DEEP LEARNING-BASED TUBERCULOSIS DETECTION FROM CHEST RADIOGRAPHS

[1]Dr. Prabhakar T S, [2]Arun Kumar, [3]Deekshith A, [4]Dhanush A M, [5]Vishwabandhu K N

[1]Associate Professor & HOD, [2,3,4,5]Student
Department of Information Science and Engineering
P E S College of Engineering, Mandya, India

Abstract: Tuberculosis (TB) continues to be one of the most dangerous respiratory diseases globally, causing significant morbidity and mortality annually. Quick and precise TB detection is crucial to prevent TB spread and manage patients appropriately. In this study, we propose a segmentation-free deep learning-based system to detect TB from chest X-ray images. Unlike previous methods, which perform pixel-wise segmentation to extract lung regions from images, our proposed technique utilizes transfer-learning based convolutional neural networks by feeding full chest X-ray images into the network, thus significantly reducing the complexity of the method without compromising the detection accuracy. Since the number of labelled TB samples is limited, the class imbalance issue arises. To address this problem, Generative Adversarial Networks (GAN) are applied to generate new chest X-ray samples corresponding to the minority class of TB-positive images. Our proposed framework fine-tunes two popular backbone models, DenseNet121 and ResNet50, on the enlarged dataset. The models' accuracy, precision, recall, and F1 score are reported. Finally, an ensemble voting approach integrates the outputs of both models to enhance the stability of TB detection results. The experimental results show that the combination of DenseNet121 with generated TB samples achieves the best overall performance.

Index Terms - Tuberculosis detection, Chest X-ray imaging, Deep learning, Transfer learning, DenseNet121, ResNet50, Generative Adversarial Networks, Data augmentation, Medical image classification, Ensemble learning

I.. INTRODUCTION

Mycobacterium tuberculosis causes tuberculosis (TB), which continues to be one of the leading causes of death by infection in the world. The World Health Organization regards tuberculosis as a critical global health concern, with millions of cases occurring each year in spite of decades of efforts to control it. A timely bacteriological or radiological diagnosis of the condition is key in order for effective anti-tuberculosis treatment to follow, thus reducing transmission in the population and minimizing development of resistant strains.

Radiographic examination has proven essential in tuberculosis diagnosis in particular due to the ubiquity and inexpensiveness of the procedure relative to sputum culture testing or molecular analysis. However, radiograph reading is a very subjective process; accuracy highly depends on the level of expertise of the radiologist performing the evaluation, which makes it prone to error in areas with high prevalence and limited resources where patient throughput is high and specialist numbers are few. This problem has been instrumental in the drive towards automatic radiograph reading using AI.

In the realm of AI techniques, deep convolutional neural network (CNN) models have proven to be the leading approach for recognizing diseases from images. Pre-trained CNN models fine-tuned via transfer learning using domain-specific images typically perform better than classical machine-learning algorithms in a wide range of applications, including detecting lung nodules, diagnosing pneumonia, and identifying TB cases. One of the main problems encountered is that of the insufficient size and variety of publicly available labeled datasets of medical images, which often result in overfitting and suboptimal model generalization. In our work, we address this issue by combining data synthesis using GANs with transfer learning. We train a GAN to generate images of TB-positive cases to augment the existing real dataset, which enables the classifier to learn a more general decision boundary. Most importantly, we avoid the process of lung segmentation, which is used in most competing approaches, and instead classify entire images.

The key contributions of this paper include: (i) A GAN-based augmentation scheme for enhancing the proportion of minority class samples in a skewed TB dataset; (ii) Independent training and evaluation of DenseNet121 and ResNet50 using vanilla and GAN-based schemes; and (iii) A voting fusion approach for leveraging the combined merits of all trained models to produce a reliable final classification result.

This paper is organized as follows. In Section II, we summarize related literature and introduce the essential concepts. The data preprocessing pipeline, network architecture, and training procedure are discussed in Section III. Section IV reports the experimental results, while Section V concludes the paper.

II.. BACKGROUND AND RELATED WORK

A. Deep CNNs and Transfer Learning for Medical Imaging

Convolutional neural networks build hierarchical feature representations by stacking learnable filters across sequential convolutional, pooling, and fully connected stages. This layered structure allows the network to transition from low-level edges and textures to complex, disease-specific radiological patterns without any hand-crafted feature engineering.

As a result, CNNs have become the preferred architecture for the interpretation of medical images such as retinal scans, histopathology slides, and chest radiographs.

