Exploiting Cascaded Ensemble of Features for the Detection of Tuberculosis Using Chest Radiographs

A. F. M. SAIF[ID][1]**, (Member, IEEE), TAMJID IMTIAZ**[ID][1]**, (Member, IEEE),
CELIA SHAHNAZ**[1]**, (Senior Member, IEEE), WEI-PING ZHU**[ID][2]**, (Senior Member, IEEE),
AND M. OMAIR AHMAD**[ID][2]**, (Life Fellow, IEEE)**

[1]Department of Electrical and Electronic Engineering (EEE), Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh
[2]Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1M8, Canada

Corresponding author: M. Omair Ahmad (omair@ece.concordia.ca)

ABSTRACT Tuberculosis (TB) is a communicable disease that is one of the top 10 causes of death worldwide according to the World Health Organization. Hence, Early detection of Tuberculosis is an important task to save millions of lives from this life threatening disease. For diagnosing TB from chest X-Ray, different handcrafted features were utilized previously and they provided high accuracy even in a small dataset. However, at present, deep learning (DL) gains popularity in many computer vision tasks because of their better performance in comparison to the traditional manual feature extraction based machine learning approaches and Tuberculosis detection task is not an exception. Considering all these facts, a cascaded ensembling method is proposed that combines both the hand-engineered and the deep learning-based features for the Tuberculosis detection task. To make the proposed model more generalized, rotation-invariant augmentation techniques are introduced which is found very effective in this task. By using the proposed method, outstanding performance is achieved through extensive simulation on two benchmark datasets (99.7% and 98.4% accuracy on Shenzhen and Montgomery County datasets respectively) that verifies the effectiveness of the method.

INDEX TERMS Computer aided detection, cascaded ensembling, bhattacharya coefficient, geometrical separability index, rotation invariant augmentation techniques.

I. INTRODUCTION

According to the World Health Organization(WHO), in 2018, more than 1.5 million people died worldwide due to Tuberculosis(TB) [1], [2]. Though effective medications have been discovered by the researcher to treat TB patients, prior detection is required to step into the treatment process. The current gold standard definitive test for TB detection is the identification of Mycobacterium tuberculosis in clinical sputum or pus sample [3]. Additionally, other techniques like sputum smear microscopy in which bacteria in sputum samples are observed under a microscope, are also used for TB detection.

Nowadays, the chest radiograph has become more acceptable and widely used technique for the detection of Tuberculosis. Another imaging technique, Computed Tomography

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao[ID].

(CT) in the chest region, can provide significantly more diagnostic information compared to a chest radiograph but it is less popular among the low income people because of its high cost of diagnosis [4], [5]. Hence, more people will be benefited if detection of TB from chest radiographs is made automatic and more reliable.

To detect Tuberculosis(TB) from chest radiographs, different computer-aided methods have been introduced by the researchers and most of these methods have provided impressive results with information about abnormalities [6]–[8]. While using computer-aided methods to detect TB from chest radiograph an important step is feature extraction process and this process involves two alternating techniques—conventional hand-engineered and deep learning features. In Tuberculosis(TB) detection, different combinations of shape, edge, and texture descriptors have been used as hand-engineered features and these features have also been applied successfully to microscopy images of cells, for
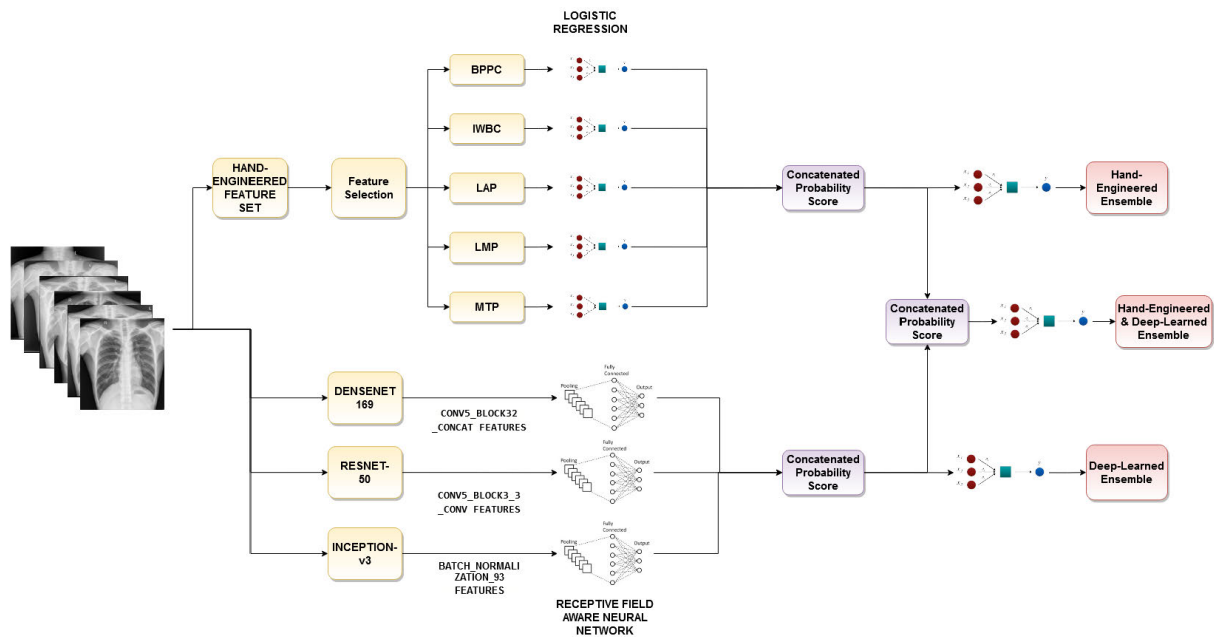
FIGURE 1. Flow chart involving major steps of the proposed method. Here first step is the feature extraction process and in the final step selected features are combined for final prediction.

which the cell cycle phase has been classified based on their appearances and patterns [9]–[14].

