IMPROVING TUBERCULOSIS DETECTION IN CHEST X-RAY IMAGES THROUGH TRANSFER LEARNING AND DEEP LEARNING: A COMPARATIVE STUDY OF CNN ARCHITECTURES

Aishwarya Keshi[1], Aarti Valsang[2]

[1]Student, Computer Science and Engineering, VVPIET, Solapur, Maharashtra, India.

[2]Assistant Professor, Computer Science and Engineering, AGPIT, Solapur, Maharashtra, India.

ABSTRACT

Introduction: Tuberculosis remains a significant global health challenge, necessitating more efficient andaccurate diagnostic methods.

Methods: This study evaluates the performance of various convolutional neural network (CNN) architectures— VGG16, VGG19, ResNet50, ResNet101, ResNet152, and Inception-ResNet-V2—in classifying chest X-ray (CXR) images as either normal or TB-positive. The data set comprised 4,200 CXR images, with 700labeled as TB-positive and 3,500 as normal. We also examined the impact of data augmentation on model performance and analyzed the training times and the number of parameters for each architecture.

Results: Our results showed that VGG16 outperformed the other models across all evaluation metrics, achieving an accuracy of 99.4%, precision of 97.9%, recall of 98.6%, F1-score of 98.3%, and AUC-ROC of 98.25%. Surprisingly, data augmentation did not improve performance, suggesting that the original dataset's diversity was sufficient. Furthermore, models with large numbers of parameters, such as ResNet152 and Inception- ResNet-V2, required longer training times without yielding proportionally better performance.

Discussion: These findings highlight the importance of selecting the appropriate model architecture based on task-specific requirements. While more complex models with larger parameter counts may seem advantageous, they do not necessarily offer superior performance and often come with increased computational costs.

Conclusion: The study demonstrates the potential of simpler models such as VGG16 to effectively diagnose TB from CXR images, providing a balance between performance and computational efficiency. This insight can guide future research and practical implementations in medical image classification.

Keywords: Tuberculosis Detection, Chest X-ray Classification, Convolutional Neural Networks, Data Augmentation, Medical Image Analysis.

1. INTRODUCTION

Tuberculosis (TB) remains one of the leading infectious diseases worldwide, affecting an estimated one-third to one-fourth of the global population with the bacillus Mycobacterium tuberculosis, the causative agent of TB [1]. In 2019, it was estimated that over 10 million individuals globally contracted TB, yet only 71% were detected, diagnosed, and reported through various countries' national TB programs, leaving approximately 29% of cases unreported [2]. According to the World Health Organization's (WHO) 2023 TB report, TB was identified as the second most common cause of death among infectious diseases [3]. Furthermore, the global incidence rate of TB remains alarmingly high at approximately 133 new cases per 100,000 people annually [3].

This situation underscores the need for prompt, effective, and affordable screening and treatment strategies to meet the WHO's ambitious goals of reducing TB incidence by 80%, decreasing TB mortality by 90%, and eliminating catastrophic financial burdens on families affected by TB by 2030 [4].

The World Health Organization (WHO) advised member countries to proactively conduct TB screenings and detection, especially within the high-risk groups, taking into account their unique epidemic scenarios and financial levels [5]. While bacteriological tests, including sputum cultures, sputum smears, and molecular diagnostics, are considered the gold standard for identifying active TB cases, their applicability on a large scale, particularly among high-risk populations, is not feasible [6]. This limitation is due to the methods being resource-intensive, logistically challenging, and associated with prolonged turnaround times [7].

As a result, chest radiography has become the most prevalent method for early TB detection [8]. However, in countries with limited resources, which also bear the highest TB burden, the availability of chest radiography screenings remains inadequate, primarily due to a shortage of radiologists [6].

In recent years, significant advancements have been made in leveraging artificial intelligence, particularly through machine learning and deep learning techniques, for analyzing chest X-ray (CXR) images to differentiate between TB-positive and TB-negative images [9-15]. This innovation has enabled individuals without radiology expertise to conduct

TB screening tests, presenting a significant shift in diagnostic approaches. These technologies have shown promising results, to the extent of outperforming radiologists in the interpretation of CXR images [14, 15]. In this research, we investigate the effectiveness of different convolutional neural network (CNN) architectures in classifying tuberculosis in CXR images.

We compare and evaluate the performance of popular CNN models including ResNet, Inception, and VGG, and examine the impact of different hyperparameters on classification accuracy. To the best of our knowledge, this study is the first to utilize a larger and more diverse dataset and conduct a comprehensive comparison of the latest CNN architectures, including ResNet101, ResNet152, and Inception-V2, assessed across different parameters. The research aims to address the following questions

1. How does the choice of CNN architecture affect the classification performance?

2. What is the optimal hyperparameter configuration for each CNN architecture?

3. Can transfer learning be leveraged to improve classification accuracy?

4. How does incorporating data augmentation techniques impact the model's performance compared to training solely on real images?

The rest of the paper is organized as follows. In the next section, we present the literature review, which provides an overview of the current state of research in the field.

This is followed by the methodology section, where we describe the deep learning models used in this research along with the techniques for improving training time such as transfer learning. We also describe the data and the data and analysis procedures used in our study such as data augmentation to mitigate against imbalance. Next, we present the results of our analysis, including any findings.

Finally, we discuss the implications of our results, conclude with a summary of our main findings, and suggest areas for future research.

2. METHODOLOGY

In this section, we provide a comprehensive overview of the methodologies used in our study, including the dataset and preprocessing, data normalization, data augmentation, the application of transfer learning methods, the architecture of CNNs utilized, and the evaluation metrics adopted to assess the performance of the models.

2.1 Implementation Overview

The implementation framework illustrated in Figure1 starts with the acquisition of a well-defined dataset, followed by comprehensive data pre-processing, which includes data augmentation, resizing, normalization, and partitioning into training, validation, and test sets. Subsequently, we embark on the development of various deep learning models.