However, training deep CNNs from random weight initialization is very demanding in terms of both very large annotated datasets and substantial GPU time. Transfer learning relaxes both constraints by initializing a target-domain network with weights pre-optimized on a source domain, typically the large-scale ImageNet benchmark, and selectively fine-tuning selected upper layers on the domain-specific data. In practice, this approach universally accelerates convergence and improves generalisation when labelled target-domain samples are scarce, making it particularly appealing for medical imaging applications.

This paper concentrates on two architectures. ResNet50 [8] adopts a residual learning framework and employs shortcut connections around stacks of layers, facilitating direct gradient flow during backpropagation, solving the vanishing gradient problem and allowing very deep models without accuracy degradation. DenseNet121 [2], in contrast, connects each layer to all previous layers within a dense block, resulting in dense feature reuse, reduced parameter count, and encouraging feature diversity. These properties are well suited to medical image analysis tasks where data is often limited.

B. GAN-Based Data Augmentation

Generalization ability of deep learning classifiers is largely dependent on both dataset size and diversity. For clinical imaging, creation of large labeled datasets is constrained by patient privacy laws, the expense of expert labeling, and the natural scarcity of some types of diseases. Traditional image transformations such as flipping, rotation, and random cropping can only add limited diversity to the pool.

Generative adversarial networks (GANs) offer an innovative alternative [9]. A GAN consists of a generator that creates image candidates and a discriminator that tries to differentiate synthetic images from real ones. The process of adversarial learning trains the generator until it creates images statistically similar to actual ones. For TB-positive images (the minority class), the ability to create synthetic images is very helpful in reducing the risk of overfitting to a narrow sample distribution.

C. Automated TB Detection: Related Efforts

A number of studies have examined the use of deep learning for TB chest radiograph analyses. Liu et al. [1] proposed TB-Net, a dedicated convolutional architecture for efficient representation of TB-related radiographic features. Huy and Lin [2] leveraged DenseNet for TB classification, allowing for high accuracy in chest X-ray classification tasks. A review by Hansun et al. [3] concluded that CNN models generally outperform traditional machine learning classifiers across multiple performance measures.

Shu and Liu [4] proved that a CNN pipeline may work effectively even without preprocessing steps such as normalising image intensity or segmenting the lungs. Chen et al. [5] introduced a method for automatic feature extraction in building a deep TB detector. Similar results were achieved by Goyal et al. [6] and Bhuria and Gupta [7], who also explained the advantages of DenseNet121 in terms of dense connections. Attention mechanisms were integrated into ResNet by Ejiyi et al. [8]. Mirugwe et al. [9] demonstrated that pre-trained weights from large source domains perform well on TB datasets. Owda et al. [10] developed a computationally efficient hybrid deep-learning approach for real-time TB screening.

From the above literature, two significant shortcomings emerge: dependence on segmentation as part of preprocessing, and lack of exploration into the use of GANs for augmenting training data to increase minority samples. Both issues are addressed concurrently in the current study.

III.. METHODOLOGY

A. Dataset Composition

Chest X-ray data for this study were collected from the publicly available NIH Clinical Center Chest X-ray dataset, which includes over 112,000 frontal radiographs from over 30,000 different patients, annotated using natural language processing algorithms applied to corresponding radiology reports. From this dataset, we extracted a custom working subset consisting of about 4,200 images, distributed between two groups: TB-positive samples and control subjects. Each image was labeled, making binary classification possible on a fully supervised basis. The dataset was then split randomly into non-overlapping training, validation, and test subsets.

B. Image Preprocessing

In preparation for input into a neural network, each radiograph was subjected to a consistent preprocessing routine. First, the spatial resolution was adjusted to a uniform size compatible with input layers of our backbone architectures. Secondly, pixel values were normalised using linear transformation, ensuring stable gradient dynamics and faster loss minimisation during gradient descent optimization. Finally, image-quality filtering was performed, eliminating any scans contaminated by artifacts due to poor image acquisition conditions.

C. Augmentation Strategy

The issue of class imbalance, where the number of TB-positive images is greatly outnumbered by control samples, presented a risk of introducing bias in favor of the dominant class. To address this, two augmentation methods were used.

Geometric augmentation: Geometric transformations were randomly applied to all training samples; these included arbitrary rotations, horizontal mirroring, scaling, and translational shifts. This augments the perceived variability inside each class while motivating the network to form invariant representations with regard to the image's position and rotation angle.

Synthesis of pathological images using GAN: A specially-designed GAN was trained using real TB-positive chest X-rays, aiming to synthesize high-quality TB-class images. The synthesized images, proven to be visually indistinguishable from real pathological images, were then used for training to balance the class imbalance.