However, recently neural network-especially the convolutional neural network (CNN) has brought significant improvement in computer-aided image classification tasks and provided more significant results than conventional machine learning approaches [15], [16]. As a result, many researchers are now using deep learning features in abnormality detection in the chest region and getting better results compared to conventional machine learning approaches [17]–[19]. In the previous CNN based Tuberculosis detection approaches, the weights of different layers of multiple existing pre-trained models are fine-tuned from scratch and then features extracted from those layers are combined to create a deep learning ensemble model [14], [20]. The main advantage of using CNN is that it extracts discriminative features automatically considering the data characteristics. As conventional machine learning and deep learning-based features have some error profiles that can degrade the output result, an approach is introduced in this paper utilizing both hand-engineered features with the features extracted from the fine-tuning pre-trained deep convolutional network (DCN) model to predict TB in a chest x-ray. The main contribution of our paper is to employ a cascaded ensemble of Hand-Engineered features with Deep Learning features extracted from multiple pre-trained models and evaluate this ensemble technique as an alternative of ensembling different handcrafted features and deep learning features separately used previously. Besides, rotation invariant augmentation techniques are utilized on the training data in the pre-processing step, which is found very effective in increasing the performance of the proposed.

II. PROPOSED METHOD

All major steps involved in the proposed method are depicted in Fig.1. At first, Images are reshaped to uniform sizes (224 × 224) followed by min-max normalization and then rotation invariant augmentation techniques are used to augment the training samples. Next, These samples are passed through two distinct types of feature extraction procedures. The first one involved hand-engineered feature extraction process of training images. Next, By applying feature selection and reduction procedure, features are prepared for further aggregation process. The second feature extraction process employed three fine-tuned transfer learned deep convolution neural networks to extract deep learning-based features from the same training dataset. In order to get the probability scores, a logistic regression classifier is applied in the hand-engineered feature-set and a Receptive-Field-Aware neural network is trained using the features of selected layers of the aforementioned transfer learned models. A Receptive-Field-Aware neural network model can enlarge the receptive field of the deep learning models which in turn increase the learning ability, and enhance the discrimination ability of the deep learning-based features. Finally, the probability scores from both of the processing hands are combined to create the feature vector which is further utilized in the final classifier to get the desired ensemble model. The detailed description of each of the processes is presented in the following subsections.

A. ROTATION INVARIANT AUGMENTATION

Although both of the datasets used in this paper are two benchmark datasets and these chest x-ray images were collected over many years, the number of chest radiographs
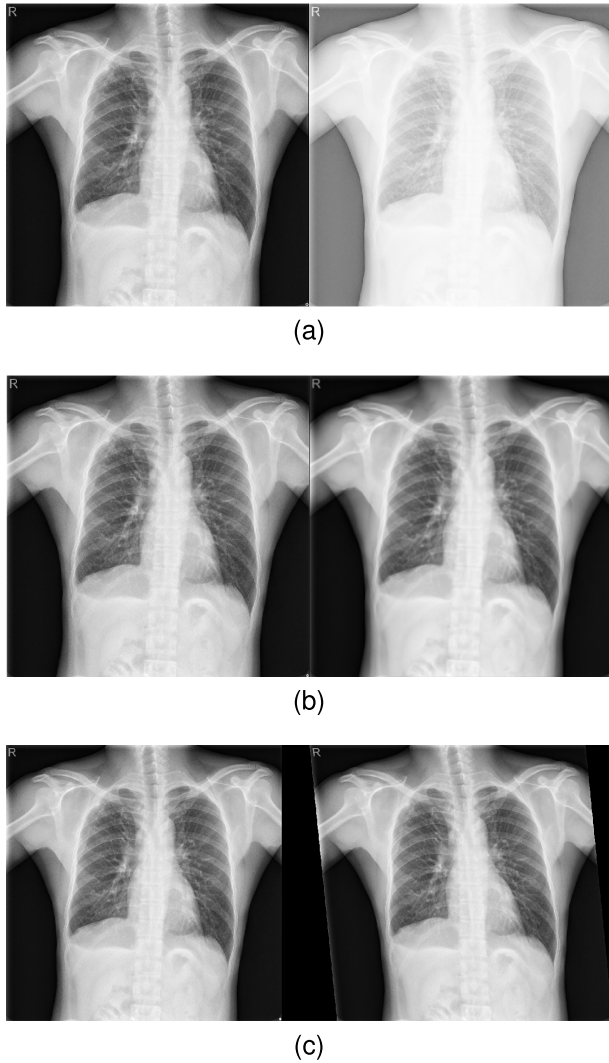
FIGURE 2. Rotation Invariant Augmentation(original image and the augmented image are shown in the left and right side respectively) (a) Elementwise Power (b) Gaussian Filtering (c) Shears.

available is limited. For this reason, data augmentation is required to increase the number of training samples and to achieve more generalizability. The most common technique usually used for the image augmentation purpose is rotation. However, rotation in chest radiographs can adversely affect the performance of subsequent automated processing steps in screening algorithms, such as lung segmentation and detection [40]. Considering the fact, we utilize three rotation invariant data augmentation techniques in the proposed method which is visualized in Fig. 2. Besides, a brief analysis about the effect of introducing rotation in the training samples is discussed in section III.

1) ELEMENT-WISE POWER
Each of the pixels in an image, I is elevated to a power, p which is defined by

$$p = f.r + 1 \qquad (1)$$

where, f is a random number taken from the gaussian distribution $X \hookrightarrow \mathcal{N}(\mu = 0, \sigma = 1)$ and r is a number less than unity. Finally the augmented image, I_a is generated by the following equation

$$I_a = sign(I) * \mid I \mid ^p \qquad (2)$$

2) GAUSSIAN FILTERING
A gaussian filter defined by the variance, σ between 0.3 and 0.9, is applied to each of the images in the training set. The radius of the kernel is chosen to be, $r = 4\,\sigma$.

3) SHEARS
Shearing is done on each images by using the following affine transformation, $\begin{pmatrix} 1 & 0 & 0 \\ s & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ where, s regulates the amount of shearing and is ranged between 0.2 and 0.4.

