■ RESEARCH ARTICLE

An Application on Chest X-Ray Images for the Detection of Tuberculosis Disease by Employing Deep Convolutional Neural Networks

Hatice Koç [a, b†], **Abdulkadir Hızıroğlu** [b], **Ahmet Emin Erbaycu** [c]

[a] Management Information Systems, Gebze Technical University, Kocaeli, Turkey
[b] Management Information Systems, İzmir Bakırçay University, İzmir, Turkey
[c] Medicine, İzmir Bakırçay University, İzmir, Turkey
[†] haticekoc@gtu.edu.tr, corresponding author

RECEIVED MARCH 1, 2023
ACCEPTED APRIL 30, 2023

CITATION Koç, H., Hızıroğlu, A., & Erbaycu. A. E. (2023). An application on chest X-ray images for the detection of tuberculosis disease by employing deep convolutional neural networks. *Artificial Intelligence Theory and Applications, 3(*1), 51-64.

Abstract

Tuberculosis is the second infectious disease causing death after COVID-19. Diagnosing it is an easy and cheap via chest radiographs. However, some countries lack medical personnel and equipment for tuberculosis detection on chest radiographs. Computer-aided diagnosis and computer-aided detection systems utilizing deep learning can be employed to identify tuberculosis on medical images. Although there are some studies, they are insufficient for unbiased systems because these systems require the datasets having different features. The aim of this study is to evaluate the performance of pretrained networks for a classification application on chest X-ray images by utilizing the dataset from the Hospital in Turkey and Montgomery Count Dataset. The predictive models were implemented with the pre-trained DCNNs such as ResNet-50, Xception, and GoogLeNet. An Xception model provides the best performance.

Keywords: tuberculosis, deep convolutional neural networks, transfer learning, classification

1. Introduction

Tuberculosis (TB) is a contagious disease and is one of the top 10 disease causing death before COVID-19 pandemic. The Global Tuberculosis Report, by WHO in 2022, expresses that TB has influenced approximately a quarter of the world's population. About 50% of TB patients have passed away since they were not treated. Besides, TB can be described as a disease of poverty due to its unfavourable impacts on 30 high TB burden countries, which these countries were influenced 87% by TB [1] [2]. 8.925 TB cases were recorded in Turkey in 2020, which the number of male patients equal 57.2% while the number of female patients has 42.8% of all TB patients [3].

Early diagnosing can prevent and treat TB. Chest X-ray (CXR) is a cheaper and effortless technology used as a part of TB diagnosis. However, it requires systematically check different points of lungs on a CXR with a physician’s interpretation [4] [5] [6]. Low-income countries have the issues arising from the lack of medical personnel and equipment for TB diagnosis. Thus, several studies evaluate if computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems assist detect TB on chest radiographs (CRs)

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN : 978-605-69730-2-4 © 2023 University of Bakırçay

in order to solve these issues. These systems can collect and manipulate medical data and report medical examination by exploring potential abnormalities on CRs. They assist physicians and radiologists by identifying the localization of lesions, diseases, and/or causes of symptoms over examinations [7] [8]. That is, they are employed as second opinion during disease diagnosis [8] and medical decision-making.

Many studies have utilized artificial intelligence (AI) to create automated activities for diagnosis process and prognosis, medical decisions, and health practise by developing CADe and CADx systems. Some studies also employ AI to process medical images, biosensors, molecular data and electronic medical records (EMR). DL is an AI approach employed to examine CRs. The typical patterns of TB are easily recognized with satisfactory accuracy during medical examinations [1] [8] by using DL. Hence, many studies have used deep convolutional neural networks (DCNNs) to identify TB since medical image classification is one of major steps to improve CADe and CADx systems [9] which enable to recognize of localization of lesions, diseases, and/or symptoms [8].

On the other hand, many challenges have been faced when implementing DL based-CADe and CADx systems. There are limited data, the dataset containing outside available populations, and/or the dataset having insufficient variables and instances [10]. An unbiased DL model requires the dataset formed of balanced distribution and several demographic variables. Therefore, some research has been performed to explore how robust DCNNs algorithm to develop CADe and CADx systems determining TB on CXR images while some have investigated whether the DCNNs-based systems help tackle or not the lack of medical personnel and equipment for TB diagnosis. Although there are several research, the studies have been performed with similar datasets. Any study has never been carried out by a dataset that is not obtained from the hospitals in Turkey. This gap in academic literature has motivated to carry out this study. Expert view also encouraged for the research problem. Briefly, the purpose of the study is to fill the gap in the study utilizing CXR images from hospital in Turkey and analyse how robust the DCNNs to identify TB.

There are two experiments in this study. Firstly, exploratory data analysis (EDA) was applied to grasp the distribution of demographic variables because it allows decision support systems (DSS) to discover tendency and unique characteristics in dataset. Then, predictive analytics was applied with transfer learning approach in order to detect TB disease since it enables to CADx and CADe systems make predictions on new observation. Before implementing the models, data augmentation was applied. Performance evaluation was performed by comparing the confusion matrix, accuracy, recall, precision, misclassification rate and F1 score of each model.

