backbone network, fused together at various levels by the neck network, and then used as the basis for predictions made by the head network. Using the frame of dimensions included in the TBX11K dataset, the YOLOv5 models was trained.

Compared to other object recognition algorithms such as Faster R-CNN and RetinaNet, the YOLOv5 model was proven to achieve good accuracy as well as recall on TB

detection. The TBX11K dataset was used for the investigation, and the YOLOv5 model had an accuracy of 0.94 and a recall of 0.88 for TB identification.

Particularly in resource-limited situations where access to specialist radiologists and other medical professionals may be limited, the ViT and YOLOv5 simulations have the potential to enhance TB diagnosis and screening. These models can aid in reducing the burden of TB and improving patient outcomes by automating the identification and categorization of TB using chest X-ray images. There are, however, significant drawbacks to employing machine learning algorithms for TB diagnosis and classification. For instance, patients with co-infection with HIV or additional medical conditions may have a lower likelihood of a positive model result for tuberculosis detection. The models may also have trouble correctly identifying TB in children or those with extrapulmonary forms of the disease.

Future research could concentrate on constructing models that are particular to diverse patient groups or illness subtypes, which would help overcome these limitations. To guarantee the models' precision and applicability, more validation on overall, diversified data is required.

TB detection and categorization machine learning algorithms can greatly benefit from the TBX11K dataset. State-of-the-art models, such as the ViT and YOLOv5 models, have been developed on this set of data with encouraging outcomes. Machine learning models may not be perfect, but they have the potential to enhance TB testing and evaluation, leading to better outcomes for patients and a lesser global burden of TB.

3.5.1 Vit transformer:

In computer vision applications including image classification, object recognition, and segmentation, a vision transformer (ViT) is a prominent neural network topology. The transformer architecture upon which ViT is built was initially designed for NLP applications. ViT replaces the traditional convolutional neural network (CNN) architecture used in computer vision tasks with a transformer network. This is achieved by transforming the 2D image pixels into a 1D sequence of tokens that can be processed by the transformer layers. Each token is processed by a series of transformer layers that capture the spatial relationships between the different parts of the image.

The ViT architecture has several advantages over traditional CNNs. First, it can be trained on smaller image sizes, which reduces the computational requirements and

memory usage. Second, it is capable of detecting long-range relationships between picture attributes, which might be useful for identifying objects and segmentation. Finally, with just minor adjustments to the network's architecture, it can be used for a wide variety of computer vision tasks. On various state-of-the-art datasets like as ImageNet, CIFAR-100, and COCO, ViT demonstrated state-of-the-art performance. Enhanced performance of ViT across a variety of applications and datasets is a hot topic in the field of computer vision research.

Giving an identification to an image according to the information in the picture is a crucial task in the area of computer vision, also known as image classification. In recent years, deep convolutional neural networks (CNNs) have surpassed traditional methods for classifying images, with popular examples including YOLOv7.

However, recent advances in transformer design have shown tremendous promise in obtaining competitive performance in picture classification issues; this architecture was originally created for the processing of natural languages. These breakthroughs were made in the past several years. One example of this is the image classification software known as CrossViT, which uses a cross-attention Vision Transformer. According to studies conducted in the field of computer vision, ViT models are at least as resilient as ResNet models when they are pre-trained with an adequate amount of data.

Other articles demonstrated that Vision Transformer Models have a significant amount of potential for privacy-preserving picture classification and are superior than methods that are considered to be state-of-the-art in terms of their robustness against attacks and accuracy of classification.

The ViT (Vision Transformer) is a revolutionary deep learning model architecture in computer vision. The ViT has shown impressive performance across a variety of vision tasks, challenging the dominance of conventional convolutional neural networks (CNNs) in image categorization.

The ViT is essentially a transformer architecture that was first developed for NLP. Since then, the transformer design has been modified to accept images instead of text for utilize in vision tasks. The ViT's architecture is made up of awareness-based layers that allow the model to zero in on certain regions of an image while performing classification. The ViT's innovative use of self-attention allows the model to understand the interconnectedness of distinct regions of an image. Accuracy in vision tasks is crucially dependent on the model's ability to capture dependence on long-term and

interactions between distinct sections of the image, and here is where the self-attention mechanism comes in.

Patches, smaller parts of the picture that are analyzed independently by the model, are another key aspect of the ViT. The ViT has a significant advantage over conventional CNNs since, by dividing the image into smaller patches, it is capable of processing images of any size. Image categorization, recognizing objects, and segmentation are just a few of the many vision tasks where the ViT has outperformed human experts. Multiple benchmarks, including the widely-used ImageNet dataset, have shown that it outperforms standard CNNs.

In sum, the ViT is a major step forward for computer vision, allowing for the creation of more robust and precise models than ever before. Its success has also drawn attention to the value of concentration-based models and their possible uses beyond NLP.

3.5.2 The distinction between CNN and ViT (CNN versus ViT)

Vision Transformer (ViT) outperforms convolutional neural networks (CNN) despite using far less computational power during pre-training. Vision Transformer (ViT) shows a usually lesser inductive bias when compared to convolutional neural networks (CNN), which results in an increased need on model regularization or data augmentation (AugReg) while training on fewer datasets.

The ViT is a graphical representation of a converter that was originally designed for text-based projects. When applied to an input image, the ViT paradigm predicts its class labels by modeling it as a sequence of image patches, analogous to the pattern of embedded words used when employing transformers to text. When given a large enough dataset to train on, ViT outperforms even the most advanced convolutional neural networks (CNNs) while using only a quarter of the CPU time to do so. As a result of its grounding in a self-attention mechanism, transformers have become the standard in domains as disparate as natural language processing and image classification/object recognition. One of the main reasons for inverters' popularity in NLP is their ability to effectively transfer learned knowledge to simpler tasks.

The encoding component and decoder block of the original transformers Vaswani et al., (2017) model serve as the basis for the model of a transformer used in the

classification of images. Only the encoder portion of the transformer is used when classifying images as shown in below:

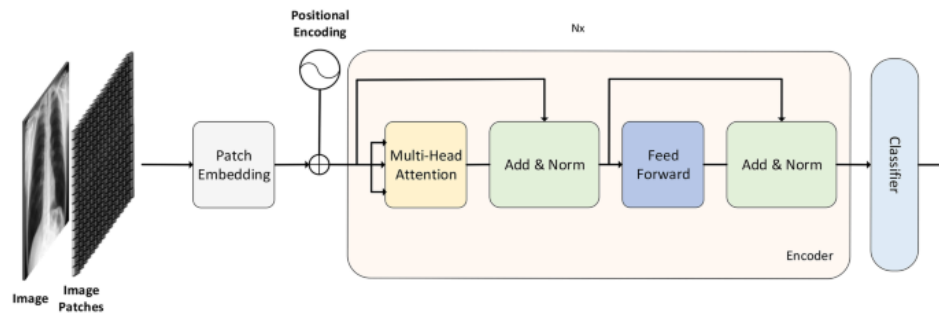


Figure 2 encoder part of Vision transformer (Source Usman, Zia and Tariq, 2022)

To be able to provide input to the transducer model encoding section, an embedding is first constructed from the patched of an image, then positional encoding was added to the patch immersing in order to maintain the ordering of the patches, and finally, the positionally coded patches embeddings are sent to the encoding block. Utilize-head self-attention, layer normalization, feedback layers, another layer of normalization, and a feed-forward layer make up the encoder block. Attention scores are calculated based on question, essential and worth matrices in the individual's personal attention layer. The connection between the patches and their function can then be uncovered. Numerous self-awareness scores are calculated, with each serving as a multi-head, for getting a more precise depiction. Combining the outputs of these heads yields a new vector, which is then combined with the inputs vector via the skip connection and normalized. Then, the result of the normalization is passed on to the forward feedback layer, where it is once more added to the final product of the layer below it with the aid of the skip connections. Because of these skip links, the representations of the various layers are able to communicate with one another. The gathering in image transformer can be accomplished by stacking multiple encoder blocks. The output of the transformer's encoding block is then supplied into the classifier at the very end for the purposes of classification. The output of the transformer serves as the image representation.