These models undergo extensive training and evaluation against different hyperparameters and evaluation metrics, to accurately predict and classify CXR images into positive or negative cases of TB.
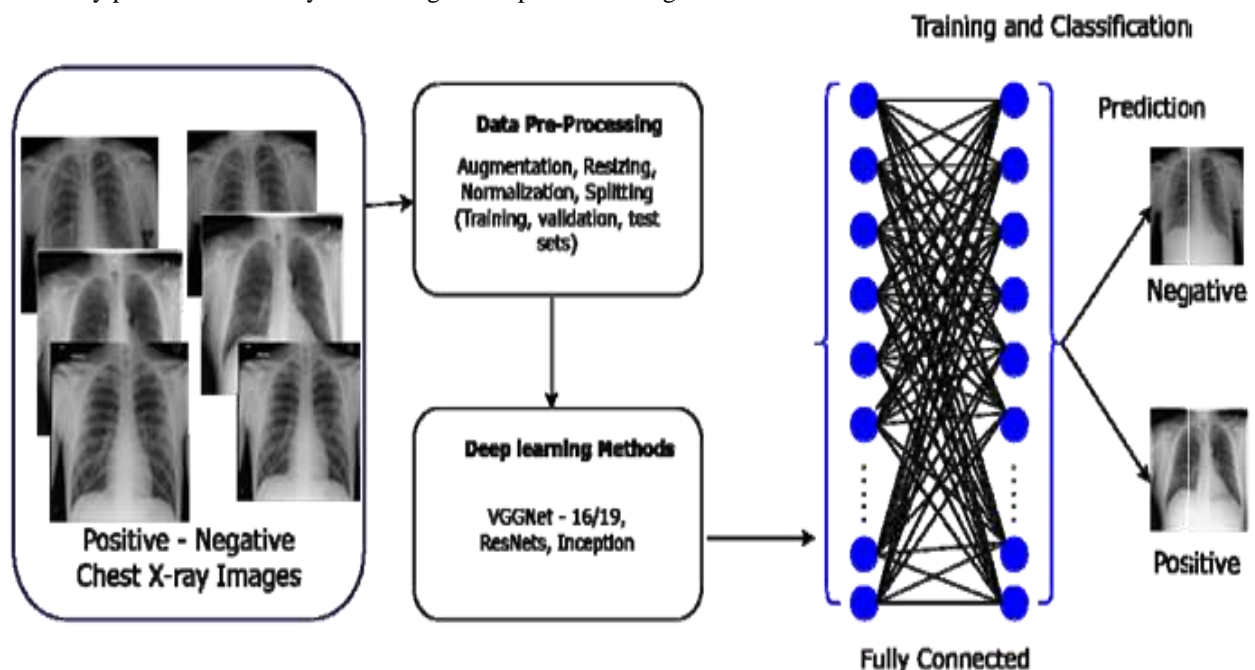

Figure 1: The implementation flow of the deep learning classification methodology

2.2 Dataset

The dataset utilized in this research comprises 4,200 CXR images sourced from a public Kaggle data repository. The data set was compiled through a collaborative effort between researchers from Qatar University (Doha, Qatar), the University of Dhaka (Bangladesh), and collaborators from Malaysia.

They worked closely with medical professionals from the Hamad Medical Corporation (Doha, Qatar) and various healthcare institutions in Bangladesh. The dataset consists of 700 CXRs images indicative of TB and 3,500 CXRs images classified as normal, with all images having a resolution of 512×512 pixels [32]. This composition provides a substantial foundation for evaluating the effectiveness of CNN models in the detection of TB from CXR images. Figure 2 presents some of the images from the dataset
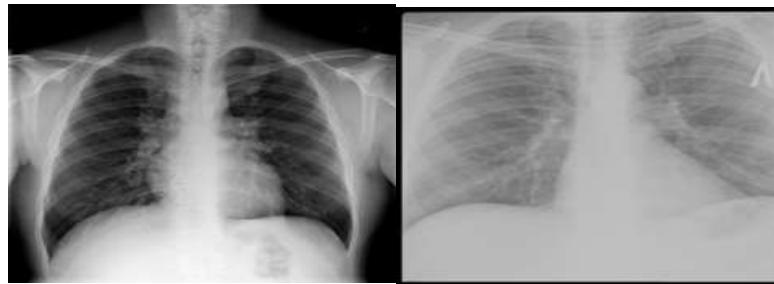
(a)   **(b)**

Figure2: The CXR sample images (a) TB-Negative (b) TB-positive

2.3 Pre-Processing

To optimize the performance and efficiency of our models,we implemented key pre-processing techniques, specifically data normalization and augmentation, prior to training the models.

Data Normalization: In the pre-processing stage of image analysis, normalization is a critical step to standardize the input data, facilitating the model's learning process. This study applies normalization to CXR images, which initially possess pixel intensity values in the range of 0 to 255, common for grayscale images [33]. The goal of normalization is to adjust these intensity values to a standardized scale that improves computational efficiency and model convergence during training. The normalization process is mathematically represented as follows:

$$I' = \frac{I - I_{min}}{I_{max} - I_{min}}$$

Where:

- I represent the original pixel intensity of the image,
- I_{min} and I_{max} are the minimum and maximum possible intensity values in the original image, rpectively, I is the normalized pixel intensity.

For gray scale images, $I_{min}=0$ and $I_{max}=255$.This equation effectively rescales the pixel intensity values to the range [0,1], making the input data more suitable for processing by the neural network layers. This normalization technique is advantageous because it ensures that each input parameter (pixel, in this case) contributes equally to the analysis, preventing features with initially larger ranges from dominating the learning process[34].It also helps to stabilize the gradient descent optimization algorithm by maintaining a consistent scale for all gradients [35].

Data Augmentation: Data Augmentation represents a powerful regularization strategy designed to artificially increase the dataset through label-preserving transformations, there by incorporating more invariant examples into the training set [36]. This approach, characterized by its computational efficiency, has been previously used to reduce overfitting when training CNNs, such as in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), where it contributed to achieving state-of-the-art results [37]. This method enhances the robustness and generalizability of deep learning models by exposing them to a wider array of variations, simulating real-world variability. In our study, to address the imbalance between TB-positive and TB-negative images and to introduce different variations, we randomly augmented 30% of the TB-positive images and 5% of TB-negative images.