The rest of the study covers four sections. In the second section, DSS, analytics, and AI were defined, and the literature was reviewed in terms of medical problems. The third section clarifies the methodology which covers five parts; model development, data collection and description, image data pre-processing, data analytics, and performance evaluation. The next section presents the results of EDA, and predictive models while the last section explains the contributions, limitations, and future of the study.

2. Literature Review

Healthcare sector requires intelligent systems that ease diagnosis, and support treatment by processing data about patients' lifestyles and having disease manifestations. CADe and CADx systems can be defined as a type of DSS that provide information of clinicians and deal with challenges from knowledge and information acquisition and management in clinical practice. Their usage grows tendency for personalized healthcare services, the request for accessibility for EMR and the distribution of convenient data [11].These systems employ analytics and internal and external data acquisition [12].

Analytics integrates actions, processes, and tools in the dimensions of descriptive, predictive, and prescriptive analytics [13]. AI provides intelligence and expertise for analytics [13]. ML, a part of AI, help develop automated and/or semi-automated processes by manipulating the data [14]. It is used to develop knowledge-based DSS employed since these systems increase accuracy rates and reduce errors in medical decisions and practices if it is intended to develop an unbiased system [15]. ANNs is one of the most commonly used ML techniques to predictive model [13] trained with structured, semi-structured, and/or unstructured data [16]. The rise in computational power and data have boosted the performance of ANNs and emerged DL. Additionally, CNNs is a type of DL technique that perform automatically feature extraction and segmentation of organs or an object [17]. Hence, several studies for CADe and CADx systems have used CNNs [18].

Some studies have analysed how robust DCNNs are when identifying TB on CXR images. They have employed transfer learning approach, implemented DCNNs models, or compared the performances of CNNs and other AI techniques. They have used open-source datasets and/or specific hospitals. Besides, the models have been implemented for either classification, feature extraction or segmentation. Their performances have been evaluated with accuracy, recall, precision, F1 score, AUC and ROCs.

The studies of [19], [20], and [21] classify lung diseases while the studies of [5], [22], [23], [24], and [25] focus on TB disease. Additionally, the studies of [22] and [25] utilized transfer learning, unlike the study of [5]. Abiyev and et. al. [19] examined the applicability of CNNs, backpropagation neural networks (BPNNs), and competitive neural networks (CpNNs) for the classification of some chest diseases. The performances of the models were compared with accuracy, MSE, and training time. The study was performed with the insufficient dataset for detecting various lung diseases, and the dataset does not include demographic variables.

Mamalakis and et. al. [20] constructed a new deep transfer learning pipeline and evaluated the performance of this pipeline on the detection of COVID-19, pneumonia, TB, and healthy patients from CXR images. The study has the dataset that contains a smaller size of cohort for COVID-19 and does not have various demographic features and multi-label lung disease.

In the study of Ölmez and et. al. [21], the biomedical classification applications were represented by using CNNs and PNNs algorithms for detecting several lung diseases in both Turkey and the World. The performances of the models were compared with accuracies. The study indicates a CNNs-based DSS can be improved for the detection of lung diseases.

On the other hand, Hwang and et. al. [5] applied the classification to determine TB disease on CXR images without transfer learning. AUROC, AUAFROC, sensitivity, and specificity were evaluated to measure the performances of the models. Besides, the performances of the models were compared with the performance of physicians. According to the results, the DLAD algorithm outperformed most physicians and enhanced the performance of non-radiology physicians as the second reader.

In the study of Lakhani and et. al. [22], a transfer learning approach was employed to examine the efficacy of DCNNs for TB detection on CRs. ROC and AUCs were compared for the performance of the models. Data augmentation was also applied when comparing the performances. The results show that greater values for AUCs were presented with the pre-trained models, the best performance was obtained with the augmented data, and the highest performance was given by the ensemble models.

Cao and et. al. [25] conducted a study to deploy a mobile device-based computing system for TB diagnostics in Peru by using a transfer learning approach and mobile health technology in order to lessen TB diagnosis time. The performances were evaluated with accuracy. The results indicated that the proposed approach is feasible.

However, in the study of [23] and [24], transfer learning was used for segmentation models, and then support vector machine (SVM) was used for the classification model. In the study of Karaca and et. al. [23], an automated DSS was proposed for TB detection. The performances of models were evaluated with accuracy and AUC. The results obtained from the study were also compared with the findings from some previous studies. Although the study was performed with insufficient data, the study indicated that data augmentation improves the accuracy of the detection of TB.

Oltu and et. al. [24] suggested an automated DSS that classifies normal and TB on CXR images. The impacts of data augmentation were also examined. Similar results were presented by MobileNet and VGG16 while the highest accuracy and AUC were presented by the MobileNet model utilized data augmentation with rotation. Poor performances were obtained when applied all data augmentation methods together and unnecessary both shifting and rotation for feature extraction.