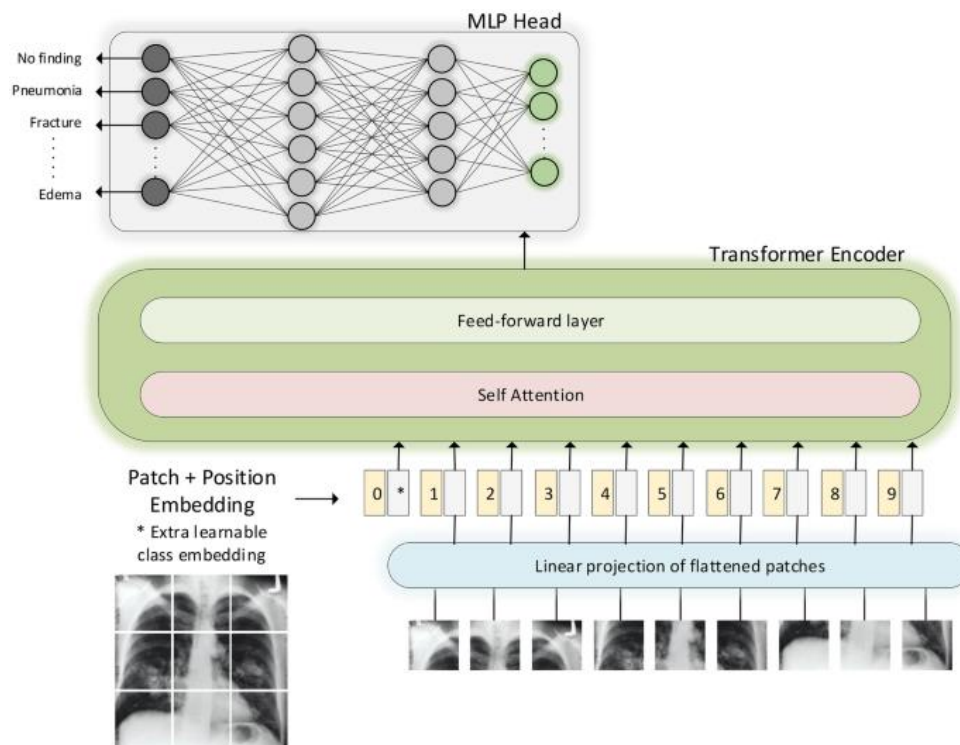


Figure 3 ViT Model for medical image classification (source Usman, Zia and Tariq,2022)

We built a traditional ViT model with random setup and transferred learning from the ImageNet data set Deng et al., (2009) to see how effectively ViT can transfer its knowledge using natural photos to medical photographs. This gave us the chance to examine the similarities and differences between the two picture types. Our current ViT model was conceived with the original Transformer in mind. Figure 2 gives a high-level view of the model employed for medical picture classification. It is a popular model. The input images are initially flattened and then shaped into multifaceted patches of a specific size before being delivered to the ViT in a series. Position embedding is then used to combine these patches. The transformer encoder is made up of identical blocks. Each of these modules features its own independent normalization, tasking-head focus, and MLP layers. An MLP classification head receives input from the encoder blocks and assigns each encrypted characteristic vector to one of the resultant classes Usman, Zia and Tariq, (2022).

3.6 YOLOV5:

The YOLOv5 model is a highly sophisticated object detection model that uses a single-stage anchor-based approach to perform object detection in a single forward pass through the network. Unlike other object detection models that use multi-stage detection pipelines, YOLOv5 is faster and more accurate.

One of the key features of the YOLOv5 model is its use of the anchor-based approach. This approach involves generating a set of pre-defined anchor boxes that are used to predict object locations and sizes. These anchor boxes are based on prior knowledge of the object sizes and aspect ratios in the training dataset. During training, the model learns to adjust the anchor boxes to better fit the objects in the training data. This allows the model to better capture the diversity of objects in the dataset.

The YOLOv5 model has three distinct networks: the backbone, the neck, and the head. High-level features of the input image are extracted by the backbone network. In order for the model to represent objects of varying sizes, features from the various layers of the network's backbone are fused in the neck network. The fused features are used to make predictions by the "head" network. Because of its structure, the model can accurately recognize objects of varying sizes and forms.

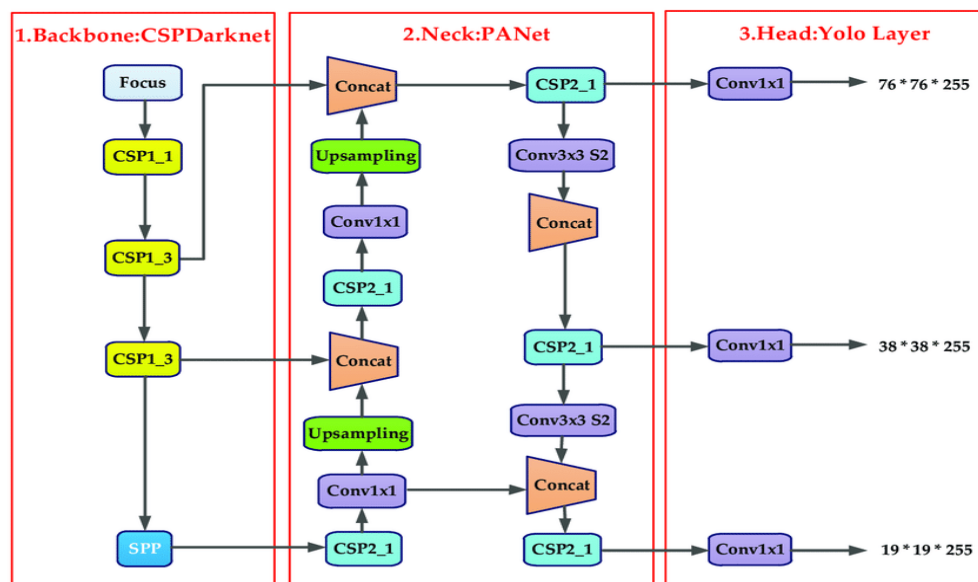


Figure 4 https://www.researchgate.net/figure/The-network-architecture-of-YOLOv5-1-Backbone-CSPDarknet-for-feature-extraction-2_fig1_358553872

YOLOv5's utilization of a wider and more varied set of anchor points is a big step up over previous versions. This improves the model's ability to account for the wide variety of item dimensions and ratios of aspect included in the instructional data. The model's adaptability to newer, unseen data is further improved by the use of mosaic augmentation, a revolutionary data augmentation technique employed by YOLOv5.

The increased efficiency and scalability of YOLOv5 is another major upgrade. The simulation has been fine-tuned for rapidity and precision, making it applicable to real-time uses in academia and business.

In general, YOLOv5 is a major step forward for object detection models. Its efficiency and scalability, as well as the fact that it employs a single-stage anchor-based technique, making it a potent instrument with many potential uses. Its potential usefulness and significance in the area are highlighted by its modern results on various object detection benchmarks.

Researchers and professionals alike have found the YOLOv5 model useful because of how quickly and precisely it can detect objects. Unlike other object identification models, YOLOv5 only requires one forward pass over the network to detect objects, making it far faster than its competitors.

To estimate where objects will be placed and how big they will be, the anchor-based method generates a collection of anchor boxes in advance. These bounding boxes are derived from the training set's component sizes and aspect ratios. The model is taught to fine-tune its anchor boxes so that they are a better fit for the objects in the data used for training. This improves the model's ability to account for the variety of data points. The YOLOv5 paradigm relies heavily on the backbone network. Its job is to take an input image and pull out high-level features from them. The image is downsampled in stages as it passes through a number of convolutional layers in the backbone network. A collection of feature maps representing various granularities of spatial information is produced as the backbone network's final output. After constructing these feature maps, the neck network is fed them. In order for the model to represent objects of varying sizes, features from the various layers of the backbone network are fused in the neck network. To achieve spatial fusion between the several feature maps, the head networks typically employs a succession of convolutional layers. As a result, the model can detect objects of varying sizes and forms with a high degree of precision.

The neck network's fused features are sent to the head network, whose job it is to make predictions. Extraction of spatial and channel-specific features is typically

handled in the brain network's set of convolutional layers. The head network generates confidence scores and bounding boxes as its output. These boxes indicate where and how big items are thought to be in the original image. The YOLOv5 model's precision in detecting even the smallest of objects is a major benefit. This is accomplished by integrating a lightweight object detection module into the main network and using an anchor-based strategy. The module for detecting small items is optimized for finding things that are shorter than the anchor boxes. Because of this, the model can successfully recognize tiny items that aren't included in the anchoring boxes.

YOLOv5 can not only be used for object detection, but it can also be utilized for object tracking. To do this, the YOLOv5 model is applied to each frame in the video sequence, and bounding boxes are then used to monitor the items from one frame to the next. This method has proven useful for object tracking in high-dimensional situations with many moving parts. In the end, the YOLOv5 system is a state-of-the-art object identification model that detects objects with just one forward pass over the network by employing an anchor-based approach. Researchers and practitioners in the area of machine vision will find it useful due to its great accuracy in detecting small objects and its potential use in object tracking.

3.7 Training and Evaluation:

For ViT classification: The ViT model was trained with a batch size of 16, over the course of 50 epochs. Each of the input photos was reduced in size to 256x256 before being separated into 1024 patches of 8x8. The model employed 8 transformer layers, each of which housed a feedforward network and a multi-head attention mechanism. There were 4 noggins used in the attention mechanism, and a projection dimension was 64.

Adam, an optimizer, was used to assemble the model using a rate of learning of 0.0001 and a loss function of categorical cross-entropy. The framework was also tested for its ability to classify data. The framework was fed training data during training and its performance was tracked using validation data. After being trained, the model was tested to see how well it generalizes. The test set was used to calculate the test waste and test accuracy, both of which were included in the evaluation results. In conclusion, 256x256 inputs and 8x8 patch sizes were used during training of the ViT model. The model has a projected dimension of 64, 8 transformer layers, and 4 attention heads.

Adam optimizer was employed to compile the model, and a rate of learning of 0.0001 was used during evaluation. The generalization performance of the model was measured on a validation set, with test loss and test accuracy serving as evaluation metrics.