B. HANDCRAFTED FEATURE EXTRACTION
Handcrafted feature extraction is a conventional image processing method that uses various algorithms to collect information present in the image itself. Many image classification tasks depend on the local characteristics in images. In order to understand the characteristics of the local information, a large number of handcrafted features are extracted that are designed manually. Most of the time the selection of these features totally depends on the accuracy and computational efficiency. The extraction process of handcrafted features for Tuberculosis detection is discussed in this section.

1) HISTOGRAM OF ORIENTED GRADIENTS
Histogram of oriented gradients is a frequently used descriptor in medical diagnosis tasks [21]. In order to extract histogram of oriented gradients, the image is first divided into small connected regions which are called cells. Let, g_x and g_y are the gradients of an image $I(x,y)$ along the x and y direction respectively and θ_g is the angle of gradient. Both the gradient and the angle of gradient at each pixel are calculated as [22]:

$$g(i,j) = \sqrt{(g_x(i,j))^2 + (g_y(i,j))^2} \qquad (3)$$

$$\theta(i,j) = \frac{g_x(i,j)}{g_y(i,j)} \qquad (4)$$

Next, for each pixels within this region, histogram of gradient directions are calculated by dividing the angles into an equal distances which is call the histogram bins and each pixel inside the cell is voted to a histogram bin according to its angle of gradient. In this method, the histogram bin size is set to 8 and the cell size is selected as 16×16.

2) BINARY PATTERN OF PHASE CONGRUENCY (BPPC)
Binary pattern of phase congruency (BPPC) employed both Phase Congruency (PC) and Local Binary Pattern (LBP) approach to describe the image data [23]. Normally, Phase Congruency information enhances the textural content of an

image such as edges and lines and is a common descriptor used in the literature and is calculated as:

$$PC(x) = \frac{E(x)}{\sum_n A_n(x)} \qquad (5)$$

where, $A_n(x)$ is the amplitude of information of the wavelet transform at a given wavelet scale, n and $E(x)$ is the local energy. For Phase Congruency (PC), the number of filter orientations and the number of filter wavelength are considered as 6 and 3. On the other hand, for LBP, the number of neighboring pixels is chosen to be 8.

3) IMPROVED WEBER BINARY CODE(IWBC)
Improved Weber Binary Code (IWBC) is a feature map which is obtained by encoding each of the components of Improved Weber Local Descriptor (IWLD) features [24]. IWLD can capture the local variation of pixels in terms of magnitude and orientation and provides a feature map of an image. With the help of this feature map, three histograms are created in a sub-region ($M_r \times M_b$)of an image and then concatenated across different regions to represent the image. For our work, the size of the sub-region is selected as 8×8.

4) CURVATURE DESCRIPTOR HISTOGRAM (CDH)
Curvature Descriptor Histogram is a shape based feature [25]. For each of the pixel of an image I(x,y) it is defined as:

$$CD = \tan^{-1} \frac{\sqrt{\lambda_1{}^2 + \lambda_2{}^2}}{1 + I(x,y)} \qquad (6)$$

where λ_1 and λ_2 are the eigen values of the hessian matrix. A 8 bin histogram is plotted where each of the bin is a feature.

5) SHAPE DESCRIPTOR HISTOGRAM (SDH)
Shape descriptor histogram is used to describe the relationship among different key points of the image [26]. For an image I(x,y), shape descriptor, SD is calculated by

$$SD = \tan^{-1} \frac{\lambda_1}{\lambda_2} \qquad (7)$$

where λ_1 and λ_2 are the eigen values of the hessian matrix. A 8 bin histogram is plotted where each of the bin is a feature as in the case of curvature descriptor histogram.

6) LOCAL ARC PATTERN (LAP)
Local Arc Pattern (LAP) is used to find the gray level intensity differences of a certain pixel with respect to the neighboring pixels in all possible directions [27]. For a particular 5×5 block two binary pattern is created. The first pattern consists of 4 bit which is found by comparing the neighboring pixel of the center pixel of that block in a particular direction. The second pattern consists of 8 bit which is found in a similar manner. Finally a histogram is created and the histograms of two separate binary pattern is concatenated to extract the LAP features.

7) LOCAL TRANSITIONAL PATTERN (LTP)
Local transitional pattern consists of two basic steps. In the first step, a Gabor filter is applied to represent the textural content of the image [28]. Gabor filter is well known for its better modeling of the visual cortex of mammalian brains and capturing the textural representation [29]. In the 2D-domain, Gabor filter is defined as follows:

$$G(x, y, \theta, \sigma, \lambda, \psi, \gamma) = \frac{\exp(-(x'^2 + \gamma^2 y'^2)}{2\sigma^2}$$
$$\times \exp(i(\frac{2\pi x'}{\lambda} + \psi)) \qquad (8)$$

where σ, ψ, λ, γ are denoted as the filter size, the phase shift, the wavelength and the spatial aspect ratio of sinusoid respectively. Then, the Local transitional binary pattern is computed by comparing the neighboring pixels of different levels at different orientation.

8) MEDIAN BINARY PATTERN (MBP)
Median Binary Pattern(MBP) is very similar to the operation of Local Binary pattern [30]. At first, the median value of the local neighboring pixels around the center pixel of a block is calculated and then each of the pixels of that block is compared against the median value. For extracting the feature a 3×3 block is considered.