The data augmentation techniques applied included random rotation within a range of 0 to 60 degrees, random width and height shifts of up to 0.2 times the image size, and random zooming of up to 0.2 times the original size, alongside horizontal and vertical flipping. To manage the newly created pixels from such transformations, a "fill mode" strategy was employed, ensuring integrity and consistency in the augmented images. These augmentations were performed using Keras's Image Data Generator, a comprehensive data augmentation suite [38]. Figure3 shows a sample of real images and their corresponding augmented outputs.
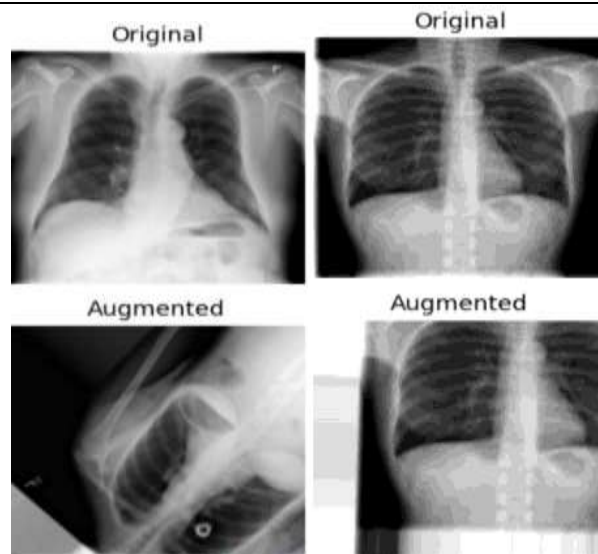
Figure 3: Sample of real and corresponding augmented images

2.4 Transfer Learning

Transfer Learning is a machine learning technique where a model developed for a specific task is repurposed as the starting point for a model on a second, related task[39].This technique leverages the knowledge gained during the initial training phase in one domain to enhance learning in another, potentially unrelated domain. It operates under the principle that information learned in one context can be exploited to accelerate or improve the optimization process in another, essentially allowing for the transfer of learned features and patterns across different but related problems [39]. This approach is particularly beneficial in situations where the dataset for the second task is too small to train a deep learning model from scratch or when computational resources are limited, offering a pathway to achieve high model performance with relatively less data and training time.

In this study, we propose an implementation that capitalizes on the transfer learning paradigm by utilizing pre- trained models such as Inception-v3, ResNet (50, 101, 152), and VGG (16, 19), which were initially trained on the Image Net dataset[37].This adaptation involves fine-tuning and customizing the models 'last layers to suit our classification task, effectively tailoring the robust, pre-learned representations of the ImageNet dataset to recognize and interpret the specific patterns and anomalies associated with TB in CXR images.

2.5 CNN Architectures

In the next subsections, we provide a brief description of the VGG and ResNet families of CNN architectures as well as the Inception ResNet architecture that are considered in this study.

VGGNet: Introduced by Simonyan and Zisserman from the University of Oxford's Visual Geometry Group in 2014, the VGGNet architecture marked a significant milestone in the field of deep learning [43]. Known for its outstanding performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) of that year, VGGNet is characterized by its use of 3×3 filters in all convolutional layers, simulating the effects of larger receptive fields. This architecture is available in two variants, VGG16 and VGG19, differing in depth and the number of layers, with VGG19 being the deeper model.

In our research, we utilized both the VGG16 and VGG19 architectures to train models on datasets consisting of solely real CXR images, and a combination of augmented and real images. This approach aimed to assess the impact of incorporating augmented images on the performance of these two architectures. Images were resized to 256×256 pixels before being input into the networks. We extended the architectures by adding a flattening layer, followed by a dense layer of 512 neurons with are lu activation function and a dropout layer with a dropout rate of 0.2 to mitigate overfitting. A softmax activation function was used in the output layer for binary classification. We employed the Adam optimizer with the binary cross entropy loss function for optimization. The training was conducted over 15 epochs with a batch size of 32 for both models. This rigorous approach ensured both architectures could classify between TB-positive and TB-negative CXR images accurately.

ResNet:[41] introduced the deep residual network (ResNet) architecture in their 2016 seminal paper. This architecture greatly improved the performance of deep neural networks and went on to win the COCO object detection challenge and the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). To date, several variants of the ResNet architecture exist, including ResNet50, ResNet101, and ResNet152, which vary in depth and number of layers. ResNet architectures are very deep models [41, 44]. The core idea behind ResNet is the use of residual connections, also known

as shortcuts, which bypass one or more layers. By resolving the vanishing gradient issue, these shortcuts maintain the gradient flow across the network and facilitate the training of much deeper networks [41].

The CXR images in this study were classified using the ResNet50, ResNet101, and ResNet152 architectures. We added three more layers to the ResNet50 model, two, each with 256 units and one with 512 units, using batch normalization and ReLU activation in each layer. To reduce overfitting, dropout layers were added with dropout rates of 0.3, 0.25, and 0.2 respectively. The binary cross-entropy loss function was used to compile the model, while the Adam optimizer was used to optimize the model at a learning rate of 0.001. Two units with a softmax activation function made up the output layer, which classified the images as either TB-positive or TB-negative. Training for this model involved 16 batch sizes and 100 epochs.

ResNet101 was trained using the same settings as ResNet50, as preliminary training showed that the same parameter values used for ResNet50, also yielded optimal results for the ResNet101 architecture. For ResNet152,a selective fine-tuning approach was adopted where, only the last 10 layers of the network were trainable, enhancing the model's focus on more feature-specific adjustments in the later stages of the network. This model shared the augmentation layers of ResNet50 but was trained for only 50 epochs, incorporating a learning rate scheduler, Reduce LR On Plateau, which adjusted the rate based on the validation loss with a factor of 0.1, patience of 5, and a minimum learning rate of 1×10^{-6}, thereby optimizing the training dynamics. The details of the models' configuration are shown in Table 1.