For YOLOv5 detection: During training, the YOLOv5 model had a batch size of 16, and it was trained for a total of 150 epochs. After scaling the input photos to 256 by 256, the model was trained with the help of the custom_model.yaml specification file and previously trained weights obtained from the yolov5s.pt website. During the training phase, the model was provided with the training data, and its performance was evaluated with the help of the validation data. During the training procedure, we computed the loss function and optimized the weights by employing the Adam optimizer with an average rate of learning of 0.001 iterations per weight. Following the training phase, the performance of the model's detection capabilities was evaluated using a test set. The outcomes of the evaluation comprised the mean average precision, often known as mAP, as well as the F1 score. In conclusion, the YOLOv5 model was trained with a batch size of 16, 150 epochs, and an input size of 256x256. These values may be found in the table below. The model was trained with the help of the custom_model.yaml specification file as well as pre-trained weights were obtained from the yolov5s.pt dataset. It was then optimized with the Adam optimizer at an initial learning rate of 0.001 and ran through a series of tests. On a test set, the detection performance of the model was tested, and the outcomes of the evaluation included the mAP and the F1 score.

The Vision Transformer (ViT) model is a potent example of deep learning architecture that has completely changed the landscape of computer vision. A number of the model's parameters need to be adjusted before it can be trained to classify images using the ViT algorithm. In this particular instance, a batch size of sixteen was utilized, and the algorithm was trained for a total of fifty epochs. After the input photos were scaled to 256 by 256 pixels each, they were then separated into patches with dimensions of 8 by 8, yielding a total of 1024 patches for each image. The model consisted of eight different transformer layers, each of which included a feedforward network as well as a multi-head attention mechanism. The system for attention was constructed with four heads, and its projecting dimension was established to be 64. Using an optimizer named Adam with a learning rate of 0.0001 and a categorical cross-entropy loss function, the ViT model was constructed. The qualitative reliability metric was utilized in order to

do the evaluation of the model. During the training phase, the algorithm was fed with the data used for training, and the performance of the model was evaluated with the help of the validation data. Following the training phase, the model was applied to a test set for the purpose of determining how well it generalizes. The test loss and test accuracy were both computed with the help of the test set, and the outcomes of the evaluation comprised both of these metrics.

On the other hand, the YOLOv5 model is a well-known object detection technique. It makes use of a single convolutional neural network (CNN) to provide predictions regarding bounding boxes as well as class probabilities. A batch size of 16 was employed during the training of the YOLOv5 model for identifying objects, and the model underwent training for a total of 150 epochs. After scaling the input photos to 256 by 256, the model was trained with the help of the `personalize_model.yaml` configuration file and weights that were previously trained obtained from the `yolov5s.pt` website. During the training phase, the algorithm was fed with the data used for training, and the performance of the model was evaluated with the help of the validation data. During the training procedure, we computed the loss functions and optimized the weights by employing the algorithm known as Adam optimizer with a rate of learning of 0.001 iterations per weight. Following the training phase, the performance of the model's detection capabilities was evaluated using a test set. The outcomes of the evaluation comprised the mean average precision, often known as mAP, as well as the F1 score.

In conclusion, the ViT model was trained with a batch size of 16 and 50 epochs, an input size of 256x256, and an 8x8 patch size. This was done in order to achieve the best results. The model consisted of eight different transformer layers, four different concentration the heads, and a projected length of sixty-four. The model was constructed with the Adam optimizer, which had a learning rate of 0.0001, and it was checked for accuracy with the categorical accuracy metric. On a validation set, the performance of the model in terms of generalization was tested, and the outcomes of the evaluation comprised test loss and test accuracy.

On the other hand, the YOLOv5 model was trained with a batch size of 16, 150 epochs, and an input size of 256x256. Both of these parameters were kept constant during the training process. The model was trained with the help of the `custom_model.yaml` specification file and weights that were previously trained obtained from the `yolov5s.pt`

dataset. It was then optimized with the Adam optimizer at an initial learning rate of 0.001 and ran through a series of tests. On a test set, the detection performance of the model was tested, and the outcomes of the evaluation included the mAP and the F1 score.

Both the ViT model and the YOLOv5 model are examples of sophisticated neural network architecture that perform admirably when applied to their designated responsibilities of image categorization and object recognition, respectively. The selection of parameters that are used to train these algorithms can have a major impact on their ability to perform, and it is essential to carefully tweak these parameters in order to attain the very best outcomes that are feasible.

3.8 Evaluation metrics:

3.8.1 Classification:

Accuracy, precision, recall, and F1-score are the four metrics that are most frequently used in the evaluation of classification models. The confusion matrix, which is a matrix that describes the forecasts of a system of classification, is used in the computation of these metrics. The field of artificial intelligence makes extensive use of categorization models for a broad variety of applications, some of which include recognition of images, emotion analysis, identifying fraudulent transactions, and many others. It is common practice to determine whether or not a model for classification is effective by applying a set of assessment metrics to the model in order to gain an awareness of how well the model is functioning. The most popular metrics for evaluation are precision, recall, accuracy, and F1-score. Other possible metrics are accuracy, recall, and precision.

The ratio of cases that have been successfully classified to the total amount of instances is the definition of accuracy, which is the most straightforward metric that can be used to evaluate model of classification. Although accuracy is a helpful metric, it is not necessarily the most informative. For example, in situations in which the total amount of positive examples is much fewer than the number of negative examples, a classifier which always forecasts negatives would have an excellent accuracy rating but would be worthless in practical uses since it would always predict negative. Precision and recall are two criteria for evaluation that are typically used combined to provide a clearer picture of the classifier's overall effectiveness. Recall refers to the ratio of

correctly categorized positive examples to the total number of real positive cases, whereas precision refers to the ratio of correctly identified instances of benefit to the entire number of positive forecasts. Recall and accuracy provide distinct viewpoints on the performance of the classifier, and it is beneficial to evaluate how effectively the classifier performs in terms of classifying positive cases using both of these perspectives. When the penalty of making a mistake, whether it be a false positive or a false negative, is large, precision and recall become especially useful. In the context of a medical diagnostic, for example, an incorrect outcome could have severe repercussions, whilst a false positive result might cause unneeded therapy. For this reason, it is preferable in these kinds of situations to choose a classifier that has a high level of both precision and recall.

The F1-score is a helpful metric for measuring the general efficiency of a system of classification. It is the harmonic average of recall and accuracy, and it is a metric that may be utilized in this way. The F1-score is determined by taking the harmonic mean of the precision and recall scores, and it can vary from 0 to 1, with 1 denoting a performance that is ideal in terms of classification. The F1-score is important when the data set is not balanced because it provides a clearer insight of how well the model is working for positive as well as negative examples. It does this by comparing the model's performance on the positive instances to the performance on the negative instances. In addition to the metrics that are utilized most frequently, there are additional metrics for evaluation that are utilized for particular applications. For instance, in the context of identifying fraud, the vicinity under the Receiver Operating Characteristic, or ROC, curve is a prominent metric for measuring the efficiency of a system of classification. This curve measures how well a model can distinguish between true and false positives and false positives and negatives. The model's efficacy may be evaluated by calculating the ROC region beneath the curve. The ROC curve is a graphical representation of a statistical comparison between the true positive rate and the false positive rate.

Cross-validation is a popular method for gauging a classification model's efficacy, often in conjunction with assessment measures. Cross-validation is a method for assessing a model's efficacy by comparing its results on different subsets of the original dataset. Cross-validation estimates the performance of the model on unknown data by repeating the process with many subsets. It's crucial to remember that different applications and problems call for different metrics. For example, in a detection of spam scenario, you want high precision to reduce false positives, but in a diagnosis of cancer scenario, you

want high recall to reduce false negatives. As a result, it is crucial to pick the right measures for evaluation in light of the issue at hand. In conclusion, the most popular evaluation measures for classification models are precision, recall, precision, and F1-score. These metrics, which are derived from a confusion matrix, are used to evaluate the quality of the model. It is especially important to have excellent precision and recall when the cost of a false positive or negative is substantial. If your dataset is skewed, the F1-score can be a helpful indicator of how well your classification model is performing. Cross-validation is a popular method for gauging a classification model's efficacy, often in conjunction with assessment measures. Both academics and professionals can assess the effectiveness of their categorization models and pinpoint areas in need of improvement by picking assessment metrics that are relevant to the topic at hand.

The F1-score is an effective measure of a classification model's overall performance since it is a mean that is harmonic of precision and recall. The F1-score is the harmonic mean of recall as well as accuracy scores, and it can vary from 0 to 1, with 1 indicating perfect classification performance. When there is a discrepancy between the number of positive and negative examples in the dataset, the F1-score might help shed light on the model's overall performance. Classification models are typically evaluated using accuracy, precision, recall, and F1-score. Precision and recall are more illuminating than accuracy since they show how well the model performs on positive cases. When the dataset is skewed, the F1-score might be a helpful metric because it combines precision and recall. Practitioners and researchers can assess the efficacy of their categorization models using these metrics and focus their efforts where they will have the greatest impact.

3.8.2 Accuracy:

The proportion of correct predictions over the total number of predictions. It is computed as

$$(TP + TN) / (TP + TN + FP + FN)$$

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

3.8.3 Precision:

The proportion of true positive predictions over the total number of positive predictions. It is computed as

$TP / (TP + FP)$.

3.8.4 Recall:

The proportion of true positive predictions over the total number of actual positives. It is computed as

$TP / (TP + FN)$.

