Comparison of convolutional neural networks for tuberculosis x-ray image classification

Yatzitl Donají Arguelles Salgado [1] and Ismael Eliezer Pérez Ruiz [1]

[1] Universidad Modelo, Mérida Yucatán 97305, MEX
14150140@modelo.edu.mx

Abstract. Tuberculosis (TB) is a leading infectious cause of death worldwide, with early diagnosis critical for effective treatment. Chest X-rays are widely used for TB detection, but their interpretation can be subjective. This study compares the performance of three deep learning models—DenseNet121, MobileNetV2, and InceptionV3—for automated TB classification in X-ray images. A balanced dataset of 1,400 images (700 TB-positive and 700 normal) was preprocessed to 224×224 pixels and split into training (75%), validation (15%), and test (10%) sets. Transfer learning was employed, fine-tuning each architecture while preserving pre-trained weights from ImageNet.

Results showed MobileNetV2 as the most balanced model, achieving 100% sensitivity (correctly identifying all TB cases) with strong precision (0.93). DenseNet121 had high overall accuracy (0.95) but produced 10 false negatives, risking missed diagnoses. InceptionV3 demonstrated robust performance but leaned toward classifying images as normal, potentially increasing false negatives. MobileNetV2's efficiency and reliability suggest it is well-suited for TB screening, particularly in resource-limited settings. These findings highlight the potential of CNNs to assist in TB diagnosis while underscoring the need for model-specific clinical validation to minimize diagnostic errors.

Keywords: Tuberculosis, Inception V3, MobileNet V2 0.1, DenseNet, Convolutional Neural Networks.

1 Introduction

Tuberculosis (TB) is a chronic infectious disease caused by the bacterium *Mycobacterium tuberculosis*, which mainly affects the lungs, although it can involve other organs [1]. It is transmitted by the airborne route, through droplets expelled when coughing, sneezing or talking, which facilitates its spread in densely populated communities or communities with limited access to health services. The diagnosis of TB is based on various tests, with chest radiography being a fundamental tool for detecting signs of active pulmonary disease. However, its interpretation can be complex, since radiological manifestations vary according to the stage of the disease and may overlap with other pulmonary pathologies, which represent a challenge for physicians when establishing an accurate diagnosis. Moreover, in some cases it can be a fatal disease, so early detection of the disease prevents the spread of the bacteria and timely treatment can be provided to the patient [2][3].

Tuberculosis remains one of the main threats to global public health, since the most recent report of the World Health Organization (WHO, 2024), in 2023 approximately 8.2 million cases of TB were diagnosed globally, of which 1.25 million resulted in deaths, thus, it remains the leading cause of death from infectious disease in the world [4][5].

In recent years, the use of deep learning with convolutional neural networks (CNNs) has transformed the field of computer vision thanks to its ability to automatically extract and model abstract features, surpassing the performance of other supervised and unsupervised algorithms [6]. These feedforward networks process high-definition RGB images through multiple layers of neurons whose weights and parameters can be adjusted during training. Their architecture comprises feature extraction modules and classification layers, including convolution operations to identify spatial patterns, clustering to reduce dimensions and batch normalisation to improve model stability. Thanks to this hierarchical configuration, CNNs can represent visual information efficiently and robustly, making them suitable for complex image analysis tasks such as those involving the Inception, MobileNet and DenseNet architectures [7].

The MobileNetV2 architecture, described in the article 'MobileNetV2: Inverted Residuals and Linear Bottlenecks", introduces key innovations for efficient convolutional networks, such as inverted residual blocks and linear bottlenecks. These innovations allow the number of parameters and operations to be reduced without significantly compromising model performance, making them ideal for image processing tasks such as classification on devices with limited resources. Each block comprises a feature expansion with 1×1 convolutions, followed by a 3×3 depthwise convolution that filters each channel separately, and a projection to a lower dimension without ReLU activation to preserve critical information. Additionally, residual connections are employed wherever possible, facilitating learning in deep networks. This architecture strikes a balance between accuracy and efficiency, adapting to different levels of complexity without requiring high-performance hardware [8].