C. RATIONAL BEHIND THE PROPOSED HAND-ENGINEERED FEATURE SELECTION PROCESS
A fact also arises that all of these extracted features may not perform well in the Tuberculosis detection process. Hence, after the extraction of the features, the quality of hand-crafted features is evaluated in terms of class separability by the standard goodness of feature measures namely –Bhattacharyya Distance (BD) and Geometrical Separability Index (GSI). The similarity of two probability distributions is measured using Bhattacharyya Distance (BD) which is closely related to Bhattacharyya Coefficient (BC). BC gives the information about the overlapping of two distribution. BD is computed as [31],

$$BD = \frac{1}{8}(\mu_2 - \mu_1)^T [\frac{1}{2}(\delta_1 + \delta_2)]^{-1}(\mu_2 - \mu_1)$$
$$+ \frac{1}{2} \ln \frac{\det \frac{\delta_1 + \delta_2}{2}}{\sqrt{\det \delta_1} \times \sqrt{\det \delta_2}} \qquad (9)$$

Here δ_i and μ_i represent the covariance matrix and mean vector of i-th cluster. Bhattacharyya coefficient (BC) is computed as,

$$BC = \exp^{-BD} \qquad (10)$$

On the other hand, GSI measures the separability of classes in the sense of nearest neighbor and it is defined as [32],

$$GSI = \frac{\sum_{i=1}^{N}(f(x_i) + f(x_{iNN}) + 1) \mod 2}{N} \qquad (11)$$

Here x_i and x_{iNN} represents a data point and its nearest neighbor respectively. And $f(.)$ is the binary decision classification function which gives the class number of data points.
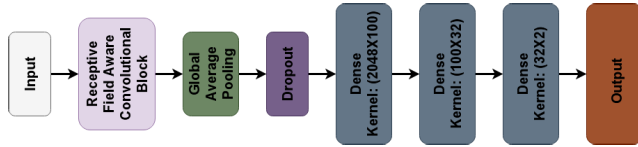
FIGURE 3. Receptive-Field-Aware Neural Network.

D. DEEP LEARNING BASED FEATURE EXTRACTION

Comparing to the conventional hand-engineered features based machine learning approaches, deep neural networks have recently achieved remarkably better performance in various detection and classification tasks. Most of the deep learning -based approaches applied in these tasks use transfer learning or deep residual network [33], [34]. Surprisingly in some cases, it outperforms the human detection capability also. In this paper, fine-tuned layers of pre-trained deep neural networks are used to extract the high-level features inherent in the data. Three deep convolution network models named Inception-V3, DenseNet-169, and Resnet-50 are utilized here to extract the deep learned features [34], [38], [39]. These deep convolution networks are different from each other in the perspective of architecture and the number of convolution layer. These neural network architectures expect 224×224 image size as input, hence the input images are reshaped to 224×224 image size. To fine-tune those pre-trained neural networks and extract features from a particular layer of a model, all the layers up to the desired layer have been frozen and then added a binary classifier layer on top to classify the image. Receptive-Field-Aware Neural Network Classifier is then used to evaluate the features collected from different layers of each of the model. Convolutional neural networks (CNN) are devised such that layer l_i feeds from layer l_{i-1} and then layer l_i feeds its outputs to layer l_{i+1}. Hence, whatever layer l_i learns, is a composition of features from the initial layers to layer l_i. Because of this property, deeper layers of a CNN model can cover an extremely larger effective area over the input image which makes them capable to extract more complex features than the earlier layers which learn basic features like edges and shapes. Hence the features of last layers of DCN can be used to train a classifier for classification task. And hence, for chest X-ray classification task, last few layers of DCN are used to train a shallow neural network. All the model have been implemented in Keras and the parameters, θ_i of the ith layer at time t are updated with Adam optimizer with learning rate α.

The extracted deep learning features and their performance in this particular task are described in Section-III. The notation used for the name of the layer of these models are derived from the notation used in keras library. Features having higher accuracy are selected here for further ensembling process with the hand-engineered features. Based on the accuracy, features of layer conv5_block3_out of densenet 50, batch_normalization_93 of inceptionV3 and conv5_block32_concat of densenet169 are selected for ensembling process.
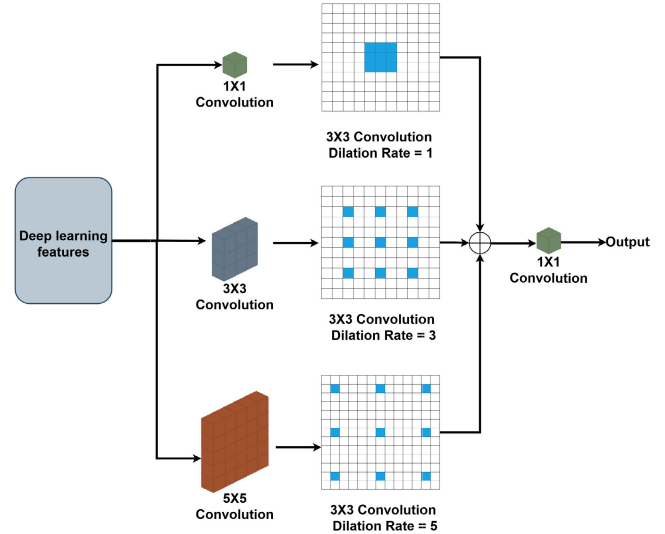

FIGURE 4. Receptive Field Aware Convolutional Block.

E. CLASSIFIER SELECTION

To choose an appropriate classification model for a dataset is a difficult work because of the applicability of the classifier on the particular data, higher feature dimensions, and the computational cost of that classification algorithm. To choose the appropriate classification model for this case, different popular models are examined, and their performance is evaluated further. Among all the classification model, Logistic Regression (LR) [35] is chosen because of its capability to perform better in low dimensionality space and binary classification problem. Moreover, it has low computational complexity as well. In this method, each of the chest x-ray images are classified using Logistic Regression into two classes based on the handcrafted and deep learning features extracted in section II-B and section II-C.