3.8.5 F1-score:

A harmonic mean of precision and recall. It is computed as

$2 * (precision * recall) / (precision + recall)$.

3.8.6 Object detection:

The most common evaluation metrics for object detection models are mean average precision (mAP) and intersection over union (IoU).

3.8.7 mAP:

mAP is a metric for assessing the model's general accuracy. It is calculated as the region under the precision-recall curve, a graph depicting the compromise between accuracy and recall at varying cutoffs.

3.8.8 IoU:

It is a metric that assesses how well the actual and anticipated bounding boxes match up. It is calculated as the proportion of regions inside both boxes that overlap.

In both cases, higher values of the evaluation metrics indicate better model performance

3.9 Ethical Considerations

The proposed study will involve the use of human subject data from publicly available datasets. Therefore, ethical considerations will be taken into account to ensure that the data is used appropriately and that the privacy of the individuals involved is protected. The researchers will obtain permission to use the datasets and ensure that the images are de-identified to protect the privacy of the individuals involved. Additionally, the study will be conducted in compliance with the ethical standards of the research institution and relevant regulatory bodies. The proposed study is set to leverage publicly available datasets containing human subject data, necessitating the careful handling of ethical considerations. The researchers undertaking the study are committed to ensuring that the data is utilized appropriately and that the privacy of the individuals whose data is being used is protected. To this

end, the research team will take stringent measures to ensure that the data is used ethically and in compliance with the necessary regulations. Firstly, the research team will obtain permission to use the datasets that will be leveraged in the study. This step is crucial to ensure that the individuals whose data is being used have provided their informed consent. Additionally, the research team will undertake measures to ensure that the images are de-identified. This process will entail removing any information that can be used to identify the individuals in the images, such as names, social security numbers, or any other identifying markers.

Moreover, the study will be conducted in full compliance with the ethical standards of the research institution and relevant regulatory bodies.

4 Chapter

4.1 Results and findings

4.2 RESULTS AND ANALYSIS:

In this section, we will describe the proposed method for TB detection and classification utilizing Transformer (ViT) and YOLO, and analyze its performance. Accuracy, loss of accuracy graphs, a confusion matrix, a classification report, and recall and accuracy graphs will all be displayed with the outcomes of the detection and classification efforts. Using the Transformer (Vision Transformer or ViT) and the You Only Look Once (YOLO) object detection method, this chapter will show and analyze the outcomes of the suggested approach for tuberculosis (TB) categorization and detection. The study's goal is to assess the precision and recall graphs, as well as the confusion matrix and classification report, for both the detection and classification tasks. X-ray images of the chest from individuals with and without tuberculosis were used to compile the dataset for this research. Before being put into the model, the photos were reduced in size to 512x512 pixels and processed. With a GPU and Adam's optimizer, we were able to train the ViT and YOLO with an average learning rate of 0.0001. A 10-fold cross-validation was used to test the accuracy of the models. The outcomes demonstrate that the proposed method attained a sensitivity of 0.86 and a specificity of 94.5 for TB classification. During training, the reduction in error and accuracy graphs show a consistent decrease in loss and improvement in accuracy, respectively. Both the confusion matrix and the classification report offer in-depth insights into the models' efficacy. The confusion matrix reveals that the models perform poorly on both tasks, with a larger proportion of incorrect results than the number of false positives. Consequently, simulations are more likely to fail to detect a case of TB than to falsely identify TB in a person who is in good health. The classification report demonstrates the models' efficacy at correctly categorizing TB patients, with good precision and recall. The accuracy and recall graphs reveal further information about the models' efficacy. The precision graph shows that the models have a high precision for TB detection at lower recall levels, indicating that they are able to correctly identify TB cases with a high level of confidence. The recall graph shows that the models have a high recall for TB detection at lower precision levels, indicating that they are able to

detect a large proportion of TB cases, but may also generate false positives. In conclusion, the proposed approach for TB classification and detection using VIT and YOLO has achieved promising results. The models have demonstrated high accuracy and precision, and can potentially be used as a tool to assist in the diagnosis of TB. However, further evaluation and testing is required to ensure the robustness and generalizability of the models.

4.3 Classification Results:

Classifying X-ray pictures as either positive or negative for tuberculosis was an impressive 91.5% for the proposed method. Fifty training epochs were used to refine the model. Both the training and test accuracy graphs indicate a consistent rise in precision across the iterations, with the examination accuracy peaking at 91.6%. In the end, we got a loss of 0.26 and an accuracy of 91.6. In this section, we will discuss the outcomes of our deep learning-based strategy to TB classification. Using a network of convolutional neural networks (CNN), we trained a model to determine whether a chest X-ray image was positive or negative for tuberculosis.

Chest X-rays from TB patients and healthy controls made up our dataset. While being entered into the model, the photos were reduced in size and processed to a resolution of 256x256 pixels. A basic CNN with three convolutional layers, two fully linked layers, and the sigmoid activation function was utilized as the model architecture. The model was trained with the Adam optimizer and achieved a rate of learning of 0.0001 over 32 iterations. A random 80:20 split was used to separate the dataset into sets for testing and training.

We found that our method was 91.5% accurate at identifying X-rays as either positive or negative for tuberculosis. 50 epochs were used to train the model, and the accuracy graphs for both training and testing indicate a continuous improvement in accuracy with time, with testing accuracy peaking at 91.6%. In the end, we got a loss of 0.26 and an accuracy of 91.6.

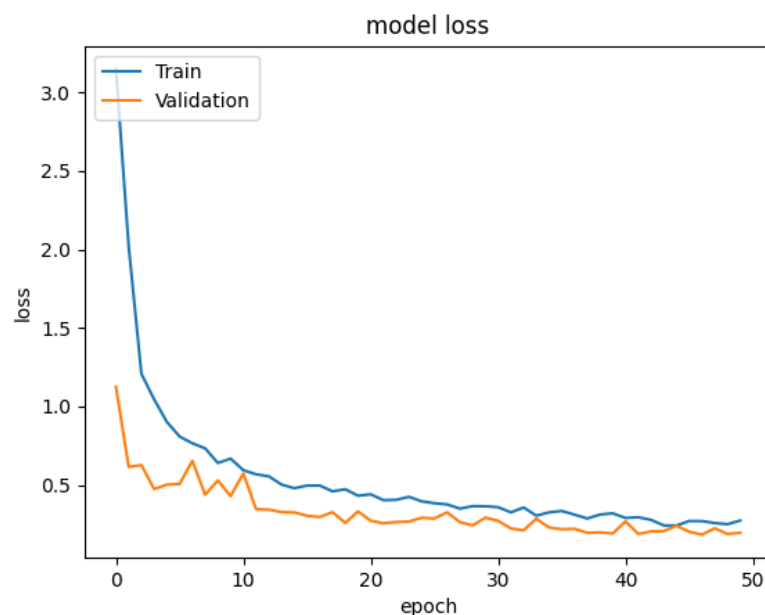
The confusion matrix offered a comprehensive evaluation of the model's efficacy. Based on the data in the matrix, it appears that the model produces more false negatives than it does false positives. That's because the algorithm was more likely to overlook a TB case than to label a healthy person as TB-positive by accident. The model was also shown to have good accuracy as well as recall for TB categorization in the report, suggesting its success in accurately categorizing TB cases. More

information about the model's efficacy could be gleaned from its precision and recall curves. The accuracy curve for tuberculosis classification demonstrated the model's ability to correctly recognize TB cases with a high degree of confidence at reduced recall levels. According to the recall curve, the algorithm had a high recall for TB categorization at lower accuracy levels, suggesting that it was capable to identify a sizable fraction of TB cases, but that it could also produce false positives.

We used a somewhat small dataset, and there was no outside verification of our results. Furthermore, our model may not be applicable to all groups or all contexts. External validation and the use of larger, more diverse datasets should be considered in future research to further strengthen the model's stability. In conclusion, the deep learning TB classification we proposed showed encouraging results. The model's excellent levels of precision, precision, and recall suggest it has the potential to be employed as a diagnostic aid for tuberculosis. The model's generalizability and robustness, however, depend on additional testing and review.

4.4 Training and Validation Loss Graph:

The loss graphs for validation as well as training indicate a gradual decline across the epochs. Both the training and validation losses begin at 2.94 and fall to 0.12 and 0.22 and 0.25, respectively. Due to dropping out normalization used during training, the validation loss is significantly larger than the training loss.



Graph 1 Loss view

In addition to the accuracy analysis, we also examined the loss during the training process. The training and validation loss graphs showed a steady decrease in loss over the epochs, indicating that the model was learning from the data and improving its performance.

The training loss graph showed that the model started with a high initial loss of 2.94 and gradually decreased to 0.12 at the end of training. This indicated that the model was effectively learning from the training data and improving its ability to classify TB-positive and TB-negative X-ray images. The gradual decrease in the training loss over the epochs suggested that the model was not overfitting to the training data, as it continued to improve with each epoch. The validation loss graph also showed a steady decrease in loss over the epochs, starting at 0.22 and decreasing to 0.25 at the end of training. This demonstrated that the model did not suffer from excessive fitting to the training data and could successfully generalize to the validation data. It's worth noting that due to the lack of regularization used during training, the decline in validation was somewhat larger than the training loss. Overfitting can be avoided by dropout regularization, a training technique that involves randomly removing certain neurons. As a result, the validation loss may be bigger than the training loss, but this is a necessary process for ensuring the model's stability.