Hyperparameter	ResNet50	ResNet101	ResNet152
Number of Layers	53(50base+3extra)	104(101base+3extra)	155(152base+3extra)
Units per Layer	256,256,512	256,256,512	256,256,512
Activation	ReLU	ReLU	ReLU
Batch Normalization	Yes	Yes	Yes
Drop out Rate	0.3,0.25,0.2	0.3,0.25,0.2	0.3,0.25,0.2
Optimizer	Adam	Adam	Adam
Learning Rate	0.001	0.001	Variable(Reduce LR On Plateau)
Loss Function	Binary Cross entropy	Binary Cross entropy	Binary Cross entropy
Training Epochs	100	100	50
Batch Size	16	16	16

Inception-ResNet: The Inception networks, introduced by Szegedy et al. [45], have greatly advanced the field of CNN as they have achieved state-of-the-art performance in a number of computer vision problems [45-47]. The original Inception V1, also known as GoogleNet, was first introduced in 2014 and won the ILSVRC of that year. The architecture introduced a novel approach, of using multiple convolutional filter sizes in parallel, allowing the network to capture various spatial features of different scales with improved utilization of computing resources [45].

In this study, we used Inception-ResNet V2 architecture, a hybrid model that combines the benefits of both the Inception and residual networks. This hybrid approach enables the architecture to learn more complex features with improved training stability and faster convergence [45]. The Inception-ResNet V2 also leverages residual connections to skip certain layers during training, which helps it improve gradient flow, accelerate training times, and reduce the likelihood of vanishing gradient problems in deep networks [48]. We selected Inception-ResNet V2 due to its demonstrated state-of-the-art results in several medical imaging tasks [47].

For our implementation, the Inception-ResNet V2 architecture was initialized with weights pre-trained on the ImageNet dataset. Similar to our approach with the ResNet152 model, all layers except the last ten were frozen to retain the pre-trained features from ImageNet.

The last ten layers were set to be trainable, enabling the model to learn specific features from the CXR images. We added three new layers: two with 256 units each and one with512 units, all using ReLU activations and Batch Normalization. Each of these layers was followed by Dropout layers with rates of 0.4, 0.35, and 0.3, respectively, to introduce non-linearity and reduce overfitting.

The final output layer consisted of two units with a softmax activation function for binary classification. The model was then compiled using binary cross-entropy as the loss function and the Adam optimizer with a learning rate of 0.0001. Training was conducted for 50 epochs with a batch size of 16.

3. MODELING AND ANALYSIS

The performance of the CNN architectures in classifying CXR images into TB-positive and TB-negative categories was assessed using several standard performance metrics including: accuracy, precision, recall, F1-score, and the

Area under the Receiver Operating Characteristic curve (AUC-ROC). Each metric provides unique insights into the model's classification abilities, considering both the true and false predictions.

Accuracy: this metric measure the proportion of true positive and true negative results among the total number of cases examined:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP (True Positives): The number of TB-positive images that are correctly identified as TB-positive by the model
- TN (True Negatives): The number of TB-negative images that are correctly identified as TB-negative by the model
- FP (False Positives): The number of TB-negative images that are incorrectly identified as TB-positive by the model
- FN (False Negatives): The number of TB-positive images that are incorrectly identified as TB-negative by the model.

Precision: Also known as positive predictive value, precision is the ratio of correctly identified TB cases to all cases that were diagnosed as TB by the model. It measures the model's accuracy in diagnosing a patient with TB when the model predicts the disease. High precision indicates a low rate of false TB diagnoses. Mathematically, it is defined as:

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall, or sensitivity, is especially critical in medical diagnostics as it quantifies the model's ability to correctly identify all actual TB cases. It represents the proportion of actual TB cases that were correctly identified by the model and aims to minimize the risk of missing a true TB case. It is computed as:

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The F1-score is the harmonic mean of precision and recall, providing a single measure that balances both the false positives and false negatives. In TB diagnosis, it is particularly useful because it creates a balance between precision (minimizing false TB diagnoses) and recall (minimizing missed TB diagnoses), which is crucial for medical screening tests. It is defined as:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

AUC-ROC: The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) measures a model's ability to discern between positive and negative classes. In the context of our problem, that specifically refers to distinguishing between TB-positive and TB-negative CXR images. The AUC-ROC curve is a plot of the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The AUC-ROC provides an aggregated measure of the model's performance across all classification thresholds, with a value of 1 representing a perfect model and a value of 0.5 representing a model with no discriminatory power. The approximate area under the receiver operating characteristic curve (AUC-ROC) is calculated by using the following formula:

$$AUC \approx \sum_{i=1}^{n} \frac{(FPR_i - FPR_{i-1}) \times (TPR_i + TPR_{I-1})}{2}$$

Where i is the current data point or threshold, FPR_i and TPR_i are the false positive and true positive rates at the i-th threshold, respectively, and n is the number of data points or thresholds used to calculate the AUC-ROC. Each term in the sum represents the area of a trapezoid where $(FPR_i - FPR_i - 1)$ is the base of the trapezoid and $(TPR_i + TPR_i - 1)/2$ is the average height of the trapezoid. The formula calculates the AUC-ROC by summing the areas of trapezoids formed by connecting consecutive points on the AUC-ROC curve.

4. RESULTS AND DISCUSSION

The study aimed to analyze and compare the performance of various CNN architectures, including VGG16, VGG19, ResNet50, ResNet101, ResNet152, and Inception-ResNet-V2, in classifying CXR images as either TB-positive or TB-negative. Additionally, we also investigated whether data augmentation could further improve the classification performance of these models by comparing the performance of models trained on only real images versus those trained on a combination of real and augmented data. We went further to examine the training time and the number of parameters for each architecture to understand the computational efficiency and resource demands for each model. This analysis is important for practical implementation, particularly in resource-constrained settings where training time and computational costs are significant considerations. By evaluating these parameters, we aimed to identify models that not only perform well but also offer a balanced trade-off between accuracy and efficiency, making them suitable for real-world applications in diverse healthcare environments. Table2 presents the models' performance evaluated across accuracy, precision, recall, and F-score for models trained on real images and those trained on a combination of real and augmented data. Figure 4 shows the performance of these models when evaluated using the AUC-ROC score metric. It is observed that the VGG16 outperformed all other architectures across all metrics, with an accuracy of 99.4%, precision of 97.9%, recall of 98.6%, F1-Score of 98.3%, and AUC of 98.25%. Its performance was superior consistently irrespective of whether the models were trained with or without data augmentation.