The studies of [26], [27], and [28] employed a lung segmentation model for TB detection. Stirenko and et. al. [26] carried out the study to prove the efficiency of the lung segmentation technique without the pretrained DCNNs. EDA was also applied to understand the distribution of disease, age, and gender in the dataset and the image heights and widths. The values of accuracy were compared to evaluate the performances. The study indicated that better performance for training on the pre-processed dataset has been acquired after lung segmentation. The segmented and augmented data increased accuracy. However, the models were trained with a small and not-well-balanced dataset. The dataset is also retrospective dataset and has non-evident outliers.

Rahman and et. al. [27] utilized a transfer learning approach, and the original and segmented lungs in X-ray images in order to detect. The study has evaluated the performance of all classification models for the detection of TB. In the study, better performance was acquired with the segmented CXR images while testing the networks. The results show that classification accuracy can be increased with image segmentation.

Heo and et. al. [28] compared two DCNNs models for the detection of TB from CRs by using CXR images and demographic variables structures and image segmentation. The performances of the models were evaluated with ROC and AUC. The study represents that the demographic variables improve the performance of the CNNs model. CADe and CADx systems employing DCNNs can enable to detect TB on CXR images easily and enhance disease management. However, the study has the limitations that are the low number of demographic features and insufficient computational power.

As summarized in Table 1, the accuracy rates were 92.4% and 95.51% in the studies of Abiyev and et. al. [19] and Ölmez and et. al. [21] respectively, while the recall rate was discovered as 98.12% by Mamalakis and et. al. [20]. However, in the studies for detecting TB with classification, Hwang and et. al. [5], and Lakhani et. al. [22] ascertained the recall rates between 94.3% and 100%, and between 92% and 97.3% while the specificity rates between 84.1% and 100%, and between 94.7% and 98.7%, respectively. Cao and et. al. [25] discovered that the accuracy rate is 89.6% for the binary classification while the accuracy rate is 62.07% for the multi-classification. The results of accuracy, recall, and specificity indicate that CNNs can be used to detect TB and other lung diseases.

Table 1. Related empirical work

Study	Purpose	Dataset	Technique & Architecture	Performance Evaluation	Findings
[19]	To show the applicability of conventional and DL techniques for the classification of chest diseases.	The dataset that contains of 112,120 CXR images belonging to 30,805 different patients from the National Institutes of Health—Clinical Centre.	BPNNs CpNNs CNNs	Accuracy MSE Training time	**BPNNs** Accuracy:80.04% MSE: 0.0025 Training time:630 sec. **CpNNs** Accuracy: 89.57% MSE:0.0036 Training time: 2500 sec. **CNNs** Accuracy: 92.4% MSE:0.0013 Training time:2500 sec.
[20]	To develop a new deep transfer learning pipeline, called as DenResCov-19, and evaluated its performance by detecting COVID-19, pneumonia, TB, and healthy patients from CXR images.	The Pediatric CXRs Dataset, The IEEE COVID-19 CXRs Dataset, Shenzhen Dataset.	DenResCov-19 constituted by concatenating four blocks from the ResNet50 network and the DenseNet121 network with their width, height, and frames.	AUC-ROC Confusion matrix Precision Recall F1 score	**For each dataset:** AUC-ROC: 0.9960, 0.9651, 0.9370, and 0.9640. F1 score: 98.21%, 87.29%, 76.09%, and 83.17%. Recall: 98.12%, 89.38%, 59.28%, and 69.7%. Precision: 98.31%, 85.28%, 79.56%, and 82.90%.
[21]	To exemplify for biomedical classification applications using DL methods for the detection of lung diseases that are widespread in Turkey and the World.	The dataset that includes 38 different features indicating laboratory examination from the patients hospitalized because of lung diseases, and 357 subjects.	PNNs CNNs	Accuracy	**PNNs** Accuracy: 91.25%. **CNNs** Accuracy: 95.51%.
[5]	To improve a DL-based automatic detection algorithm (DLAD) for active TB on CRs as a second opinion and measure its performance with different datasets and by comparing it with	Seoul National University Hospital Dataset, Boramae Medical Center Dataset, Kyunghee University Hospital at Gangdong Dataset, Daejeon Eulji Medical Center Dataset, Montgomery Count Dataset, Shenzhen Dataset.	DCNNs	ROC Sensitivity Specificity	**For internal validation** AUROC: 0.988. AUAFROC: 0.977. **For external validation** AUROC: 0.977 to 1.000. AUAFROC: 0.973 to 1.000 Sensitivity: 94.3%-100%. Specificities: 91.1%-100%.