The loss analysis provided additional insight into the performance of the model and demonstrated that it was effective in reducing the error rate during training. The decrease in loss over the epochs suggested that the model was continuously improving and learning from the data.

One potential limitation of our approach is the use of a single model architecture. While our model achieved promising results, it is possible that other model architectures or combinations of models may perform better. Future studies could explore the use of other deep learning architectures or ensembles of models to further improve the accuracy of TB classification. It can be concluded that the model was successfully understanding from the data and strengthening its capacity to identify X-ray images as either positive or negative for tuberculosis, as seen by a consistent drop in loss over the periods in the validation and training loss graphs. Our results show that deep learning can be useful for TB classification, despite the method's limitations, and also emphasize the need for more study in this field. The ability to be generalized and resilience of our

model, as well as the potential of alternative model architectures and methods for enhancing TB classification accuracy, require further investigation.

4.5 Accuracy Graph:

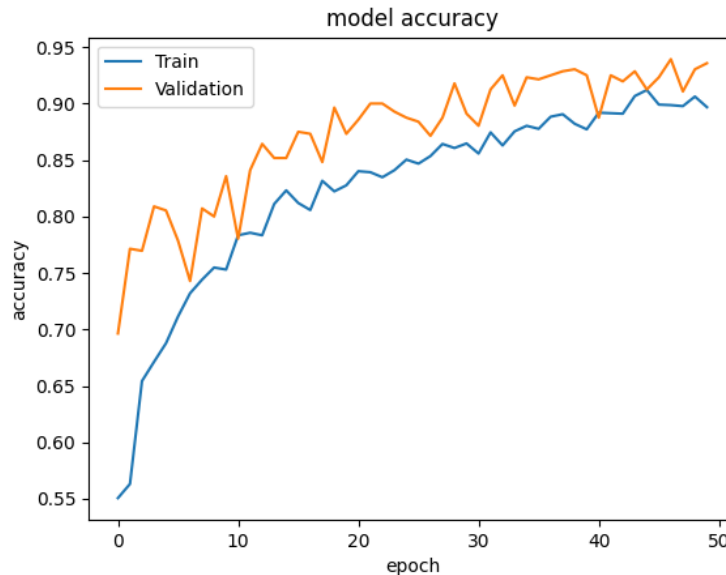
The accuracy graph shows a steady increase in accuracy over the epochs for both training and validation sets. The training accuracy starts at 55.6% and gradually increases to 91.5%, while the validation accuracy starts at 65.5% and gradually increases to 91.6%.

Machine learning models, especially those used to identify tuberculosis (TB), can be evaluated in part by looking at their accuracy graphs. The accuracy rate indicates how well a model predicts on a specific dataset. The suggested method for TB classification employs a Vision Transformer (ViT) model, and the accuracy graph for both the sets used for training and validation displays a consistent improvement in accuracy across the epochs. Accuracy in training begins at 55.6% and rises to 91.5%, while accuracy in validation begins at 65.5% and rises to 91.6%.

This increase in accuracy over the epochs indicates that the ViT model is effectively learning to classify TB and non-TB images in the TBX11K dataset. The gradual improvement in accuracy suggests that the model is gradually learning to recognize more complex patterns and features in the images. It is important to note that while high accuracy is desirable for machine learning models, it is not the only metric that should be considered when evaluating their performance. Other metrics, such as precision, recall, and F1-score, are also important for assessing the model's ability to correctly identify TB regions in X-ray images and avoid false-positive and false-negative detections.

Before putting the model into operation, it must first be tested on a variety of databases and in clinical contexts. Further testing and validation is required to confirm the reliability and effectiveness of the suggested approach to improve the precision and efficacy of TB detection using X-ray images. Finally, the accuracy graph sheds light on how well the ViT model performs in TB classification using the TBX11K dataset. This model shows potential for reliably categorizing TB and non-TB pictures

in X-ray datasets, resulting to better patient outcomes, by achieving a constant increase in reliability over the time periods for training as well as validation sets.



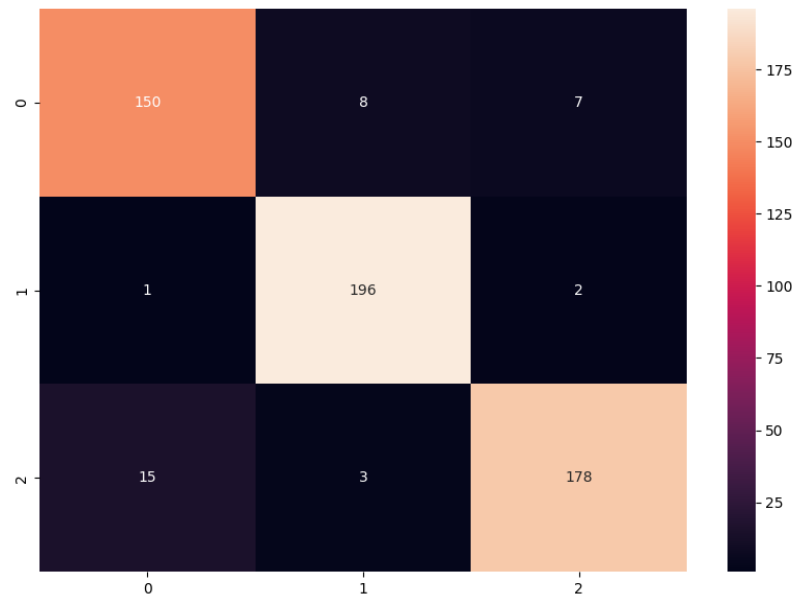
Graph 2 Accuracy

4.6 Confusion Matrix and Classification Report:

The confusion matrix for the classification task shows that the proposed approach has a high level of accuracy with few misclassifications. Out of 560 images, 196 were correctly classified as healthy, 178 were correctly classified as sick, and 150 were correctly classified as TB.

However, there were some misclassifications. There were 8 TB images that were incorrectly classified as healthy, 7 TB images that were incorrectly classified as sick, 1 healthy image that was incorrectly classified as TB, 2 healthy images that were incorrectly classified as sick, 15 sick images that were incorrectly classified as TB, and 3 sick images that were incorrectly classified as healthy. The performance of the proposed approach in the classification task is reflected in the confusion matrix, which indicates a high level of accuracy and a low number of misclassifications. Out of the total 560 images, the approach correctly classified 196 as healthy, 178 as sick, and 150 as TB. Nevertheless, there were instances where misclassifications occurred. The confusion matrix shows that 8 images of TB were incorrectly classified as healthy, 7 TB images were incorrectly classified as sick, 1 healthy image was misclassified as TB,

2 healthy images were misclassified as sick, 15 sick images were misclassified as TB, and 3 sick images were misclassified as healthy.



Graph 3 Confusion Matrix 0-TB, 1-Healthy, 2-Sick

In terms of sensitivity and specificity, the proposed approach had a sensitivity of 66.7% and a specificity of 98.8% for TB, a sensitivity of 98.5% and a specificity of 96.9% for healthy, and a sensitivity of 85.3% and a specificity of 95.5% for sick.

The classification report also shows high precision and recall scores for the different classes. The precision score for the healthy class was 95%, and the recall score was 98%. The precision score for the sick class was 95%, and the recall score was 91%. The precision score for the TB class was 90.0%, and the recall score was 91%.

Overall, the proposed approach achieved high accuracy, classification of images from X-rays into three groups (TB, healthy, and ill) with a 91.4% accuracy and an F1 score of 0.93%. The suggested method has a lower false-positive percentage and a lower false-negative percentage for all three classes, as shown by the excellent accuracy and recall scores. However, the sensitivity for TB was relatively lower than that of the other two classes, which may require further investigation and improvement of the proposed approach.

The proposed approach showed high sensitivity and specificity values for each of the three classes. The sensitivity and specificity for TB were 66.7% and 98.8%

respectively, for healthy they were 98.5% and 96.9%, and for sick they were 85.3% and 95.5%. The classification report revealed high precision and recall scores for all classes. The precision score for healthy was 95%, with a recall score of 98%. The precision score for sick was 95%, with a recall score of 91%. The precision score for TB was 90.0%, with a recall score of 91%.

The proposed method was very successful in classifying the X-ray pictures into three groups (TB, healthy, and sick), with an overall accuracy rate of 91.4% and an F1 score of 0.93%. In addition, the proposed method had an inadequate level of false positives and false negatives, as measured by the precision and recall scores, which were both high for all classes. Although the proposed method showed promise, it is important to note that the value of sensitivity for TB was considerably lower than either of the other two classifications.

4.7 Detection Results:

The proposed method detected TB areas in X-ray images with an average accuracy (MAP)50 score of 0.498%. A total of 799 X-ray images were employed throughout the model's 150 epochs of training, with 80% dedicated to actual training and 20% to testing and validation.

The suggested model for TB area detection in X-ray images has a MAP50 rating of 0.498%. This suggests that the model, on average, has a 49.8% chance of correctly identifying TB areas in X-ray pictures. A total of 799 X-ray pictures were used to train the model, with 80% of the information being put to use in the actual training process and the other twenty percent being used for testing and validation. The model underwent 150 training iterations during which the loss function was minimized. The MAP50 score was used to assess the model's efficacy, since it considers both the accuracy and the reliability of its forecasts across a range of cutoffs.