On the other hand, for deep learning features, a Receptive-Field-Aware Neural Network Classifier is used as shown in Fig.3. This neural network classifier consists of one receptive field aware block (shown in Fig.4), one Global Average Pooling (GAP) layer, one Dropout layer, and three Dense layers. Receptive field aware block is consists of three parallel convolutional blocks and these convolutional blocks have kernel size of 1×1, 3×3, and 5×5 respectively. Each of these convolutional block is followed by another convolutional block which has kernel size of 3×3 but dilation rate of 1, 3, and 5 respectively. At the end, these convolutional blocks are concatenated which is then followed by a convolutional block of 1×1 kernel size. GAP and Dropout layers help to reduce the overfitting of the model and interpret the feature maps as categories confidence maps. This type of architecture utilises the multi branch pooling of varying kernels that strengthens the deep learned features [44]. On top of that, dilated convolution controls their eccentricities and generate a good final representation. Since *dropoutrate* > 0.2 for input data highly affects the training model by adding a penalty to the loss function, a dropout rate of 0.2 is used. The three consecutive dense
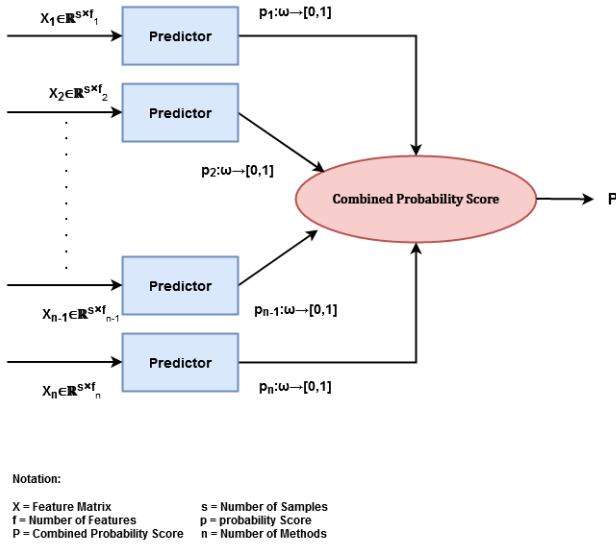
Notation:

X = Feature Matrix   s = Number of Samples
f = Number of Features   p = probability Score
P = Combined Probability Score   n = Number of Methods

FIGURE 5. Cascaded Ensembling Approach.

layers help to reduce underfitting of the network and increase the probability of extracting global relationship between the features. The activation functions used in the first dense layer and last dense layer are Rectified Linear Unit (ReLu) and Sigmoid Activation Function respectively.

F. CASCADED ENSEMBLE OF SELECTED FEATURES

Cascaded ensemble is a technique where outputs from multiple learning algorithms are considered and combined sequentially to make the final prediction. In this paper, instead of concatenating multiple features, we proposed a cascaded ensemble of hand-engineered features along with different deep learned features to make multiple ensemble models and utilizing them to generate the final predictive model. The reason for selecting this approach is to find a better predictive model than any single one. Referred to Fig.5, at first, each logistic regression classifier produces probability scores matrix of

$$p : \omega \rightarrow [0, 1]; \quad \sum_{n \in \omega} p(n) = 1$$

where ω is the sample space and in our case $\omega = \{normal, abnormal\}$. Hence, the dimension of the probability score matrix will be two times the number of samples and they are represented by p_{BPPC}, p_{IWBC}, p_{LAP}, p_{LMP}, and p_{MTP} for the hand-engineered feature-sets. Then, the probability scores from selected Hand-engineered features are combined by the following equation:

$$P = \frac{\prod_{i=1}^{n} p_i^{w_i}}{\prod_{i=1}^{n} p_i^{w_i} + \prod_{i=1}^{n} (1 - p_i)^{w_i}} \quad (12)$$

Here, n is the number of methods used, w_i is a weighting factor and its value is considered to be $\frac{1}{n}$ when no prior information exists about class probabilities, and p_i is the

scores obtained from i^{th} method. This equation produces the final probability score of hand engineered features, P_{HE} which is then considered as the input of a Logistic Regression Classifier to produce the final "Hand-Engineered Ensemble" model.

Similarly, from the other hand, each shallow neural network produces similar predictive scores matrix and they are represented by $p_{DENSE-NET}$, p_{RESNET}, and $p_{INCEPTION}$. These probability scores from deep learned features are also combined using (12) to produce final predicted score P_{DL}. This predicted score is then used as an input to Logistic Regression Classifier to obtain the "Deep Learning Ensemble" model.

Finally, the combined scores P_{HE} and P_{DL} are once again combined using (12) and the final predicted score P_{Final} is obtained. It is then classified through Logistic Regression classifier to produce the final ensemble model.

III. RESULTS AND ANALYSIS

In this section, a detailed overview of simulations in different stages of our proposed method are described. Details of the data-set, Feature Quality Analysis, performance of the hand-engineered, deep learned and proposed cascaded ensemble model are discussed in the following subsections.

A. DATASET

The proposed method is evaluated on two standard chest X-ray datasets. The first one is Montgomery County (MC) dataset which is made publicly available by Department of Health and Human Services, Montgomery County, Maryland, USA [9]. It comprises of 138 frontal chest X-ray in which 80 of them are normal and the rest are with manifestation of tuberculosis. Each of the image in this set has a resolution of either 4,020 × 4,892 or 4,892 × 4,020 pixels.

The second one that is used for the evaluation purpose is Shenzen dataset which is a standard digital image database for Tuberculosis and is created by the National Library of Medicine, Maryland, USA in collaboration with Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China [12]. It has total 662 cases in which 336 cases are with manifestation of tuberculosis and the rest are normal. These X-rays are of variable pixel size but all of them have approximately 3000 × 3000 pixel size. This dataset also provides a clinical reading of each of the CXRs. The data were captured within one month period as a daily routine at Shenzhen Hospital using Phylips DR digital diagnost system. A typical example of a chest x-ray image of a normal patient and a patient suffered from tuberculosis is depicted in Fig.6.

In order to improve the diagnosis capabilities of the transfer learnt models used in the proposed method, pretraining the models serves a very important role. It is worth mentioning that most of these models are pretrained with Imagenet dataset comprising of different natural images. However, for our objective, it is necessary to pretrain the models with that kind of images which is very similar to the nature of chest X-ray images. Hence, a third dataset is utilized in
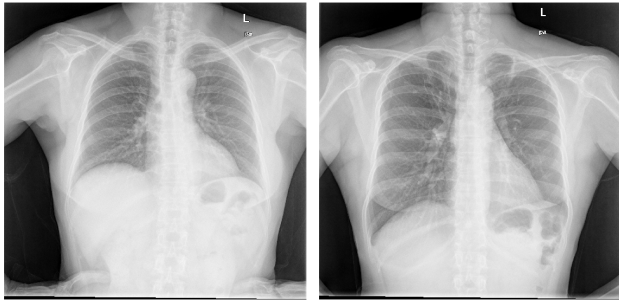
(a) A normal chest x-ray from from Shenzhen Dataset. (b) A chest x-ray from Shenzhen Dataset with manifestation of Tuberculosis.