	physicians' performance.				
[22]	To analyse the efficacy of DCNNs for TB detection on CRs.	Belarus TB Public Health Program Dataset, Thomas Jefferson University Hospital Dataset, Montgomery Count Dataset, Shenzhen dataset.	AlexNet GoogLeNet.	ROC AUCs	**AlexNet** AUC: 0.90, 0.95, 0.98, and 0.98 Sensitivity: 92 % Specificity: 94.7% **GoogLeNet** AUC: 0.88, 0.94, 0.97, and 0.98. Sensitivity: 92 % Specificity: 98.7% **The ensemble of both** AUC: 0.99 Sensitivity: 97.3 % Specificity: 94.7%
[25]	To deploy a mobile device-based computing system that screens CXR images for TB diagnostics in Peru.	4701 CXR images were employed, which provided by Dr. Peinado in Peru.	GoogLeNet	Accuracy	**For binary classification** Accuracy: 89.6% **For multiclass categorization** Accuracy: 62.07%.
[23]	Proposed an automated DSS for TB detection.	Montgomery Count Dataset	VGG16 VGG19 DenseNet121 MobileNet InceptionV3 SVM	Accuracy AUC	**With data augmentation** Accuracy: 98.7%, 98%, 98.9%, 98.8% and 96.5%. AUC: 1.000, 0.990, 1.000, 1.000 and 0.999. **Without data augmentation** Accuracy: 87%, 86.2%, 80.4%, 74.6% and 79%. AUC: 0.900, 0.910, 0.870, 0.810 and 0.880.
[24]	Presented an automated DSS identifying TB and analysed how data augmentation influences the analysis was examined.	Montgomery Count Dataset, Shenzhen Dataset.	VGG16 MobileNet. SVM	AUC Accuracy	**MobileNet** Accuracy: 91.30%, 96.40%, 96.50%, 0.-96.60%, 96.50%, 91.40%, 91.40% and 87.80%. AUC: 0.970, 0.990, 0.990, 0.990, 0.990, 0.990, 0.970 and 0.950. **VGG16** Accuracy: 91.40%, 96.70%, 95.70%, 95.60%, 95.60%, 91.60%, 91.30% and 87.60%. AUC: 0.960, 0.990, 0.990, 0.990, 0.990, 0.970, 0.960 and 0.930.
[26]	To prove the efficiency of the lung segmentation technique with lossless and lossy data augmentation for predicting TB disease.	Shenzhen Hospital Dataset, Montgomery Count Dataset, JSRT Dataset	DCNNs Lung segmentation masks EDA	Loss Accuracy Intersection-Over-Union (IoU) or Jaccard Index F1-score	The highest training rate with lung segmentation Lower training rate for the segmented dataset with both lossless and lossy data augmentations.

[27]	To detect TB with transfer learning and the original and segmented lungs in CXR and evaluate the performance of all classification models for the detection of TB.	Kaggle CXR images dataset and corresponding lung mask dataset for the lung segmentation. For classification: National Library of Medicine datasets (Montgomery Count and Shenzhen datasets), Belarus dataset, NIAID TB dataset, and RSNA CXR dataset.	ResNet18 RestNet50 ResNet101 ChexNet InceptionV3 VGG19 DenseNet201 SqueezeNet MobileNet U-Net Modified U-Net		**ChexNet** Accuracy: 96.47% Precision: 96.62% Sensitivity: 96.47% Specificity: 96.51% F1-score: 96.47%. **DenseNet201** Accuracy: 98.6% Precision: 98.57% Sensitivity: 98.56% Specificity:98.54% F1-score: 98.56%.
[28]	To compare two DCNNs models for the detection of TB, which these models were I-CNN trained with only CXRs and D-CNN trained with demographic variables and employed the health examination data of annual workers.	Dataset contains the medical surveillance data acquired from workers at Yonsei University.	VGG19 InceptionV3 ResNet50 DenseNet121 InceptionResNet V2 U-Net	ROC AUC	**VGG19** D-CNN: ROC: 0.9213 AUC: 0.97147. I-CNN ROC:0.9075 AUC: 0.9570. **InceptionV3** D-CNN: ROC: 0.9045 AUC:0.9616. I-CNN ROC:0.8821 AUC:0.9523. **ResNet50** D-CNN: ROC:0.8955 AUC: 0.9219. I-CNN ROC: 8780 AUC:09219. **DenseNet121** D-CNN ROC:0.8864 AUC: 0.9472. I-CNN: ROC:0.8605 AUC: 0.9315. **InceptionResNetV2** D-CNN ROC: 0.8864 AUC: 0.9455. I-CNN ROC:0.8881 AUC: 0.9482. Sensitivities: D-CNN Sensitivity: 81.50%. I-CNN Sensitivity: 77.50%.

These studies reveal that DCNNs can be an alternative to determine from CXR images so as to implement automated CADe and CADx systems as a second opinion. However, an unbiased AI-based system requires several datasets covering diverse populations, instances, and variables. Hence, the DCNNs models were implemented without the lung segmentation and the feature extraction by contrast with [26], [27], [28], [23] and [24]. The binary classification was applied within in this study, unlike the studies of [19], [20], [21], and [25].

3. Methodology

3.1. Model development

The five steps are generally followed to implement a ML model, which are collecting data, data pre-processing, splitting data for developing the model, training the model, and testing and validating the model. Although feature engineering and domain expertise are parts of implementing a ML model, DL models may require some different processes [29]. The steps of a typical ML project are applied with additional processes related to using a DL technique embedded within that. The steps in Figure 1 were followed in this study in order to implement the models.