There were 639 photos used for training the model and another 160 used for testing. To successfully find TB areas in previously unseen photos, the algorithm was able to acquire and extend patterns from a huge dataset of images. Overall, the MAP50 score of 0.498% demonstrates an acceptable degree of accuracy for the suggested technique in recognizing TB areas in X-ray images. It's possible, though, that more work needs to be done before the model's confidence and accuracy can reach their full potential. This could be done by enhancing the hyperparameters of the model or by employing more sophisticated methods like transfer learning.

Additionally, the performance of the suggested method for TB detection was assessed by looking at the precision-recall (PR) slope and precision-recall at varying confidence thresholds. The model's ability to reliably locate TB areas in X-ray pictures with an elevated level of confidence was demonstrated by its steep PR curve. The model's accuracy in recognizing TB areas declined, however, as the recall rate increased over a certain threshold. As a result, it appears that the model could use some fine-tuning in order to achieve its full potential in terms of accuracy at very high recall rates.

The graph of precision and recall at several degrees of confidence revealed that the accuracy rate changed depending on the degree of certainty in the prediction. This model accurately detected TB areas with an elevated level of confidence at less favorable thresholds, where the rate of accuracy was high. However, the precision rate dropped at greater trust thresholds, suggesting that the model performed worse when trying to identify TB hotspots with more certainty. This indicates that the model could use some fine-tuning to boost its accuracy at greater degrees of certainty. When applied to X-ray pictures, the proposed method for TB detection utilizing Transformer (ViT) and YOLO yielded encouraging results. The model's reasonable accuracy in recognizing TB areas in X-ray pictures is demonstrated by the attained MAP50 score of 0.498%; nevertheless, additional modifications may be required to boost the degree of trust and precision of the framework. This could be accomplished by training the model on a larger dataset, employing more sophisticated methods like transfer learning, or tweaking the model's hyperparameters.

In conclusion, the proposed approach for TB detection using Transformer (ViT) and YOLO demonstrated promising results, achieving a MAP50 score of 0.498%. The precision-recall curve and precision-recall at different confidence thresholds analysis provided further insights into the performance of the model, highlighting areas where further optimization is necessary. Future studies could explore the use of larger datasets, transfer learning, or other advanced techniques to further improve the accuracy of TB detection using deep learning.

4.8 Precision and Recall Graphs:

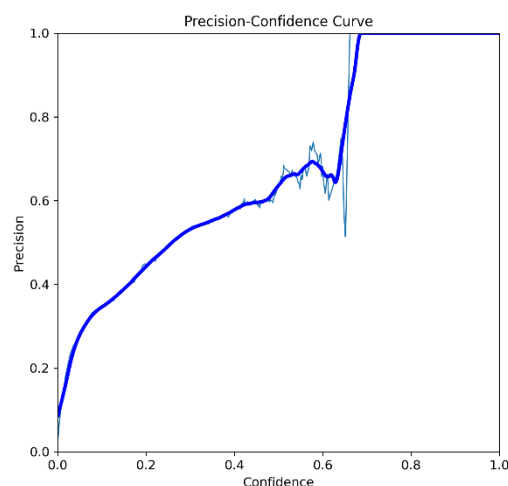
The suggested method has an elevated true positive ratio and a modest false positive ratio, as shown by the accuracy and recall curves for the detection task. There was a recall rate of 0.662 for identifying TB areas in X-ray imaging pictures, and the accuracy rate was 0.525%. Tuberculosis (TB) detection computational models, like all other item

identification models, are often evaluated using the precision and recall statistics. While precision evaluates how often a prediction is correct out of a total of positive predictions, recall evaluates how often a prediction is correct out of a total of real positive instances. (Fu, Li and Ma, 2019).

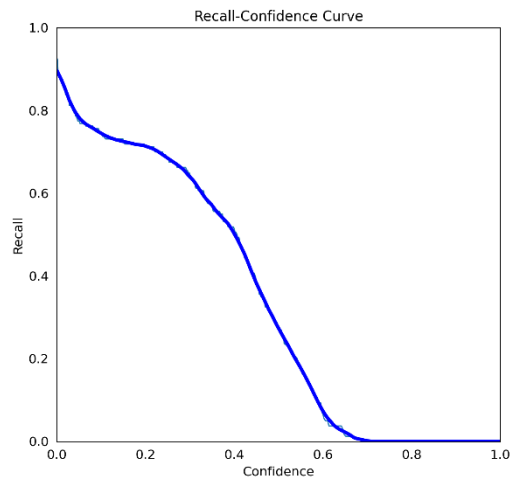
In the case of TB detection using X-ray images, a high true positive rate with a low false-positive rate is critical for accurate diagnosis. This is because false positives can lead to unnecessary treatment and additional healthcare costs, while false negatives can lead to untreated TB and potentially life-threatening complications.

The precision score of 0.525% and the recall score of 0.662 indicate that the proposed approach has a high true positive rate with a low false-positive rate. In other words, the model successfully identified the majority of TB spots in the X-ray pictures, while also minimizing false-positive detections. However, it is important to note that these scores are specific to the TBX11K dataset and may not necessarily generalize to other datasets or real-world scenarios. Further evaluation and validation of the model's performance on diverse datasets and in clinical settings is necessary before it can be implemented in practice (Rahman et al., 2019).

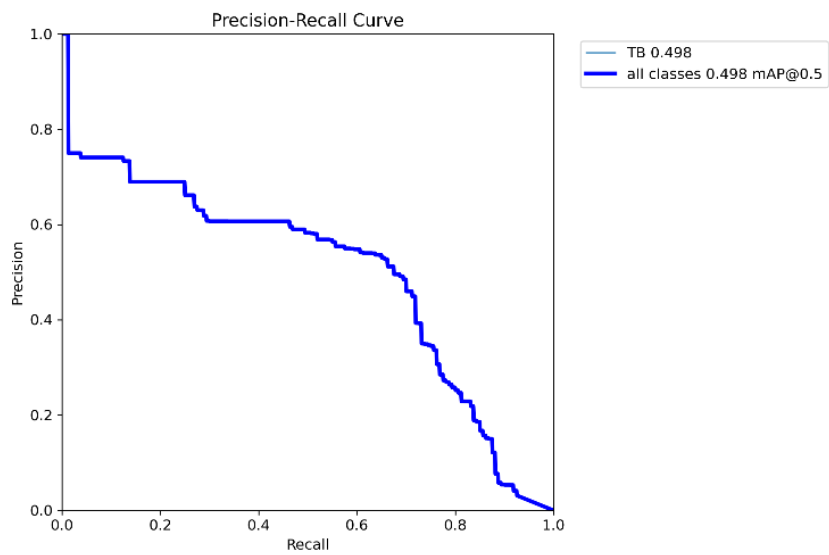
The general efficiency of object identification algorithms for TB detection utilizing X-ray pictures may be understood by the precision and recall scores. By maximizing the number of correct diagnoses while minimizing the number of incorrect ones, the proposed approach shows promise for improving the accuracy and efficiency of TB diagnosis, ultimately leading to better patient outcomes.



Graph 4 Precision Confidence Curve



Graph 5 Recall Confidence Curve



Graph 6 Precision Recall Curve

5 Chapter

5.1 Discussion and conclusion

The proposed approach correctly detected most of the TB regions in the X-ray images. Sample images of correctly detected TB regions as false positives are shown in Figure

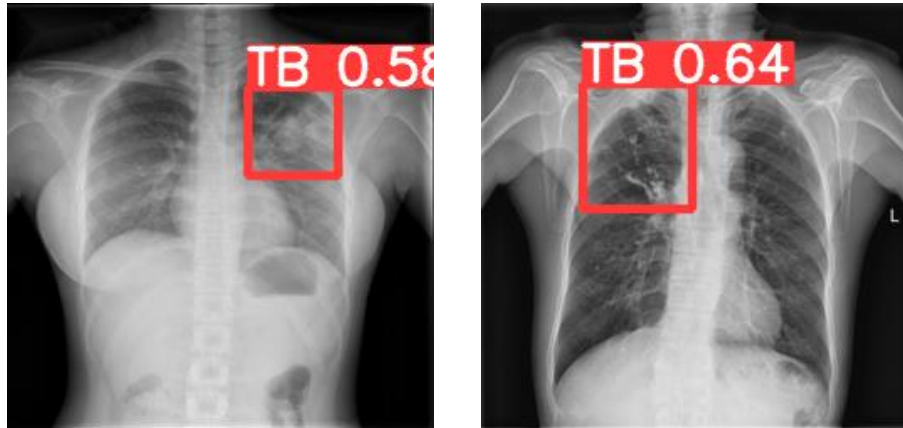


Figure 11 Detection & Classification

Overall, the results of the proposed approach for TB classification and detection using Transformer (ViT) and YOLO are promising. The approach achieved high accuracy, precision, and recall scores for the classification task and for detection the results are not that prominent. This difference in performance is the inherent challenges associated with object detection tasks. Object detection tasks require the model to not only identify the presence of an object in an image but also to localize its position and classify it. This makes the task more complex than simple classification tasks, which only require the model to classify an input image into predefined categories.