Conversely, the Inception-v3 architecture employs Inception modules, which execute multiple convolutional operations in parallel, including 1×1, 3×3, and 5×5 filters, and concatenate them in the output channel. This enables the network to capture information at various scales. It also introduces key optimisations, such as the factorisation of large convolutions (e.g., a 5×5 convolution is replaced by two 3×3 convolutions) and the use of asymmetric convolutions (such as 1×7 followed by 7×1), which reduces computational cost without affecting representation capacity. Furthermore, it incorporates techniques such as label smoothing, batch normalisation, and auxiliary regularisation to improve generalisation and accelerate training. This architecture has been extensively adopted due to its optimal balance between performance, accuracy, and computational efficiency in the domain of computer vision [9].

The Densely Connected Convolutional Networks (DenseNet) architecture signifies a substantial advancement in the domain of deep convolutional networks, with the introduction of dense connections between layers. In contrast to the conventional approach of passing the output of a single layer to the subsequent layer, this network employs a concatenation strategy. Specifically, it integrates the outputs of all preceding layers as inputs to each subsequent layer within a designated dense block. This

approach fosters enhanced feature reuse and substantially enhances gradient propagation during the training process. Consequently, it enables the training of more complex networks with reduced risk of performance degradation. The network is organised into dense blocks, followed by transition layers. These layers apply a 1×1 convolution and a pooling operation. These operations control the size of the feature map and prevent exponential growth of the channels. The modular configuration of the network enables the depth to be adapted to the task at hand, thereby achieving optimal results in image classification, detection, and segmentation with noteworthy computational efficiency. As DenseNet has been demonstrated to enhance accuracy, it is also advantageous in terms of memory and training time [10].

This research contributes to the study of processing chest X-ray images, with a focus on the automatic classification of healthy and sick patients for the detection of tuberculosis. This approach is an effective complementary tool for supporting clinical diagnosis. The study's primary objective is to compare the performance of DenseNet, MobileNet and InceptionV3 convolutional neural network architectures by evaluating metrics such as sensitivity, recall and F1 score when classifying X-ray images of patients with and without tuberculosis.

2 Methods

2.1 Dataset "Tuberculosis (TB) Chest X-ray Database"

Initially, the dataset was procured from the Kaggle repository, which houses a compendium of chest X-ray images that have been classified as positive for tuberculosis, in addition to images corresponding to healthy patients. The dataset, designated "Tuberculosis (TB) Chest X-ray Database" encompasses a total of 4,200 images, of which 3,500 are from patients without tuberculosis (normal) and 700 images correspond to patients with tuberculosis. In order to maintain class balance and avoid bias during model training, 700 images were randomly selected for the "Normal" category. In this instance, data augmentation techniques were not applied, thus ensuring the exclusive use of real data.

2.2 Preprocessing the dataset

Given the heterogeneity of the image dimensions within the set, a standardisation process was implemented to ensure consistency, resulting in a final resolution of 224 × 224 pixels. The selection of this resolution is supported by the findings of Hooda et al. [11], which utilised this size for the DenseNet and Inception architectures. In the case of MobileNet, the same resolution was utilized because, according to the official documentation [12], when the 'include_top' parameter is set to 'False', the input images must be standardized to 224 × 224 pixels.

The dataset was then divided into three distinct sets: a training set comprising 75% of the total data, a validation set consisting of 15%, and a test set containing the residual 10%. This approach was adopted with the objective of attaining an equilibrium between

the model's learning process, the refinement of its parameters, and the subsequent evaluation of its final performance. The 75% allocated to training enables the neural network to acquire a sufficient number of examples to discern relevant patterns in the images. The 15% allocated for validation is instrumental in regulating the training process, thereby facilitating the identification of overfitting and the requisite adjustment of hyperparameters. The residual 10% designated for testing ensures an objective evaluation of the model's final performance. This ratio has been frequently used in research involving moderate data sets, and its application is well-documented in the literature as a means of maintaining model integrity and avoiding bias in evaluation [13].