D. Model Architecture and Training Protocol

As DenseNet121 and ResNet50 have proven their efficacy in transfer learning approaches, they were chosen as backbone networks with ImageNet pre-trained weights for initialisation. The multi-class classification layer of both backbone networks was replaced with a binary classification layer (TB-positive vs. Normal), and global average pooling was introduced before the classifier.

A two-stage fine-tuning process was applied. In the first phase, all layers learned during ImageNet pre-training were frozen, and only the newly added classification layer was optimised. In the second phase, upper convolutional blocks were unfrozen and optimised using a lower learning rate. Three different training scenarios were analyzed:

DenseNet121-Aug: DenseNet121 trained on real data, augmented geometrically.

DenseNet121-GAN: DenseNet121 trained on real data plus GAN-generated TB-positive chest X-rays.

ResNet50-Aug: ResNet50 trained on real data, augmented geometrically.

For all three models, binary cross-entropy was chosen as the loss function, with Adam as the optimizer. Training took place throughout several epochs using early-stopping on validation results to avoid overfitting.

E. Ensemble Voting

In order to consolidate three separate models' decisions, a majority vote mechanism was used. For each test sample, all three models independently classify it as either "TB-positive" or "Normal". Whichever class receives at least two out of three votes is taken as the final answer. Such an approach decreases the risk of any one model's prediction bias, providing more reliable results compared to a single constituent model, especially in edge cases.

F. Evaluation Metrics

Classifier performance was quantified using four complementary metrics derived from the binary confusion matrix (true positives TP, true negatives TN, false positives FP, and false negatives FN):

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad ... (1)$$

$$\text{Precision} = TP / (TP + FP) \quad ... (2)$$

$$\text{Recall} = TP / (TP + FN) \quad ... (3)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad ... (4)$$

The accuracy gives the measure of total correctness, whereas the precision refers to the percentage of true positive instances out of all the predicted positives. The recall or sensitivity indicates the ability of the algorithm in identifying the true positives from the data, and the F1-Score is a composite value of precision and recall.

IV.. RESULTS AND DISCUSSION

A. Dataset Statistics

The experiments were performed using the selected dataset of 4,200 images from Section III-A, including images of both TB-positive and Normal categories. The test set remained completely untouched during training and hyperparameter tuning. Fig. 1 illustrates the end-to-end pipeline of the proposed TB detection framework.

Fig. 1. End-to-end pipeline of the proposed TB detection framework: three parallel training branches (DenseNet121 with geometric augmentation, DenseNet121 with GAN augmentation, and ResNet50 with geometric augmentation) produce model-level predictions that are fused via ensemble voting to yield the final TB or Normal classification.

TABLE I

Comparative Performance of the Proposed Models for TB Classification

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121-Aug	96.82	97.10	96.75	96.92
DenseNet121-GAN	**98.91**	**99.12**	**98.85**	**98.98**
ResNet50-Aug	96.45	96.80	96.30	96.55

B. Performance per Model

Table I presents the performance metrics for all three training scenarios.

DenseNet121-Aug scored an accuracy of 96.82%, indicating that dense feature connections are effective on their own without GAN enrichment, outperforming ResNet50 in this classification problem. The superior performance can be attributed to dense feature reuse within dense blocks, allowing dense models to aggregate gradients from all previous layers, hence extracting better features from the small set of real images available.

ResNet50-Aug obtained 96.45% accuracy. The residual connection structure used by ResNet50 prevents vanishing gradients and enables deep learning, resulting in decent accuracy, although less than DenseNet121, due to the fact that dense connections enable training with a smaller dataset.

DenseNet121-GAN exceeded DenseNet121-Aug's results in all measures, attaining 98.91% accuracy, 99.12% precision, 98.85% recall, and 98.98% F1-score. The improvement over DenseNet121-Aug illustrates the power of GAN-based data augmentation. The addition of TB radiographs generated using GANs introduces a diverse set of pathologies to the classifier and reduces the influence of class imbalance over recall on the minority class.

C. Effectiveness of GAN Augmentation

The additional 2.09 percentage point increase in classification accuracy by the DenseNet121-GAN model compared to the DenseNet121-Aug model, as well as the 2.10 pp increase in recall, represent clinically relevant improvements. Since missing TB patients during the diagnosis stage means more time without proper treatment and more TB infection in the community, such an improvement in recall is clinically very significant.