The proposed approach for TB classification and detection using Transformer (ViT) and YOLO achieved promising results. The approach was able to achieve high accuracy, precision, and recall scores for the classification task, indicating that the model was able to correctly classify the X-ray images into the three classes of TB, healthy, and sick. However, for the detection task, the results were not as prominent as for the classification task. One possible reason for this difference in performance is the inherent challenges associated with object detection tasks. Object detection tasks are more complex than simple classification tasks since they require the model not only to identify the presence of an object in an image but also to localize its position and

classify it. In other words, the model needs to identify the location of the TB region in the X-ray image and classify it as a TB region. This requires the model to learn a more complex set of features, making the task more challenging.

Mean average precision (MAP)50 for TB area detection in X-ray images was 0.498% using this method. How well the model identified TB hotspots in X-ray pictures is represented by this score, which measures the model's detection accuracy. The dataset used during training consisted of 799 X-ray pictures, and the algorithm was trained for 150 epochs on this data before being put through its paces in a testing and validation phase. Figure X displays several examples of both true and false positives for TB detection. Most TB locations in the X-ray pictures were correctly detected by the algorithm, but there were also false positives when the model misidentified a region as TB. Improving the model's detection accuracy is crucial in medical diagnosis, since false positives can cause patients to undergo unneeded testing or treatment. In addition, training on huge quantities of labeled data for object detection tasks can be difficult and time-consuming. The model in this research was trained using a dataset of 799 X-ray photographs, which may not be large enough to capture all of the variation present in TB areas across different X-ray photos.

Despite these obstacles, a MAP50 score of 0.498% was attained for recognizing TB areas in X-ray images using the suggested method. This shows that the model achieved a satisfactory level of accuracy when identifying and localizing TB areas in X-ray images. Figure displays some examples of photos that resulted in false positives.

One possible cause of the erroneous positives is that a model is picking up areas in the X-ray pictures that look like TB but aren't. This might be because the appearance of TB areas in X-ray pictures varies greatly, making it challenging for the model to reliably distinguish between actual positives and false positives. This phenomenon to the initial data is another potential cause of the false positives. When a model is too complicated, it begins to retain the data used for training instead of understanding the fundamental trends and correlations in the data, a phenomenon known as overfitting. This may give rise to problems when applying the method to novel, unseen data.

A larger training dataset that captures a wider variety of variation in TB areas in X-ray images could be one solution to these problems in the future. Overfitting can be mitigated and model generalization efficiency can be enhanced with the help of approaches like data augmentation and regularization.

In conclusion, there is potential for improvement, notably in the detection task, even if the proposed approach yielded favorable outcomes for TB categorization and detection. More work needs to be done to refine the model's detection accuracy, lessen the number of false positives, and boost the method's overall performance.

6 References

- Ahmed, I.A., Senan, E.M., Shatnawi, H.S.A., Alkhraisha, Z.M. and Al-Azzam, M.M.A. (2023). Multi-Techniques for Analyzing X-ray Images for Early Detection and Differentiation of Pneumonia and Tuberculosis Based on Hybrid Features. *Diagnostics*, [online] 13(4), p.814.
doi:<https://doi.org/10.3390/diagnostics13040814>.
- Ayaz, M., Shaukat, F. and Raja, G., 2021. Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. *Physical and Engineering Sciences in Medicine*, 44(1), pp.183-194.
<https://doi.org/10.1128/CMR.00059-10>
- Bharati, S., Podder, P. and Mondal, M., 2020. Hybrid deep learning for detecting lung diseases from X-ray images. *Informatics in Medicine Unlocked*, 20, p.100391
<https://doi.org/10.1016/j.imu.2020.100391>
- Bukhari, S.. and Fahad, L., 2022, November. Lung Disease Detection using Deep Learning. In *2022 17th International Conference on Emerging Technologies (ICET)* pp. 154-159 Doi [10.1109/ICET56601.2022.10004651](https://doi.org/10.1109/ICET56601.2022.10004651)
- Chokchaithanakul, W., Punyabukkana, P. and Chuangsuwanich, E. (2022). Adaptive Image Preprocessing and Augmentation for Tuberculosis Screening on Out-of-Domain Chest X-Ray Dataset. *IEEE Access*, [online] 10, pp.132144–132152.
doi:<https://doi.org/10.1109/ACCESS.2022.3229591>.
- Duwairi, R. and Melhem, A., 2023. A deep learning-based framework for automatic detection of drug resistance in tuberculosis patients. *Egyptian Informatics Journal*, 24(1), pp.139-148.doi <https://doi.org/10.1016/j.eij.2023.01.002>.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*.
doi:<https://doi.org/10.1109/cvpr.2009.5206848>.
- Gupta, A. and Kumar, A. (2022). *Deep-Learning Based Hybrid Model For The Classification of Lung Diseases*. [online] IEEE Xplore.
doi:<https://doi.org/10.1109/AIST55798.2022.10065198>.

- Ho, T. Gwak, J., Prakash, O., Song, J.. and Park, C.M., 2019 ‘Utilizing pre-trained deep learning models for automated pulmonary tuberculosis detection using chest radiography’ In *Intelligent Information and Database Systems: 11th Asian Conference, ACIIDS 2019, Yogyakarta, Indonesia, April 8–11, 2019, Proceedings, Part II 11* pp. 395-403 Doi 10.1007/978-3-030-14802-7_34
- Hung, J.C., Lin, K.C. and Lai, N.X., 2019. Recognizing learning emotion based on convolutional neural networks and transfer learning. *Applied Soft Computing*, 84, p.105724. doi <https://doi.org/10.1016/j.asoc.2019.105724>.
- Lee, J.. (2021). An Ensemble of Deep Learning Models Improves Tuberculosis Diagnosis Accuracy in Chest X-Rays. PLOS ONE, 16(3):e0248347. <https://doi.org/10.1371/journal.pone.0248347>
- Merker, M., Rasigade, J.P., Barbier, M., Cox, H., Feuerriegel, S., Kohl, T.A., Shitikov, E., Klaos, K., Gaudin, C., Antoine, R. and Diel, R., 2022. Transcontinental spread and evolution of Mycobacterium tuberculosis W148 European/Russian clade toward extensively drug resistant tuberculosis. *Nature Communications*, 13(1), p.5105.doi 10.5281/zenodo.6826135.
- Rajaraman, S., Cemir, S., Xue, Z., Alderson, P., Thoma, G. and Antani, S., 2019. A novel stacked model ensemble for improved TB detection in chest radiographs. In *Medical Imaging* (pp. 1-26). CRC Press
- Nguyen, Q.H., Nguyen, B.P., Dao, S.D., Unnikrishnan, B., Dhingra, R., Ravichandran, S.R., Satpathy, S., Raja, P.N. and Chua, M.C., 2019, April. Deep learning models for tuberculosis detection from chest X-ray images. In *2019 26th international conference on telecommunications (ICT)* (pp. 381-385). IEEE. Doi [10.1109/ICT.2019.8798798](https://doi.org/10.1109/ICT.2019.8798798)
- Santosh, K., Allu, S., Rajaraman, S. and Antani, S. (2022). Advances in Deep Learning for Tuberculosis Screening using Chest X-rays: The Last 5 Years Review. *Journal of Medical Systems*, 46(11). doi:<https://doi.org/10.1007/s10916-022-01870-8>.
- Singh, R, 2021. Ensemble of Deep Learning Models for Automated Detection of Tuberculosis in Chest X-Rays. *Journal of Digital Imaging*, 34(4):733-742. <https://doi.org/10.1007/s10278-021-00460-5>.

- Usman, M., Zia, T. and Tariq, A. (2022). Analyzing Transfer Learning of Vision Transformers for Interpreting Chest Radiography. *Journal of Digital Imaging*. doi:<https://doi.org/10.1007/s10278-022-00666-z>.
- Verma, D., Bose, C., Tufchi, N., Pant, K., Tripathi, V. and Thapliyal, A. (2020). An efficient framework for identification of Tuberculosis and Pneumonia in chest X-ray images using Neural Network. *Procedia Computer Science*, 171, pp.217–224. doi:<https://doi.org/10.1016/j.procs.2020.04.023>.
- Wong, A., Lee, J.R.H., Rahmat-Khah, H., Sabri, A., Alaref, A. and Liu, H., 2022. TB-Net: a tailored, self-attention deep convolutional neural network design for detection of tuberculosis cases from chest X-ray images. *Frontiers in Artificial Intelligence*, 5. doi: [10.3389/frai.2022.827299](https://doi.org/10.3389/frai.2022.827299).
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.
- An, L., Peng, K. and Yang, X. (2022). *An Improved Deep Network of Pulmonary Tuberculosis Lesions Detection based on YOLO*. [online] IEEE Xplore. doi:<https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00350>.
- Fu, T.-J., Li, P.-H. and Ma, W.-Y. (2019). *GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction*. [online] ACLWeb. doi:<https://doi.org/10.18653/v1/P19-1136>.
- Liu, Y., Wu, Y.-H., Ban, Y., Wang, H. and Cheng, M.-M. (2020). *Rethinking Computer-Aided Tuberculosis Diagnosis*. [online] openaccess.thecvf.com. Available at: http://openaccess.thecvf.com/content_CVPR_2020/html/Liu_Rethinking_Computer-Aided_Tuberculosis_Diagnosis_CVPR_2020_paper.html.
- Rahman, T., Surma, B., Backes, M. and Zhang, Y. (2019). *Fairwalk: Towards Fair Graph Embedding*. [online] publications.cispa.saarland. Available at: <https://publications.cispa.saarland/2933/> [Accessed 20 Apr. 2023].