2.3 Dataset training

Three deep convolutional network architectures were utilised in this research study: MobileNetV2, InceptionV3, and DenseNet are all based on the transfer learning technique. This strategy enabled the reutilisation of models that had been trained with the ImageNet dataset, with subsequent adaptation to the specific domain of medical images. In each instance, the initial top classification layers were eliminated to construct a new classification head that was adapted to the number of classes in the dataset. The input images were previously normalised and preprocessed using the functions corresponding to each architecture, thus ensuring compatibility with the pretrained weights.

During the process of fine-tuning, a consistent approach was adopted across the three architectures, based on the freezing of the initial layers and the updating of the final layers. The procedure under discussion is founded upon the principles of transfer learning, which indicate that the layers in closest proximity to the network input learn general visual patterns such as edges and textures, while the deep layers acquire more specific representations of the original training domain, as has been established in previous studies [14][15]. In the case of MobileNetV2, the majority of the architecture was kept frozen, with training being permitted only at the level of the final layers. Conversely, InceptionV3 underwent modifications by unfrozen all layers post-number 290, while preserving the integrity of the remaining model components, in accordance with the recommendations outlined in the Keras documentation for transfer learning [16].

The three networks under consideration were compiled using the Adam optimizer, a widely recognised software component for its computational efficiency, dynamic learning rate adaptation, and robustness against sparse gradients. The validation of this optimiser has been previously undertaken in a range of studies, including that of Haal et al., where it demonstrated efficacy in the fine-tuning of models for the automatic detection of tuberculosis in chest X-rays. In all cases, a loss function designed for multiclass classification was employed, and the models were trained for fifteen epochs, using separate sets for training and validation. In order to address the potential for imbalances in class composition, a system of class weights was implemented with the objective of promoting more equitable learning conditions. During the training process, the accuracy and loss metrics for both the training and validation sets were monitored and

graphically represented to facilitate analysis of model convergence and early detection of overfitting.

3 Results

The performance of the DenseNet121, MobileNet, and Inception architectures was evaluated using chest X-ray images for the purpose of tuberculosis detection. The metrics of precision, sensitivity, F1 score, and accuracy are summarised in Table 1, alongside the respective confusion matrices.

Table 1. A comparison of evaluation metrics between CNN architectures is presented herein. Values close to 1 indicate greater classification effectiveness. The arithmetic mean is denoted by 'Macro Avg', while the mean weighted by support is denoted by 'Weighted Avg'.

Table 1. A comparison of evaluation metrics between CNN architectures is presented herein. Values close to 1 indicate greater classification effectiveness. The arithmetic mean is denoted by 'Macro Avg', while the mean weighted by support is denoted by 'Weighted Avg'.

Table 1. A comparison of evaluation metrics between CNN architectures is presented herein. Values close to 1 indicate greater classification effectiveness. The arithmetic mean is denoted by 'Macro Avg', while the mean weighted by support is denoted by 'Weighted Avg'.

Métrica	Clase	MobileNet	Inception	DenseNet
Precisión	NORMAL	1	1	0.89
	TUBERCULOSIS	1	0.94	1
Recall	NORMAL	1	0.93	1
	TUBERCULOSIS	1	1	0.9
F1-Score	NORMAL	1	0.96	0.94
	TUBERCULOSIS	1	0.97	0.95
Soporte	NORMAL	92	82	81
	TUBERCULOSIS	92	102	103
Exactitud (Accuracy)	-	1	0.97	0.95
Macro Avg	-	1	0.97	0.95
Weighted Avg	-	1	0.97	0.95

Although DenseNet121 showed high overall accuracy, the confusion matrix reveals a concerning tendency to misclassify some tuberculosis cases as normal. While the model correctly identified all normal cases, it recorded ten false negatives in the pathological class, which could affect its clinical applicability despite its favourable overall metrics. By contrast, MobileNet offered a better balance between sensitivity and accuracy, correctly identifying all tuberculosis cases and slightly decreasing the detection of normal images. This behaviour suggests a controlled bias towards the pathological class, which is desirable in medical diagnostic tasks.