FIGURE 6. Chest x-ray images collected from Shenzen Hospital dataset(a,b).

TABLE 1. Feature quality of different hand-engineered features in terms of BC and GSI evaluated on the Shenzhen dataset.

HE features	HOG	BPPC	IWBC	CDH	SDH	LAP	LTP	MBP
BC	0.68	0.260	0.26	0.62	0.77	0.24	0.33	0.27
GSI	0.28	0.77	0.84	0.34	0.32	0.71	0.74	0.79

our method only for the purpose of pretraining the models. This dataset is collected from the NIH Chest X-ray dataset [41] which comprises of 108948 X-ray images for 8 thorax abnormalities.

B. FEATURE QUALITY ANALYSIS

As stated earlier, two standard measures–Bhattacharya Coefficient(BC) and Geometric Separability Index(GSI)–are utilized to select the optimum feature set from the feature pool. For any feature with a higher BC value indicates that it has less capability to separate the data into clusters. Hence, it is desired to choose a feature with a lower BC value. Opposite scenario is seen in the GSI case where a lower GSI value indicates that it has less capability to separate the data into clusters. For our preliminary selected hand-engineered features, the computed BC and GSI are shown in Table.1. From the table, it is found that Histogram of Oriented Gradients(HOG), Curvature Descriptor Histogram(CDH), and Shape Descriptor Histogram(SDH) have the highest BC and lowest GSI value and hence these three features are excluded from the selection process.

C. EXPERIMENTAL SETUP

All the experiments of this study have been implemented the google cloud platform with NVIDIA P-100 GPU as the hardware accelerator with 5 and 10 fold cross validation schemes. In the intermediate layer of the transfer learnt neural network classifier, a L2 regularizer with a factor of 0.01 is imposed on both kernel and bias terms to learn sparse features.

TABLE 2. Perormance of Hand Engineered features on two standard TB datasets.

Dataset	Evaluated Metrics	Models					
		BPPC	IWBC	LAP	LMP	MTP	HE Ensemble
MC	Accuracy(%)	90.4	88.0	85.4	88.1	86.1	92.8
	AUC(%)	90.1	89.0	86.5	89.9	87	94.4
	Sensitivity(%)	91.0	88.8	86.1	89.8	88.2	94.0
	Specificity(%)	90.9	89.2	86.9	88.3	88.9	94.8
Shenzhen	Accuracy(%)	90.9	89.2	87.9	89.3	87.9	94.8
	AUC(%)	92.2	89.4	88.2	88.9	90.7	95.8
	Sensitivity(%)	91.2	90.1	89.8	89.4	90.3	95.2
	Specificity(%)	90.9	89.2	88.9	89.8	91.9	94.9

The Adam optimizer is employed with an initial learning rate of 0.001 which is decayed at a rate of 0.99 after every 5 epochs. To evaluate the performance, Four standard evaluation metrics have been used in this method and they are accuracy, sensitivity, specificity and area under the ROC curve (AUC). The accuracy, sensitivity and specificity are defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (14)$$

$$Specificity = \frac{TN}{TN + FP} \quad (15)$$

respectively. Where,

- TP = True Positive,
- TN = True Negative,
- FP = False Positive,
- FN = False Negative.

On the other hand, AUC is a measurement of the performance of a classification model which expresses the capability of a model to distinguish between the classes. If $x_1, x_2, x_3, \ldots, x_m$ be the output of a classifier on the positive examples and $y_1, y_2, y_3, \ldots, y_n$ be the output of that classifier on the negative examples, then AUC can be calculated by the following equation [27]

$$AUC = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (S(x_i, y_j))}{mn} \quad (16)$$

where

$$S(x_i, y_j) = \begin{cases} 1, & \text{if } x_i > y_j \\ \frac{1}{2}, & \text{if } x_i = y_j \\ 0, & \text{if } x_i < y_j \end{cases}$$

The higher the AUC, the better the model can predict.

D. EXPERIMENTAL RESULTS
1) PERFORMANCE ANALYSIS OF HAND-ENGINEERED FEATURES AND THEIR ENSEMBLE

Generally, performance of hand-engineered features totally depends on the quality of that particular feature and on the task it is applied to. The performance of the individual hand-crafted features used in this method is shown in Table.2. From the Table, a significant amount of improvement is found

TABLE 3. Accuracy and AUC of different features extracted from different layers of fine-tuned DCN using Shenzhen Dataset.

	Layer	batch_normalization_93	mixed9_1	concatenate_1	activation_93	mixed10
InceptionV3	Accuracy	**93.8%**	91.8%	84.7%	90.4%	89.3%
	AUC	**95.2%**	94.3%	89.7%	91.5%	92.1%
	Sensitivity	**94.5%**	92.8%	90.2%	91.5%	90%
	Specificity	**94.8%**	94.2%	89.6%	90.8%	90.3%
	Layer	conv5_block32_1_relu	conv5_block32_2_conv	conv5_block32_concat	bn	relu
Densenet169	Accuracy	84.9%	82.5%	**94.7%**	90.5%	91.3%
	AUC	91.2%	87.4%	**96.5%**	94.6%	94.7%
	Sensitivity	88.9%	84.2%	**95.2%**	89.8%	92.9%
	Specificity	88.2%	84%	**95.6%**	91.7%	92.4%
	Layer	conv5_block3_2_relu	conv5_block3_3_conv	conv5_block3_3_bn	conv5_block3_add	conv5_block3_out
Resnet-50	Accuracy	85.6%	84.2%	85.7%	90.4%	**93.5%**
	AUC	88.2%	86.4%	87.5%	93.45%	**95.1%**
	Sensitivity	88.6%	84.3%	85.7%	91.4%	**94.4%**
	Specificity	87.7%	84.9%	89.6%	91.6%	**95.7%**

TABLE 4. Evaluation of the performance of DL ensemble in association with pre-training and Augmentation.