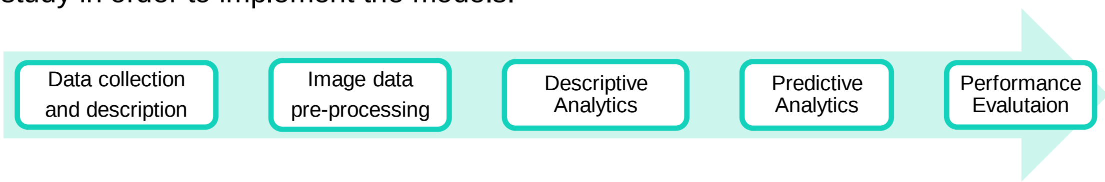

Figure 1. The steps for implementing DCNNs models

3.2. Data collection and description

Willemink and et. al. [30] proposed eight phases that are ethical approval, data access, querying data, data de-identification, downloading and storing data, data quality control, structuring data, and labeling data so as to prepare medical image for ML [30]. CXR images were prepared to implement DCNNs models by the similar phases in this study.

CXR images and demographic variables concerning these images were obtained from the Hospital in Turkey after ethical approval, which are about TB patients and healthy people. 444 of CXR images were included, which each image has the size of 3032 width and 2520 height. The total 222 data is about TB patients while the rest of data is related to healthy people. Demographic variables consist of age and gender, which were recorded “csv” file and the corresponding attributes “ID”, “Gender”, “Age”, “Images”, and “Class”. There are different two classes that are Healthy and TB. The class of TB is expressed with 1 while the class of healthy is expressed with 0. Furthermore, Montgomery Count Dataset were also ultilized. 110 CXR images were included, which each image has the size of 4020 width and 4892 height or 4892 width and 4892 heights. An example for CXR images’ view and summarizing demographic variables are given in Figure 2.

ID	Gender	Age	Image	Class
1	F	57	0_1.jpg	0
2	M	24	0_2.jpg	0
3	F	29	0_3.jpg	0
4	F	24	0_4.jpg	0
...	...	...	...	...
393	F	59	1_393.jpg	1
394	F	61	1_394.jpg	1
395	F	53	1_395.jpg	1
396	M	42	1_396.jpg	1
...	...	...	...	...

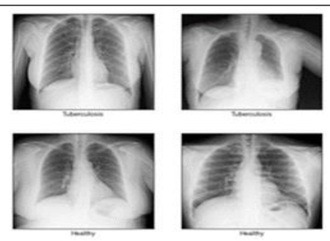

Figure 2. An example for demographic variables and CXR images’ views

3.3. Image data pre-processing

Image data pre-processing was performed in two stages which are data transformation and data augmentation in by coding in MATLAB. In the stage of data transformation, CXR images were resized as the dimensions of 224x224x3 before training the models implemented with the ResNet-50 and GoogLeNet architectures while the images were resized as the dimensions of 299x299x3 in order to train the models implemented with the Xception architecture. Besides, CXR images were transformed into colour images since these architectures require colour input images.
In the stage of data augmentation, rotation, reflection and shear intensity techniques were applied as the following: random rotation between -5 and 5, random reflection as 1, random shear intensity between -0.05 and 0.05.

3.4. Data analytics

3.4.1. Exploratory data analysis

Exploratory data analysis (EDA) is a technique used in descriptive analytics and utilizes retrospective data, statistical techniques and visualization tools before implementing a ML/DL model in order to explore useful information, inform results and support decision-making. In this study, EDA was performed to analyse demographic variables so as to understand the distributions of gender and age in terms of each class. Descriptive statistics and data visualization were employed by coding in Python on Jupyter Notebook.

3.4.1. Predictive analytics

The total of 22 models were implemented to discover the best performance for diagnosing TB disease. The models for binary classification were constituted by coding in MATLAB R2022b to predict TB disease on CXR images. Moreover, transfer learning approach was employed because implementing new DCNNs requires a large dataset as well as substantial time and cost [23]. Some characteristics of the architectures such as ResNet-50, Xception, and GoogLeNet are provided in Table 2.

Table 2. A brief for ResNet-50, Xception, and GoogLeNet

Characteristics	**ResNet-50**	**Xception**	**GoogLeNet**
Input dimensions	224x224x3	299x299x3	224x224x3
Class	1000	1000	1000
Layers	50	71	22
Last three layers	‘fc1000’, ‘fc1000_softmax’, 'ClassificationLayer_fc1000'	'predictions', 'predictions_softmax', 'ClassificationLayer_predictions'	‘loss3-classifier’, ‘prob’ ‘output’

The purpose of improving 22 models is to explore the best performance for diagnosing TB disease by comparing their performances. The first 11 models were trained with the data from the hospital by applying data augmentation. Nevertheless, the rest 11 model were trained with the combination of Montgomery Count Dataset and the hospital dataset after applying data augmentation. Additionally, the two datasets were split into training dataset and test dataset as 80% and 20%, respectively. Each model was built with 0.001 learning rate. The values of epoch, mini-batch size, and the types of optimizers were changed for each model, which are detailed in Table 3. The last three layers of each pre-trained network were also frozen, and then, a fully connected layer that contains two classes, a Softmax layer and classification output were added.