Finally, the Inception-based model performed well overall, although its sensitivity in detecting tuberculosis was lower. The confusion matrix shows a tendency to classify images as normal, which could increase the risk of false negatives. Overall, MobileNet is the most balanced option for tuberculosis detection in medical images, while DenseNet and Inception have specific limitations that must be considered depending on the clinical context of application.

References

1. Medicina Integral. (n.d.). Radiological manifestations of pulmonary tuberculosis. https://www.elsevier.es/es-revista-medicina-integral-63-articulo-manifestacionesradiologicas-tuberculosis-pulmonar-13029945
2. Rojas, P., & Contreras, A. (2004). Imaging diagnosis of pulmonary tuberculosis. Revista Chilena de Neuropsiquiatría, 42(4). https://www.scielo.cl/scielo.php?pid=S0717-93082004000400006&script=sci_arttext&utm_source=chatgpt.com
3. MedlinePlus. (n.d.). Tuberculosis test. https://medlineplus.gov/spanish/pruebas-de-laboratorio/prueba-de-tuberculosis/?utm_source=chatgpt.com
4. World Health Organization. (2024). Global Tuberculosis Report 2024. https://iris.who.int/bitstream/handle/10665/379339/9789240101531-eng.pdf?sequence=1
5. Pan American Health Organization. (2024). Tuberculosis re-emerges as leading cause of death from infectious disease. https://www.paho.org/es/noticias/1-11-2024-tuberculosis-resurge-como-principal-causa-muerte-por-enfermedad-infecciosa
6. Iturbe Herrera, A. (2023). Computational diagnosis of tuberculosis using neural networks [Bachelor's thesis, Tecnológico Nacional de México]. https://rinacional.tecnm.mx/bitstream/TecNM/6615/1/DC_Alberto_Iturbe_Herrera_2023.pdf
7. Morales, J., et al. (2022). Medical image recognition using deep learning. Revista I+D Tecnológico, 18(3). https://www.redalyc.org/journal/5122/512261374010/html/
8. Zhang, X., et al. (2018). ShuffleNet: An extremely efficient convolutional neural network for mobile devices. arXiv preprint arXiv:1801.04381. https://arxiv.org/pdf/1801.04381v4
9. Howard, A.G., et al. (2015). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1512.00567. https://arxiv.org/pdf/1512.00567v3
10. Sandler, M., et al. (2016). MobileNetV2: Inverted residuals and linear bottlenecks. arXiv preprint arXiv:1608.06993. https://arxiv.org/pdf/1608.06993v5
11. Zhang, Y., et al. (2019). Transfer learning with convolutional neural networks for diabetic retinopathy detection. In Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 317–324. https://sci-hub.se/https://link.springer.com/chapter/10.1007/978-3-030-14802-7_34
12. Keras Applications. (n.d.). MobileNet. https://keras.io/api/applications/mobilenet/?utm_source=chatgpt.com
13. Brownlee, J. (2020). Train/test split for evaluating machine learning algorithms. Machine Learning Mastery. https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/
14. Szegedy, C., et al. (2014). Going deeper with convolutions. arXiv preprint arXiv:1411.1792. https://arxiv.org/pdf/1411.1792

15. Tan, M., & Le, Q. V. (2018). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1805.08974. https://arxiv.org/pdf/1805.08974
16. Keras Documentation. (n.d.). Transfer learning and fine-tuning. https://keras.io/guides/transfer_learning/