Table 2: Evaluation of CNN architectures across key evaluation metrics: This table summarizes the performance of various CNN architectures according to precision, recall, and F1-Score. Models marked with an asterisk (*) were trained using a combination of real and augmented data, showcasing the impact of data augmentation on model performance.

Architecture	Accuracy	Precision	Recall	F1-Score
VGG16	99.4	97.9	98.6	98.3
* VGG16	99.3	96.6	99.3	97.9
VGG19	99.2	96.6	98.6	97.6
* VGG19	99.2	96.6	98.6	97.6
ResNet50	96.1	81.3	96.9	88.4
* ResNet50	89.0	97.5	30.0	45.9
ResNet101	96.9	94.8	84.6	89.3
* ResNet101	97.3	92.1	90.0	91.1
ResNet152	97.9	93.6	93.6	93.6
* ResNet152	97.5	87.6	96.6	92.1
InceptionResNet-v2	99.0	95.9	98.6	97.2
* InceptionResNet-v2	99.2	97.2	97.9	97.5

Surprisingly, increasing the dataset size through data augmentation did not correspond with an increase in the performanceofthemodelsacrossallarchitectures,asseeninTable2.Thiswasalsoobservedinothermodelssuch as Resnet50, where when augmented data was included, the AUC-ROC score dropped significantly from 85.65% to 63.75% as shown in Figure 4.
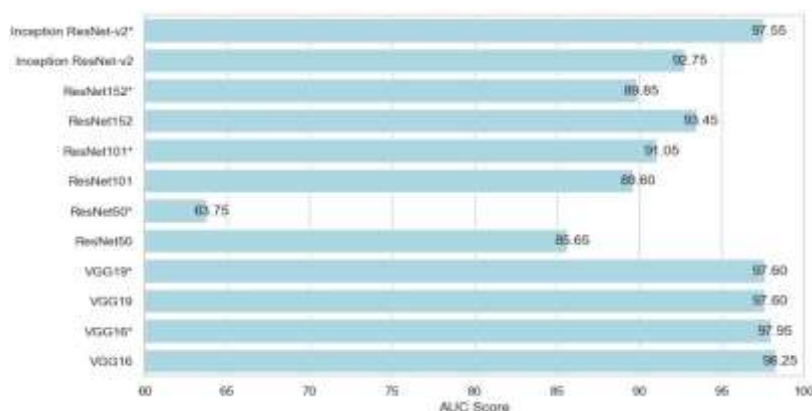

Figure 4: The models' are a under a curve scores

4.1 Training Time

We also tracked each model's training time with a combination of data augmentation and real images versus training with only real images, as shown in Figure 5. As expected, training with data augmentation requires more time due to the increased size of the dataset. For example, training the ResNet152 with data augmentation took

356.6 minutes whereas training without augmentation took 345.7 minutes. This observation highlights the trade- off between longer training times and the potential benefits of data augmentation. However, data augmentation did not improve performance in our case, indicating that the additional training time did not translate into better model generalization.
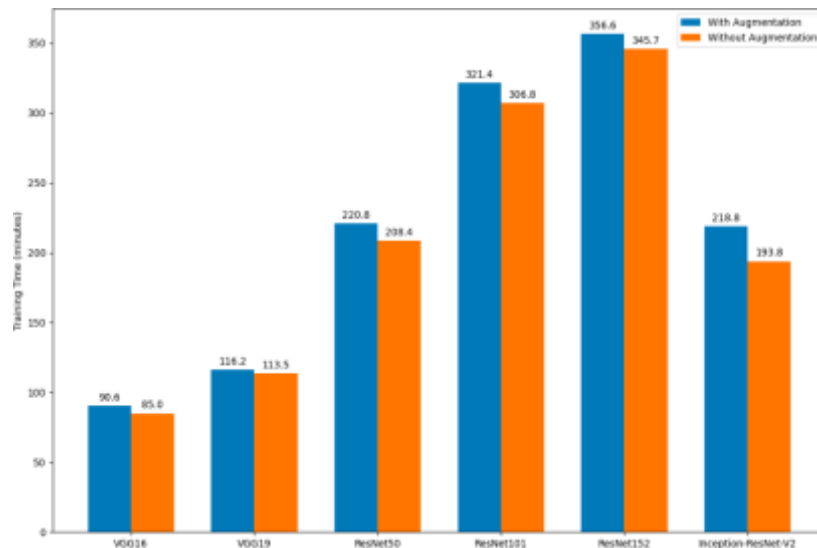