Table 3. The value of epoch, mini-batch size, and the types of optimizers

Dataset	Model	Optimizer	Epoch	Mini-Batch
The dataset from the hospital	ResNet-50 Model 1	Adam	10	25
	ResNet-50 Model 2	Sgdm	10	25
	Xception Model 1	Adam	10	25
	GoogLeNet Model 1	Adam	10	25
	GoogLeNet Model 2	Adam	10	32
	GoogLeNet Model 3	Adam	25	32
	GoogLeNet Model 4	Adam	32	64
	GoogLeNet Model 5	Sgdm	10	25
	GoogLeNet Model 6	Sgdm	10	32
	GoogLeNet Model 7	Sgdm	25	32
	GoogLeNet Model 8	Sgdm	32	64
The combined dataset	ResNet-50 Model 3	Adam	10	25
	ResNet-50 Model 4	Sgdm	10	25
	Xception Model 2	Adam	10	25
	GoogLeNet Model 9	Adam	10	25
	GoogLeNet Model 10	Adam	10	32
	GoogLeNet Model 11	Adam	25	32
	GoogLeNet Model 12	Adam	32	64
	GoogLeNet Model 13	Sgdm	10	25
	GoogLeNet Model 14	Sgdm	10	32
	GoogLeNet Model 15	Sgdm	25	32
	GoogLeNet Model 16	Sgdm	32	64

3.5. Performance Evaluation

Performance evaluation was performed by comparing the values of confusion matrix and performance measures of each model. For this, False Positive (FP), False Negative (FN), accuracy rate, recall rate, precision rate, F1 score, and misclassification rate were employed. Moreover, the positive class is the TB class while the negative class is the healthy class in this study.

4. Results

4.1. Exploratory Data Analysis

EDA was performed with three datasets which are the dataset from the hospital, Montgomery Count Dataset, and the dataset combined. The demographic variables were analysed with EDA. The findings from descriptive statistics are given in Table 3.

Table 4. Descriptive statistics for each dataset

Distribution of variables in the dataset	The dataset from hospital	Montgomery count dataset	The combined dataset
The total number of people	444	110	554
The total number of TB patients	222	55	277
The total number of females	242	58	300
The total number of males	202	52	254
The total number of female patients	86	19	105
The total number of male patients	136	36	172
The range for the distribution of age of TB patients	17 to 89	15 to 89	15 to 89
The range for the distribution of age of female patients	16 to 83	25 to 89	16 to 89
The range for the distribution of age of male patients	18 to 89	15 to 73	15 to 89
The average age of all people	43.19	41.07	42.77
The average age of females	39.76	42.12	40.21
The average age of males	47.30	39.90	45.78
The average age of patients	50.76	49.10	50.43
The average age of female patients	47.38	59.26	49.53
The average age of male patients	52.90	43.72	50.98

According to these findings, the age of the oldest patient for each gender in all data is the same although the age of the youngest patient for each gender is different. The average age of female patients and male patients are 50 and 51, respectively, while the average age of all TB patients is 50. The distribution of gender of TB patients indicates that the total number of male patients is 62% of all TB cases. The number of TB cases among male is 68% while the number of TB cases among female is 38%. Besides, the findings from the combined dataset show that TB cases were mostly common in adults and the number of TB cases is higher in males. This is parallel with the "Global Tuberculosis 2022" Report by WHO as the report expresses that the number of men affected by TB disease is more than the number women affected by TB disease [2].

4.2. Predictive models

The models were assessed by comparing their performances. The results of performance measures for each predictive model are presented in Table 5 while Figure 3 gives the accuracy and loss graphs for the first three best performances in terms of accuracy rate obtained from training dataset.

Table 5. A summary of comparing values of performance measures for each model

Model	Accuracy	Recall	Precision	F1 score	Misclassification	FN	FP
ResNet Model 1	78.41%	85.71%	68.18%	75.95%	21.59%	5	14
ResNet Model 2	80.68%	88.57%	70.45%	78.48%	19.32%	4	13
Xception Model 1	-	-	-	-	-	-	-
GoogLeNet Model 1	50%	-	-	-	-	-	-
GoogLeNet Model 2	50%	-	-	-	-	-	-
GoogLeNet Model 3	50%	-	-	-	-	-	-
GoogLeNet Model 4	50%	-	-	-	-	-	-
GoogLeNet Model 5	78.41%	100%	56.82%	72.47%	21.59%	0	19
GoogLeNet Model 6	71.59%	100%	43.18%	60.32%	28.41%	0	25
GoogLeNet Model 7	86.36%	90%	81.82%	85.72%	9.10%	4	8
GoogLeNet Model 8	81.82%	78%	88.64%	82.98%	5.68%	11	5
ResNet Model 3	80.90%	80.36%	81.82%	81.08%	10%	10	11
ResNet Model 4	88.18%	95.65%	80%	87.13%	10%	2	11
Xception Model 2	99.09%	100%	98.18%	99.10%	0.91%	0	1
GoogLeNet Model 9	50%	-	-	-	-	-	-
GoogLeNet Model 10	50%	-	-	-	-	-	-
GoogLeNet Model 11	50%	-	-	-	-	-	-
GoogLeNet Model 12	50%	-	-	-	-	-	-
GoogLeNet Model 13	50%	-	-	-	-	-	-
GoogLeNet Model 14	50%	-	-	-	-	-	-
GoogLeNet Model 15	50%	-	-	-	-	-	-
GoogLeNet Model 16	50%	-	-	-	-	-	-