Dataset	Models	Evaluated Metrics	Without Pretraining	With Pretraining	
				Without Augmentation	With Augmentation
MC	DL Ensemble	Accuracy(%)	93.2	94.8	96.5
		AUC(%)	92.4	95.3	97.2
		Sensitivity(%)	93.4	94.8	96.3
		Specificity(%)	92.8	95.3	97
Shenzhen	DL Ensemble	Accuracy(%)	93.6	95.4	98
		AUC(%)	94.5	96.2	97.9
		Sensitivity(%)	94.5	95.4	97.5
		Specificity(%)	94.7	95.9	98.4

in both of the datasets where ensemble of features is used. for the MC Dataset,the ensemble of HE provides a 2.65, 4.77, 3.2 and 4.3 percent improvement in Accuracy, AUC, Sensitivity and Specificity respectively in comparison with the best performance found from an individual HE model. Same scenario is found from the Shenzhen dataset as well.

2) PERFORMANCE ANALYSIS OF DEEP LEARNING FEATURES AND THEIR ENSEMBLE

It is already mentioned in section II-C that deep learning features have achieved significant performance in most of the disease detection tasks in different diagnostic modalities. To evaluate the performance of chest abnormality detection, we have employed three fine-tuned deep convolution model. A deep convolutional network (DCN) learns features automatically by looking at pixels in context and every different layers different type of features. So each layer of a DCN extracts unique characteristic features which helps to categorize an image. Hence, to evaluate the performance of each fine-tuned layer, a receptive field aware neural network is used. In order to select the best deep learning features for our proposed method, several features extracted from the fine-tuned layers of three selected DCNs are individually feed to the Receptive-Field-Aware Neural Network classifier. The results obtained from fine tuning models have been demonstrated in Table.3.

In Table.3, performance of the individual transfer learnt neural network features are summarized. The features of conv5_block3_out extracted from fine-tuned Resnet50 architecture, features of conv5_block32_concat of transfer learnt DenseNet169 and features of batch_normalization_93 of Inception-V3 provide prominent results on Shenzen dataset. Similarlym, for Montgomery County(MC) dataset features extracted from same layers provide noticable classification performance.

From Table.4, it is found that the ensemble of the selected features, extracted from the transfer learnt models provides an improvement when it is associated with pretraining and our proposed augmentation techniques. Pretraining of the models with chest X-ray data provides 2.05, 3.05, 1.94 and 3.04 percent improvement in Accuracy, AUC, Sensitivity, and Specificity respectively, and when it is combined with augmentation techniques, 4.1, 5.22, 3.7 and 5.21 percent improvement in Accuracy, AUC, Sensitivity, and Specificity is found respectively.

3) PERFORMANCE ANALYSIS OF CASCADED ENSEMBLE OF HAND-ENGINEERED AND DEEP LEARNING FEATURES

Finally, We have evaluated our proposed model with respect to the Hand-engineered(HE) ensemble and Deep Learning(DL) ensemble model. A perceptible amount of
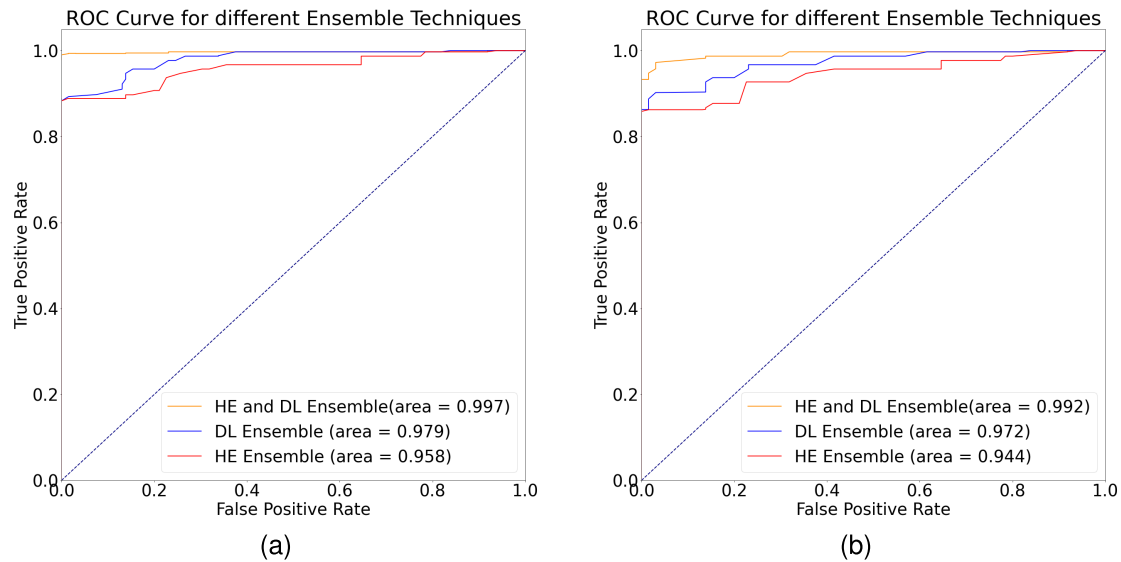
FIGURE 7. ROC curve for Hand-Engineered, Deep Learning and HE & DL ensemble model in (a) Shenzhen, (b) MC Dataset.
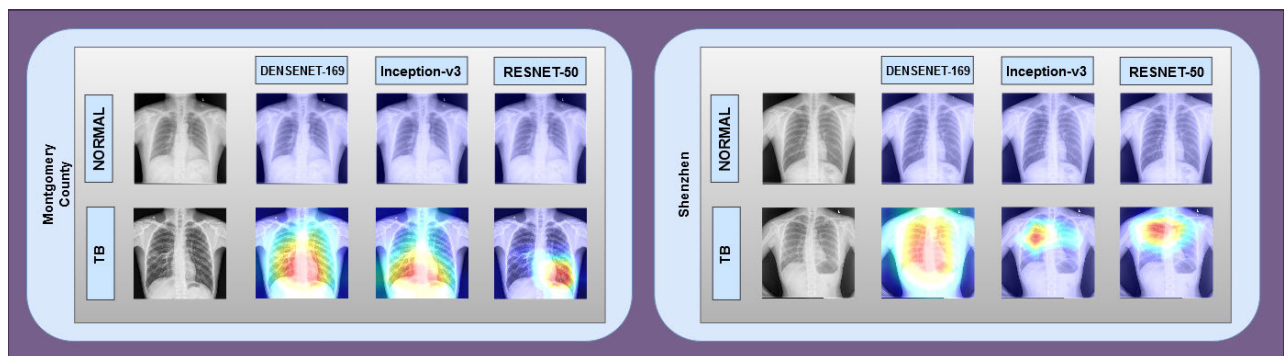