Figure 5: Training time for the models

4.2 Model Parameters

In addition to our analysis, We provide a detailed breakdown of the parameter count for each model used in our study, as shown in Figure6. The number of parameters in a model reflects its complexity and capacity to learn from data. Consequently, it has a direct impact on both training time and the computational resources required, influencing the model's overall efficiency and scalability.
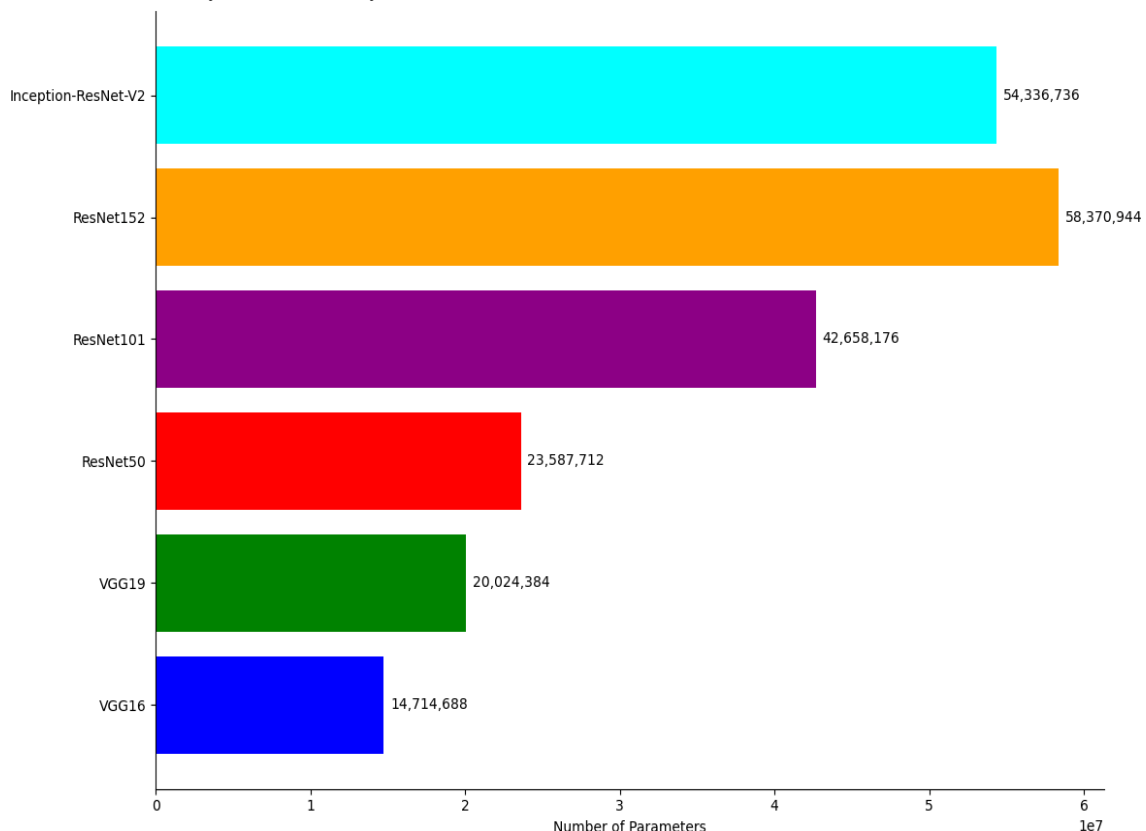

Figure 6: Parameters of each model

The findings from this study provide significant insights into the performance and efficiency of several CNN architectures in the classification of CXR images for TB detection. The architectures evaluated included VGG16, VGG19, ResNet50, ResNet101, ResNet152, and Inception-ResNet-V2. Of these, the VGG16 consistently achieved the highest performance across all metrics such as: accuracy, precision, recall, and F1-score. This consistent performance suggests that VGG16 effectively captures the necessary features for distinguishing between TB- positive and TB-negative CXR images, even with fewer parameters compared to the deeper models.

The findings also high light the fact that, while data augmentation is often used to improve the performance of CNN models by expanding the dataset and introducing variability, it does not necessarily lead to performance improvements if the base dataset already provides sufficient diversity for training. In our study, the original data set appeared robust enough, and the addition of augmented data did not enhance model performance. This suggests thatthebenefitsofdataaugmentationarecontext-dependentandmaynotalwaysbenecessaryoreffectivewhen the existing dataset is already well-suited for the task.

It is commonly observed in several studies that models with a higher number of parameters, such as ResNet152 and InceptionResNet-V2, are capable of capturing more deep patterns in the data [41, 45]. However, this comes at the cost of requiring more computational resources and longer training times. Interestingly, in our study, despite having fewer parameters, VGG16 outperformed the more complex models. This suggests that for our specific task of classifying CXR images into TB-positive and Tb-negative categories, VGG16 efficiently captured the relevant features without necessitating excessive complexity. This finding highlights the importance of selecting the appropriate model architecture based on the specific characteristics and requirements of the task at hand, rather than simply opting for the model with the most parameters. This result also aligns with the principle that simpler models can often perform competitively when they are well-matched to the data and the problem domain [43].

5. CONCLUSION

This study presents a comprehensive evaluation of several CNN architectures - VGG16, VGG19, ResNet50, ResNet101, ResNet152, and Inception-ResNet-V2 — in classifying CXR images as either TB-positive or TB-negative. The findings showed that the VGG16 architecture consistently outperformed the other models across all the evaluation metrics, achieving superior performance despite having fewer parameters compared to the more complex architectures such as ResNet152 and InceptionResNet-V2.

Our results also showed limited benefits of data augmentation in this context, suggesting that the original dataset provided sufficient diversity for effective training. This highlights the importance of context in the application of data augmentation techniques. Furthermore, the study demonstrated the significant trade-offs between model complexity, training time, and performance. Models with more parameters, such as ResNet152 and Inception- ResNet-V2, required longer training times and more computational resources without corresponding improvements in classification performance across all the evaluation metrics. This result continues to emphasize the need for careful selection of model architecture based on the specific characteristics and requirements of the task at hand. Overall, our research contributes to the growing body of evidence supporting the effectiveness of deep learning models in medical image classification and provides insights into optimizing these models for TB detection in CXR images.

6. REFERENCES

[1] Assefa Y, WoldeyohannesS, Gelaw YA, Hamada Y, and Getahun H, Screening tools to exclude active pulmonary TB in high TB burden countries: systematic review and meta-analysis, he international journal of tuberculosis and lung disease, pp. 728–734, 2019,vol. 23, no. 6.

[2] Chakaya Jeremiah, Khan Mishal, Ntoumi Francine, Aklillu Eleni, Fatima Razia, Mwaba Peter, Kapata, Nathan Mfinanga Sayoki, Hasnain Seyed Ehtesham, Katoto Patrick DMC and others, Global Tuberculosis Report 2020– Reflections on the Global TB burden, treatment and prevention efforts, International journal of infectious diseases, pp. S7–S12, 2021,vol. 113, Elsevier.

[3] World Health Organization, Global tuberculosis report2023,Geneva: World Health Organization, 2023,Licence: CC BY-NC-SA.3.0 IGO. https://www.who.int/teams/global-tuberculosis.

[4] Plekar Mukund, Weil Diana, Lonnroth Knut, Jaramillo Ernesto, Lienhardt Christian, Dias Hannah Monica, Falzon Dennis, Floyd Katherine, Gargioni Giuliano, Getahun Haileyesus, and others, WHO's new end TB strategy, The Lancet, pp. 1799–1801, 2015,vol. 385, Elsevier.

[5] World Health Organization, Systematic screening for active tuberculosis: an operational guide, Geneva: World Health Organization, https://apps.who.int/iris/.