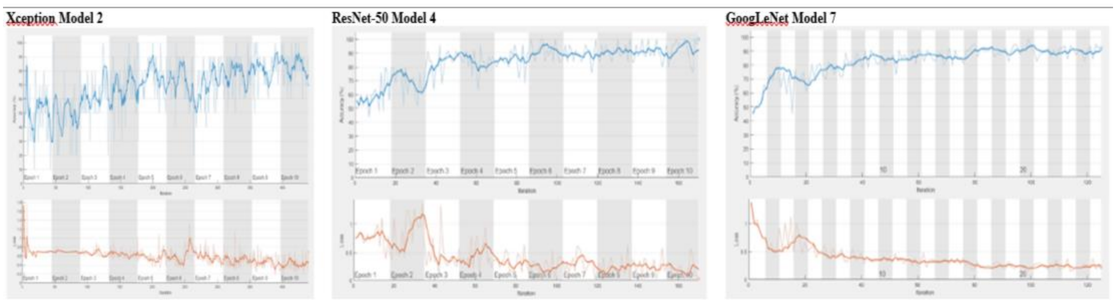

Figure 3. The accuracy and loss graphs for the first three best performances

These results show that the ResNet-50 Model 4 presents the highest accuracy rate, recall rate, and F1 score among the ResNet-50 models while it gives the lowest value of FN. When investigating GoogLeNet models, the highest accuracy rate, and F1 score are provided from the GoogLeNet Model 7. The highest precision rate is obtained from the GoogLeNet Model 8 while the GoogLeNet Model 8 gives the lowest misclassification rate. Besides, the lowest value of FP is provided from the GoogLeNet Model 8. However, the Xception Model 2 provides the best performance when evaluating in terms of accuracy, recall, misclassification, F1 score, FN, and FP among all the models.

5. Conclusion and Discussion

CADe and CADx systems are clinical DSS which can also utilize DL to manipulate CRs, deal with medical problems, support medical decision making. They enhance the productivity in diagnosis and treatment because DL rises accuracy rates and diminishes errors as well as reducing costs and time for detection. Furthermore, unbiased AI-based CADe and CADx systems require the dataset that covers balanced and different varied population and several features. Therefore, the study was performed to fill the gap in the study employed CXR images from hospital in Turkey when examining how robust the DCNNs to identify TB disease by comparing the performance of DCNNs architectures.

Experiments were carried out with EDA and predictive analytics. The distribution of demographic variables in the dataset were investigated with EDA while predictive analytics was performed for implementing the binary classification models with ResNet-50, Xception, and GoogLeNet. Before implementing the models, the data were augmented but lung segmentation were not applied. The performances of these DCNNs models were evaluated on the trained dataset from the Hospital in Turkey and Montgomery Count CXR datasets. For this, the accuracy, recall, precision, misclassification rate, F1 score, FP and FN values were compared.

The findings showed that the Xception Model 2 has the best performance, and the pretrained networks can be useful to improve CADx and CADe systems in determining TB disease on CXR images. However, this generalization should not be made because all types of DCNNs architectures have not been tested, and more than data is needed to avoid overfitting and develop a reliable system. An unbiased system can ignore any TB case or cause to apply TB treatment for any healthy individual. On the other hand, the study can encourage new research and practices to solve the issues arising from the lack of personnel and equipment for TB diagnosis. The research is also significant to reduce diagnosis time and propose an individualized treatment.

There are several limitations for the study: the data could be insufficient and may not be varied in terms of region, and age, and insufficient GPU power and time were available for a reliable DCNNs model. Hence, further experimental studies can be conducted with multi-classification or object detection on these datasets or tested on new data.

References

[1] World Health Organization, «Global tuberculosis report,» World Health Organization, Geneva, 2020.

[2] World Health Organization, «Global tuberculosis report,» World Health Organization, Geneva, 2022.

[3] T.C. Sağlık Bakanlığı, «Ulusal tüberküloz kontrol programı,» T.C. Sağlık Bakanlığı Yayın No: 1129, Ankara, 2022.

[4] T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü, «Tüberküloz tanı ve tedavi rehberi,» T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü, Ankara, 2019.

[5] E. J. Hwang, S. Park, K.-N. Jin, J. I. Kim, S. Y. Choi, J. H. Lee, J. M. Go, J. Aum, J.-J. Yim, J.-J. Yim ve C. M. Park, «Development and validation of a deep learning–based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs,» Clinical Infectious Diseases, cilt 65, no. 5, pp. 739-747, 2019.

[6] World Health Organization, «Chest Radiography in Tuberculosis Detection: Summary of current WHO recommendations and guidance on programmatic approaches,» World Health Organization, Geneva, 2016.