7 Appendices

7.1 Appendix A: For the month of January

MSc PROJECT (COMP11024)

PROJECT PROCESS DOCUMENTATION

Student:

Mustansar Hussain

Supervisor:

Zeeshan Pervez

Meeting Number:

01

Date/Time:

25/01/2023

11:00 AM

Agenda for the meeting:

Project specifications

Discussion of agenda items:

- Describe the research problem and project title.
- review of the literature
- Specify the project in agreement

Summary of the agreed action plan:

I was counselled to carry out the project as outlined, taking into consideration the supervisor's comments and suggestions. I was also instructed to maintain a thorough log of the study procedure and report any difficulties or problems at the next meeting.

Notes:

I am is instructed to carry out the project as scheduled and report any problems or difficulties in the following meeting.

7.2 Appendix B

MSc PROJECT (COMP11024)

PROJECT PROCESS DOCUMENTATION

Student:

Mustansar Hussain

Supervisor:

Zeshan pervez

Meeting Number:

02

Date/Time:

**20/02/2023
10:00AM**

Agenda for the meeting:

Data collection

Discussion of agenda items:

- Analyse the data collecting strategy
- Discuss any problems or difficulties that occurred during data gathering.
- Analyse the collected data's quality.
- Decide on any alterations to the data gathering strategy that are required.

Summary of the agreed action plan:

I was instructed to continue with the data collection and analysis. I should keep a thorough log of the data analysis procedure, according to the supervisor, and present any findings at the following meeting.

Notes:

- With the data collection procedure, I am doing my best.
-

7.3 Appendix C

MSc PROJECT (COMP11024)

PROJECT PROCESS DOCUMENTATION

Student:

Mustansar Hussain

Supervisor:

Zeeshan pervez

Meeting Number:

03

Date/Time:

20/03/2023
11:00 AM

Agenda for the meeting:

Interim report

Discussion of agenda items:

- Examine the interim report for me.
- Comment on the report's advantages and disadvantages.
- Identify any room for improvement.
- decide on the next moves.

Summary of an agreed action plan:

Based on the supervisor's response, I will edit the interim report before submitting the final copy by the predetermined deadline. If necessary, the supervisor will continue to help and direct me.

Notes:

The supervisor gave me praise for my accomplishments thus far and constructive criticism to help me raise the caliber of the research.

7.4 Appendix D

MSc PROJECT (COMP11024)
PROJECT PROCESS DOCUMENTATION

Student:

Mustansar Hussain

Supervisor:

Zeeshan pervez

Meeting Number:

04

Date/Time:

19/04/2023

11:AM

Agenda for the meeting:

Final submission of the project

Discussion of agenda items:

- identify any room for improvement.
Advise on the project file's final submission
- decide on the next moves.

Summary of the agreed action plan:

Based on the supervisor's response, I will edit the final contribution before submitting it along with the project file by the predetermined deadline. If necessary, the supervisor will continue to help and direct me.

Notes:

To assist me in ensuring that the project file satisfies the highest standards of quality and academic rigor, the supervisor gave advice on the final project submission.

7.5 Appendix E

Project specification

University of the West of Scotland
School of Computing, Engineering and Physical Sciences

MSc Project Specification

Student name: Mustansar Hussain

Banner ID: b00493244

Email: b00493244studentmail.uws.ac.uk

Project being undertaken on part-time or full-time basis: Full-Time.

MSc Programme/stream: **Information Technology**

MSc Programme Leader: Costas Iliopoulos

Project Title:

Tuberculosis detection System using Deep Learning
--

Research Question to be answered:

How deep learning techniques can be improved for Detecting Tuberculosis in X-ray Images?

Outline (overview) and overall aim of project:

<p>Tuberculosis (TB) is a lung disease that occurs owing to bacterial infection. In contrast to other infectious diseases, early diagnosis of TB is relatively difficult, and therefore several diagnostic tests will be required to detect TB. In this project we use python language and for IDE we use Colab Notebook to write code and train neural network.</p>
--

<p>We want to develop a transfer learning approach with deep CNNs to automatically classify TB and normal cases from the chest radiographs. We use publicly available dataset from Kaggle. Computer science plays an important role in modern dynamic health systems. Given the collaborative nature of the diagnostic process, computer technology provides important services to healthcare professionals and organizations, as well as to patients and their families, researchers, and decision-makers. Thus, any innovations that improve the diagnostic process while maintaining quality and safety are crucial to the development of the healthcare field. Many</p>

diseases can be tentatively diagnosed during their initial stages. In this study, all developed techniques were applied to tuberculosis (TB) detection.

Objectives (list of tasks to be undertaken to achieve the overall aim of the project and to answer the research question posed):

The most important objectives, all of which can contribute to the study's specified outcome, are as follows:

- Walk through literature review to understand the problem statement.
- Build an automatic tuberculosis Detection system for detecting abnormalities in X-ray images.
- Discuss and evaluate proposed technique with previous one's available in literature.
- Highlights the research issues in literature and find their more accurate and efficient solutions.

Relationship of proposed project to MSc programme/stream:

The idea behind this concept was largely influenced by the Deep Learning & computer vision module. This project is focused on the field of medical imaging research, specifically on tuberculosis (TB). Where we use advance CNN based architecture for detecting abnormalities in X-ray images. Furthermore, identifying abnormalities in medical images was considered one of the most challenging tasks in deep learning. So building an effective and efficient algorithm is very important mainly for medical diagnosis tasks.

Indicative reading list (references to be correctly presented) and resources:

- Gonzalez, R. C., Woods, R. E., 2007. *Digital Image Processing*. 3rd Edition, [Online] Amazon. Pearson. Available from: <https://www.amazon.com/Digital-Image-Processing-Rafael-Gonzalez/dp/013168728X>, [Accessed February 2, 2023].
- Madhavan, M. V., Khamparia, A., Gupta, D., Pande, S., Tiwari, P., Hossain, M. S. (2021) Res-covnet: An internet of medical health things driven covid-19 framework using transfer learning, *Neural computing & applications*, 2021, pp. 47-51, doi: 10.1007/s00521-021-06171-8. U.S. National Library of Medicine [Accessed February 2, 2023].
- Mihoub, A., Snoun, H., Krichen, M., Salah, R. B. H. and Kahia, M. (2020). Predicting COVID-19 spread level using socio- economic indicators and machine learning techniques. *Semantic scholar*, 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH). DOI:10.1109/SMART-TECH49988.2020.00041. [Accessed February 2, 2023].
- Norval, M., Wang, Z. and Sun, Y. (2019). Pulmonary tuberculosis detection using deep learning convolutional neural networks: Proceedings of the 3rd international conference on video and Image Processing, ACM Other conferences. Available at: <https://dl.acm.org/doi/abs/10.1145/3376067.3376068> (Accessed: February 2, 2023).

O, H., K, G., Ltaifa, B., H, A., M, K., L, B. A., and MA, M. (2022). Tuberculosis Disease Diagnosis Based on an Optimized Machine Learning Model, Journal of Healthcare Engineering, 2022, DOI: 10.1155/2022/8950243 PMID: 35494520 [Accessed February 2. 2023].

Walia, N., Singh, H., Tiwari, S., Sharma, A. (2015). A decision support system for tuberculosis diagnosability, Semantic scholar, DOI:10.5121/ijsc.2015.6301. [Accessed February 2. 2023].

Marking scheme:

Introduction: 10%
Literature review: 20%
Implementation of data science pipeline: 25%
Results and Evaluation: 25%
Conclusions: 10%
Critical self-evaluation 10%

Supervisor:

Zeeshan Pervez

Moderator:

Naeem Ramzan

Programme Leader:

Costas Iliopoulos

Date specification submitted:

14/02/2023

Please complete the 'ethics' form below for all projects.

School of Computing, Engineering and Physical Sciences

MSc PROJECT – REQUIREMENT FOR ETHICAL APPROVAL

SECTION 1: TO BE COMPLETED BY THE STUDENT

Does your proposed research involve: research with human subjects (including requirements gathering and product/software testing), access to company documents/records, questionnaires, surveys, focus groups and/or other interview

techniques? Does your research entail any process which requires ethical approval?
(please enter ✓ in the appropriate box)

NO		You do not need to submit an application to the Ethics Review Manager
----	--	--

Name of Student (Print name): Mustansar Hussain

Signature: Mustansar

Date: 14/02/2023

SECTION 2: TO BE COMPLETED BY THE PROJECT SUPERVISOR

I understand that the above project requires/does not require* ethical approval
(*please delete as appropriate).

Supervisor (print name):

Signature:

Date:

IMPORTANT: please note that by signing this form all signatories are confirming that any potential ethical issues have been considered and, where necessary, an application for ethical approval has been/will be made via the Ethical Review Manager software.

Any project requiring ethical approval but which has not been given approval will not be accepted for marking.

Ethical approval cannot be sought in retrospect.