[6] Liao Qinghua, Feng Huiying, Li Yuan, Lai Xiaoyu, Pan Junping, Zhou Fangjing, Zhou Lin, and Chen Liang, Evaluation of an artificial intelligence (AI) system to detect tuberculosis on chest X-ray at a pilot active screening

project in Guangdong, China in 2019, Journal of X-ray Science and Technology, pp. 221–230, 2022,vol. 30, no. 2, IOS Press.

[7] Van't Hoog Anna H, Meme Helen K, Laserson Kayla F, Agaya Janet A, Muchiri Benson G, Githui, Willie A, Odeny Lazarus O, Marston Barbara J, and Borgdorff Martien W, Screening strategies for tuberculosis prevalence surveys: the value of chest radiography and symptoms, PloS one, pp. e38691, 2012,vol. 7, no. 7, Public Library of Science San Francisco, USA.

[8] Pande Tripti, Pai Madhukar, Khan Faiz Ahmad, and Denkinger Claudia M, Use of chest radiography in the 22 highest tuberculosis burden countries, European Respiratory Journal, pp. 1816–1819, 2015,vol. 46, no. 6, Eur Respiratory Soc.

[9] Kant Sonaal, andSrivastavaMuktabhMayank,Towards automated tuberculosis detection using deep learning, 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1250–1253, 2018,IEEE.

[10] Hwang Sangheum, Kim Hyo-Eun, Jeong Jihoon, and Kim Hee-Jin, A novel approach for tuberculosis screening based on deep convolutional neural networks, Medical imaging 2016: computer-aided diagnosis, pp. 750–757, 2016,vol. 9785, SPIE.

[11] Kim Tae Kyung, Paul H Yi, Hager Gregory D, and Lin Cheng Ting, Refining dataset curation methods for deep learning-based automated tuberculosis screening, Journal of Thoracic Disease, pp. 5078, 2020,vol. 12, no. 9, AME Publications.

[12] Lakhani Paras and Sundaram Baskaran, Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks, Radiology, pp. 574–582, 2017, vol. 284, no. 2, Radiological Society of North America.

[13] Hooda Rahul, Sofat Sanjeev, Kaur Simranpreet, Mittal Ajay, and Meriaudeau Fabrice, Deep-learning: A potential method for tuberculosis detection using chest radiography, 2017 IEEE International Conference on signal and image processing applications (ICSIPA), pp. 497–502, 2017, IEEE.

[14] Ho Thi Kieu Khanh, Gwak Jeonghwan, Prakash Om, Song Jong-In, and Park Chang Min, Utilizing pre-trained deep learning models for automated pulmonary tuberculosis detection using chest radiography, Intelligent Information and Database Systems: 11th Asian Conference, ACIIDS 2019, Yogyakarta, Indonesia, April 8–11, 2019, Proceedings, Part II 11, pp. 395–403, 2019, Springer.

[15] Kazemzadeh Sahar, Yu Jin, Jamshy Shahar, Pilgrim Rory, Nabulsi Zaid, Chen Christina, Beladia Neeral, Lau Charles, McKinney Scott Mayer, Hughes Thad, and others, Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists, Radiology, pp. 124–137, 2023, vol. 306, no. 1, Radiological Society of North America.

[16] Zulvia Ferani E, Kuo Ren-Jieh, and Roflin Eddy, An initial screening method for tuberculosis diseases using a multi-objectivegradientevolution-basedsupportvectormachineandC5.0decisiontree,2017IEEE41stannual computer software and applications conference (COMPSAC), pp. 204–209, 2017, vol. 2, IEEE.

[17] Saybani Mahmoud Reza, Shamshirband Shahaboddin, Hormozi Shahram Golzari, Wah Teh Ying, Aghabozorgi Saeed, Pourhoseingholi Mohamad Amin, and Olariu Teodora, Diagnosing tuberculosis with a novel support vector machine-based artificial immune recognition system, Iranian Red Crescent medical journal, 2015, vol. 17, no. 4, Brieflands.

[18] HussainySyed Fahadulla, Zaffar Fareed, Zaffar Muhammad Adeel, Khaliq Aasia, Khan Imran H, Ahmad Raza, Decision-tree inspired classification algorithm to detect Tuberculosis(TB), 2017

[19] Mithra K Sand Emmanuel WR Sam, FHDT: fuzzy and Hyco-entropy-based decision tree classifier for tuberculosis diagnosis from sputum images, Sadhan¯ a¯, 2018, vol. 43, Springer.

[20] Ayas Selen and Ekinci Murat, Random forest-based tuberculosis bacteria classification in images of ZN-stained sputum smear samples, Signal, Image and Video Processing,pp. 49–61, 2014, vol. 8, Springer.

[21] Zheng Chi, Liu Jingxin, and Qiu Guoping, Tuberculosis bacteria detection based on Random Forest using fluorescent images, 2016 9thinternational congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI), pp. 553–558, 2016, IEEE.

[22] Rahman Muhammad, Cao Yongzhong, Sun Xiaobing, Li Bin, and Hao Yameng, Deep pre-trained networks as a feature extractor with XG Boost to detect tuberculosis from chest X-ray, Computers & Electrical Engineering, pp. 107252, 2021, vol. 93, Elsevier.

[23] Sebhatu Sıraj, Pooja, Nand Parmd, Intelligent System for Diagnosis of Pulmonary Tuberculosis Using XGBoosting Method, International Conference on Ubiquitous Computing and Intelligent Information Systems, pp.493–511,Springer.

[24] Kotei Evans, and Thirunavukarasu Ramkumar A Comprehensive Review on Advancement in Deep Learning Techniques for Automatic Detection of Tuberculosis from Chest X-ray Images, Archives of Computational Methods in Engineering,vol.31, pp.455–474, Springer.

[25] Yadav Ojasvi, Passi Kalpdrum, and Jain Chakresh Kumar Using deep learning to classify X-ray images of potential tuberculosis patients, 2018 IEEE International Conference on Bio informatics and Biomedicine (BIBM), pp.2368–2375, 2018, IEEE.