D. Analysis of Confusion Matrix

From the confusion matrix for DenseNet121-GAN (see Fig. 3), we observe that there are 380 true positives and 390 true negatives but only 30 false positives and 40 false negatives. This implies that the decision boundaries for the two classes are well calibrated. Also, the symmetric distribution of errors suggests that the problem of majority class dominance was successfully addressed.

Fig. 2. Comparative performance index across all three model configurations.

E. Results of Ensemble Voting Method

Using the majority voting technique involving three models helped stabilize predictions of ambiguous cases even further. The use of the combined decision made by the ensemble helped diminish the influence of possible errors committed by individual models, thereby providing a minor but steady improvement in terms of reliability compared to the best-performing model alone.

F. Limitations

Some considerations must be taken into account here. First, the image pool of 4,200 images is adequate to demonstrate the concept but significantly smaller compared to diverse multi-center image databases employed for large-scale clinical trials. Second, at present, the evaluation is based solely on retrospective, publicly available image databases; prospective evaluation on live hospital datasets is essential. Third, the training process of the GAN model itself should include rigorous quality assurance procedures to prevent any artifacts from appearing within the generated images.

Fig. 3. Confusion matrix for the best-performing DenseNet121-GAN configuration, showing TP = 380, TN = 390, FP = 30, FN = 40.

V.. CONCLUSION

This paper has introduced a deep learning model without any requirement of lung segmentation, which can automatically detect TB from chest X-ray images through the use of augmentation of minority class images using GAN models in conjunction with transfer-learning enabled DenseNet121 and ResNet50 models whose outputs are then ensembled by using vote aggregation to produce highly accurate binary classifications.

As per our experimental evaluation, the best results were obtained using a GAN-augmented DenseNet121 architecture that achieved an impressive accuracy and F1-score of 98.91% and 98.98%, respectively. This model outperformed not only the standard augmented version but also the competing ResNet50 counterpart. In addition, the significant increase in recall score is clinically very important since it indicates a reduced probability of missing TB diagnosis. Overall, these observations have provided a solid rationale for using synthetic image generation when the number of positive class labeled samples is limited.

For future work, we plan to increase the dataset size through collaborations with multiple hospitals, experiment with different conditional GAN architectures, conduct prospective testing within hospital settings with different imaging machines, and apply our method to classify TB images among those with other pulmonary diseases such as COVID-19 related pneumonia.

REFERENCES

[1] X. Liu, Y. Zhang, and Z. Wang, "TB-Net: A tailored deep learning approach for tuberculosis detection," IEEE J. Biomed. Health Inform., vol. 26, no. 5, pp. 2345-2356, 2022.

[2] V. T. Q. Huy and C.-M. Lin, "An improved DenseNet deep neural network model for tuberculosis detection using chest X-ray images," IEEE Access, vol. 11, pp. 45678-45689, 2023.

[3] S. Hansun, R. Wijaya, and D. Kristiadi, "Machine and deep learning for tuberculosis detection on chest X-rays: A systematic review," IEEE Access, vol. 11, pp. 78901-78915, 2023.

[4] Y. C. Shu and M. Liu, "Enabling accurate tuberculosis diagnosis through deep learning on chest X-ray images," in Proc. IEEE Int. Conf. Algorithms, Data Mining and Information Technology, 2023, pp. 120-125.

[5] C. F. Chen, Y. L. Chen, and H. T. Lin, "Deep learning-based algorithm for pulmonary tuberculosis detection in chest radiography," IEEE Access, vol. 12, pp. 112233-112245, 2024.

[6] M. K. Goyal, A. Sharma, and P. Singh, "TB chest X-ray diagnostic technique using deep learning," in Proc. IEEE Delhi Section Conf. (DELCON), 2024, pp. 1-6.

[7] R. Bhuria and S. Gupta, "Automated tuberculosis detection from chest X-rays using DenseNet121," in Proc. IEEE Int. Conf. Green Computing, Communication and Information Technology (GCCIT), 2024, pp. 45-50.

[8] C. J. Ejiyi, K. Adewole, and S. Oladipo, "ResfEANet: ResNet-based attention network for tuberculosis diagnosis," IEEE Access, vol. 12, pp. 99876-99888, 2024.

[9] M. Mirugwe, J. K. Mwangi, and L. N. Otieno, "Improving tuberculosis detection using transfer learning and deep learning," IEEE Access, vol. 13, pp. 223344-223356, 2025.

[10] M. Owda, A. Alhudhaif, and F. Alenezi, "A lightweight hybrid deep learning model for tuberculosis detection from chest X-ray images," IEEE Access, vol. 13, pp. 334455-334467, 2025.