[7] R. A. Castellino, «Computer aided detection (CAD): an overview,» Cancer Imaging, pp. 17-19, 23 August 2005.

[8] K. Suzuki, «Computer-aided detection of lung cancer,» Image-based computer-assisted radiation theraphy, Singapore, Springer, 2017, pp. 9-40.

[9] J. Zhang, X. Yutong, Q. Wu ve Y. Xia, «Medical image classification using synergic deep learning,» Medical image analysis, cilt 54, pp. 10-19, 2019.

[10] D. W. v. S. T. I. M. E. Matheny, «Artificial intelligence in health care a report from the national academy of medicine,» JAMA,, cilt 323, no. 6, pp. 509-510, 2020.

[11] M. A. Musen, B. Middleton ve R. A. Greenes, «Clinical decision support system,» Biomedical informatics, Cham,, Springer, 2021, pp. 795-840.

[12] K. C. Laudon ve J. P. Laudon, Management information systems: managing the digital firm, London: Pearson, 2020.

[13] R. Sharda, D. Dursun ve E. Turban, Analytics, data science, & artificial intelligence: systems for decision support, Hoboken: Pearson, 2020.

[14] E. Alpaydın, «Introduction,» Introduction to machine learning, Cambrige, MIT Press, 2010, pp. 1-19.

[15] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker ve C. Mooney, «Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systemeatic review,» Applied Sciences, cilt 11, no. 11, 2021.

[16] A. K. Jain, J. Mao ve K. M. Mohiuddin, «Artificial neural networks: a tutorial,» Computer, cilt 29, no. 3, pp. 31-44, 1996.

[17] R. Yamashita, M. Nishio, R. K. G. Do ve K. Togashi, «Convolutional neural networks: an overview and application in radiology,» Insights into imaging, cilt 9, no. 4, pp. 611-629, 2018.

[18] S. Kulkarni ve S. Jha, «Artificial intelligence, radiology, and tuberculosis: a review,» Academic radiology, cilt 27, no. 1, pp. 71-75, 2020.

[19] R. H. Abiyev ve M. K. S. Ma'aitah, «Deep convolutional neural networks for chest diseases detection,» Journal of healthcare engineering, 2018.

[20] M. Mamalakis, A. J. Swift, B. Vorselaars, S. Ray, S. Weeks, W. Ding, R. H. Clayton, L. Mackenzie ve A. Banerjee, «DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays,» Computerized Medical Imaging and Graphics, no. 102008, p. 94, 2021.

[21] E. Ölmez, O. Er ve A. Hızıroğlu, «Deep learning in biomedical applications: detection of lung dise convolutional neural networks,» Deep learning in biomedical and health informatics, Boca Raton, CR 2021, pp. 97-115.

[22] P. Lakhani ve B. Sundaram, «Deep learning at chest radiography: automated classification of p tuberculosis by using convolutional neural networks,» Radiology, cilt 284, no. 2, pp. 574-582, 2017.

[23] B. K. Karaca, S. Güney, B. Dengiz ve M. Ağıldere, «Comparative Study for Tuberculosis Detection Deep Learning,» 2021 44th International Conference on Telecommunications and Signal Processir Brno, 2021.

[24] B. Oltu, S. Güney, B. Dengiz ve M. Ağıldere, «Automated tuberculosis detection using pre-trained an 2021 44th International Conference on Telecommunications and Signal Processing (TSP), Brno, 202

[25] Y. Cao, C. Liu, M. J. Brunette, N. Zhang, T. Sun, P. Zhang, J. Peinado, E. S. Garavito, L. L. Garcia Curioso, «Improving tuberculosis diagnostics using deep learning and mobile health technologie resource-poor and marginalized communities,» 2016 IEEE first international conference on connecte applications, systems and engineering technologies (CHASE), Washington, 2016.

[26] S. Stirenko, Y. Kochura, O. Alienin, O. Rokovyi, Y. Gordienko, P. Gang ve W. Zeng, «Chest X-ray a tuberculosis by deep learning with segmentation and augmentation,» 2018 IEEE 38th International Co on Electronics and Nanotechnology (ELNANO), Kyiv, 2019.

[27] T. Rahman, A. Khandakar, M. Abdulkadir, K. R. Islam ve K. F. Islam, «Reliable tuberculosis detecti chest X-ray with deep learning, segmentation and visualization,» IEEE Access, cilt 8, pp. 191586-1916

[28] S.-J. Heo, Y. Kim, S. Yun, S.-S. Lim, J. Kim, C.-M. Nam, E.-C. Park, I. Jung ve J.-H. Yoon, «Deep algorithms with demographic information help to detect tuberculosis in chest radiographs in annual health examination data,» International journal of environmental research and public health, cilt 16, 250, 2019.

[29] T. M. Navamani, «Efficient deep learning approaches for health informatics,» Deep learning and computing environment for bioengineering systems, St. Louis, Academic Press, 2019, pp. 123-137.

[30] M. J. Willemink, W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R. M. Sum L. Rubin ve M. P. Lungren, «Preparing medical imaging data for machine learning,» Radiology, cilt 2 pp. 4-15, 2020.