[26] Cao Yu, Liu Chang, Liu Benyuan, Brunette Maria J, Zhang Ning, Sun Tong, Zhang Peifeng, Peinado Jesus, Garavito Epifanio Sanchez, Garcia Leonid Lecca, and others Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities, 2016 IEEE first international conference on connected health: applications, systems and engineering technologies (CHASE), pp.274–281, 2016, IEEE.

[27] Panicker Rani Oomman, Kalmady Kaushik S, Rajan Jeny, and Sabu MK Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods, Biocybernetics and Biomedical Engineering, vol.38, pp.691–699, 2018, no.3, IEEE.

[28] Stirenko Sergii, Kochura Yuriy, Alienin Oleg, Rokovyi Oleksandr, Gordienko Yuri, Gang Peng, and Zeng WeiChest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation, 2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO), pp.422–428, 2018, IEEE.

[29] Nijiati Mayidili, Zhou Renbing, Damaola Miriguli, Hu Chuling, Li Li, Qian Baoxin, Abulizi Abudukeyoumujiang, Kaisaier Aihemaitijiang, Cai Chao, Li Hongjun, and others Deep learning based CT images automatic analysis model for active/non-active pulmonary tuberculosis differential diagnosis, Frontiers in Molecular Biosciences,vol.9, pp.1086047, 2022, Frontiers.

[30] Meraj Syeda Shaizadi, Yaakob Razali, Azman Azreen, Rum SN, Shahrel Azree, Nazri, Ahmad, and Zakaria Nor Fadhlina Detection of pulmonary tuberculosis manifestation in chest X-rays using different convolutional neural network (CNN) models, Int. J. Eng. Adv. Technol.(IJEAT), vol.9, no.1, pp.2270–2275, 2019

[31] Jaeger Stefan, Candemir Sema, Antani Sameer, Wang Y´ `1-Xiang J, Lu Pu-Xuan, and Thoma George´ Two public chest X-ray datasets for computer-aided screening of pulmonary diseases, Quantitative imaging in medicine and surgery, vol.4, no.6, pp.475, 2014, AME Publications

[32] Rahman Tawsifur, Khandakar Amith, Kadir Muhammad Abdul, Islam Khandaker Rejaul, Islam Khandakar F, Mazhar Rashid, Hamid Tahir, Islam Mohammad Tariqul, Kashem, Saad Mahbub, Zaid Bin and others Reliable tuberculosis detection using chestX-ray with deep learning, segmentation andvisualization,IeeeAccess,vol.8, pp.191586–191601, 2020, IEEE

[33] [33]Sachin Rajan, Sowmya V,Govind D, and Soman KP Dependency of various color and intensity planes on CNN based image classification, Advances in Signal Processing and Intelligent Recognition Systems: Proceedings of Third International Symposium on Signal Processing and Intelligent Recognition Systems(SIRS-2017),September13-16,2017, Manipal,India,pp.1167–177,2018, Springer

[34] Araujo Teresa, Aresta Guilherme, Castro Eduardo, Rouco Jos´ e, Aguiar Paulo, Eloy Catarina, Pol´onia Ant´onio, and Campilho Aur´elio´ Classification of breast cancer histology images using convolutional neural networks, PloS one, vol. 12, no.6, pp.e0177544, 2017, Public Library of Science SanFrancisco,CAUSA

[35] Santurkar Shibani, Tsipras Dimitris, Ilyas Andrew, and Madry Aleksander How does batch normalization help optimization?, Advances in neural information processing systems, vol. 31, 2018

[36] Taylor Luke and Nitschke Ge off Improving deep learning with generic data augmentation, 2018IEEE symposium series on computational intelligence (SSCI), pp.1542–1547, 2018, IEEE

[37] Russakovsky Olga, Deng Jia, Su Hao, Krause Jonathan, Satheesh Sanjeev, Ma Sean, Huang Zhiheng, Karpathy Andrej, Khosla Aditya,Bernstein Michael,and others Imagenet largescale visual recognition challenge, International journal of computer vision,vol.115,pp.211–252,2015,Springer

[38] Gulli Antonio and Pal Sujit Deep learning with Keras, 2017, Packt Publishing Ltd

[39] Hussain Mahbub, Bird Jordan J, and Faria Diego R A study on CNN transfer learning for image classification, Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK, pp.191–202, 2019, Springer

[40] LeCun Yann, Bottou Leon, Bengio Yoshua, and Haffner Patrick´ Gradient-based learning applied to document recognition, Proceedings of the IEEE,vol.86, no.11, pp.2278–2324, 1998, IEEE

[41] He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and patternrecognition,pp.770–778,2016

[42] Huang Gao, Liu Zhuang, Van Der Maaten Laurens and Weinberger, Kilian Q Densely connected convolutional networks, Proceedings of the IEEE conference on computer vision and pattern recognition, pp.4700–4708, 2017, IEEE

[43] Simonyan Karen and Zisserman Andrew Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014

[44] He Kaiming, Zhang Xiangyu, RenShaoqing, and Sun Jian Identity mappings in deep residual networks, Computer Vision –ECCV 2016:14th European Conference, Amsterdam, The Netherlands, October 11–14-2016, Proceedings,PartIV14,pp.630–645,2016,Springer

[45] Szegedy Christian, Liu Wei, Jia Yangqing, Sermanet Pierre, Reed Scott, Anguelov Dragomir, Erhan Dumitru, Vanhoucke Vincent, and Rabinovich Andrew Going Deeper With Convolutions, booktitle = Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015

[46] Szegedy Christian, Vanhoucke Vincent, Ioffe Sergey, Shlens Jon, Wojna Zbigniew Rethinking the Inception Architecture for Computer Vision, booktitle = Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

[47] Neshat Mehdi, Ahmed Muktar, Askari Hossein, Thilakaratne Menasha, and Mirjalili Seyedali Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception-ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs, Procedia Computer Science, vol.235, pp.1841–1850, 2024,Elsevier

[48] Szegedy Christian, Ioffe Sergey, Vanhoucke Vincent, and Alemi Alexander Inception-v4, inception-resnet and the impact of residual connections on learning, Proceedings of the AAAI conference on artificial intelligence, vol.31, no.1, 2017