FIGURE 8. Grad-CAM: class activation visualization of three selected neural network layers on two datasets.

TABLE 5. Classification performance analysis of different classifier in Shenzhen Dataset using HE and DL ensemble technique.

Performance Parameter	Logistic Regression	SVM	ANN	KNN	Random Decision Forest
Accuracy(%)	**99.7**	96.0	94.5	92.7	91.9
AUC(%)	**99.7**	94.3	93.9	93.1	92.7
Sensiticity(%)	**99.6**	94.1	94.3	92.9	92.1
Specificity(%)	**99.6**	93.8	93.5	93.5	92.6

TABLE 6. Performance Comparison of different ensemble techniques.

| Dataset	Evaluated Metrics	Models		
		HE Ensemble	DL Ensemble	HE and DL Ensemble
MC	Accuracy(%)	92.8	96.5	98.4
	AUC(%)	94.4	97.2	99.2
	Sensitivity(%)	94.0	96.3	99.1
	Specificity(%)	94.8	97	99.4
Shenzhen	Accuracy	94.8	98	99.7
	AUC	95.8	97.9	99.7
	Sensitivity	95.2	97.5	99.6
	Specificity	94.9	98.4	99.6

improvement has been found by applying the proposed HE and DL ensemble model. The AUC has increased from the highest AUC of 0.974 found in the ensembled DL model to 0.995 as shown in Fig.7(a). Hence, a 2.15 percent higher AUC is found compared to the DL ensemble model in Shenzhen Dataset. On the other hand, a 2.17 percent improvement in AUC is found in the Montgomery County dataset which is shown in Fig.7(b). The performance of all ensemble models are recapitulated in Table.6. It is clear that our proposed handcrafted and deep learning feature fusion is appeared to be very effective in the tuberculosis detection task. To find

the performance of different classification scheme in terms of accuracy, AUC, Sensitivity, and Specificity, different classifiers are evaluated in HE and DL ensemble model and their performance is visualized in Table.5. It is found from the table that Logistic Regression provides best classification performance than the other classifiers.

Since cross validation techniques are used to evaluate the performance of the proposed ensemble models, box and whisker plots of their accuracies in two different datasets

TABLE 7. Performance comparison of different methods of Tuberculosis detection Task.

Methods	Features	Dataset	Cross-Validation	Accuracy(%)	AUC(%)	Specificity(%)	Sensitivity(%)
Jaeger et. al.(2017) [36]	Hand Engineered: Frangi, HOG, LBP, GIST	Shenzhen	-	0.90	0.949	-	-
Govindarajan et. al.(2019) [14]	Hand Engineered: BoF, SURF	MC	10 Fold	87.8	94.0	85.9	87.7
Chandra et. al.(2020) [37]	Hand Engineered: geometrical and first order statistical features	MC	10 Fold	95.6	95	-	-
		Shenzhen	10 Fold	99.4	99.0	-	-
Abideen et. al.(2020) [42]	Convolutional Neural Network	MC	5 Fold	96.42	-	-	-
		Shenzhen	5 Fold	86.46	-	-	-
Pasa et. al.(2019) [43]	Deep Neural Network	MC	5 Fold	79	81.1	-	-
		Shenzhen	5 Fold	84.4	90.0	-	-
Our Proposed Method	Hand Engineered & Deep Neural Network	MC	5 Fold	**97**	**98.1**	**98.1**	**97.3**
			10 Fold	**98.4**	**92.2**	**99.4**	**99.1**
		Shenzhen	5 Fold	**97.8**	**98.1**	**98.5**	**98.2**
			10 Fold	**99.7**	**99.7**	**99.6**	**99.5**
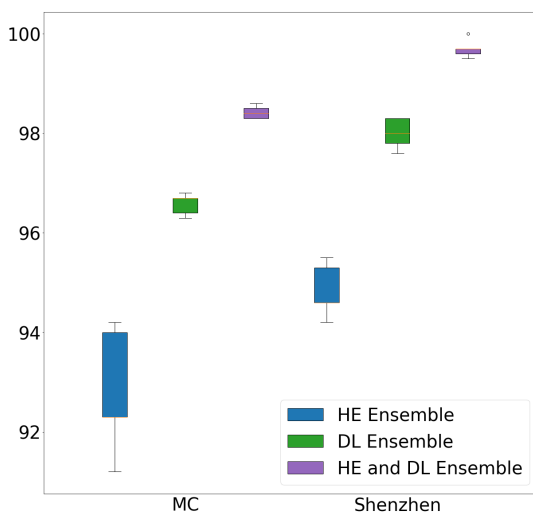
FIGURE 9. Accuracy of the three proposed ensemble techniques.

are shown in Fig.9. It is found from the plot that HE and DL ensemble technique has a low standard deviation in case of accuracies among different folds of the two datasets in comparison to the other models. In case of Shenzhen dataset, the standard deviation among the accuracies found using HE and DL ensemble technique is 0.187 whereas it is 0.30 and 0.54 for the other two models. The similar scenario is observed in the MC dataset.

Finally, in order to localize the particular area of interest in the images that provides valuable information in decision making process of the models, Gradient Based Class Activation Mapping(Grad-CAM) of two random samples are provided in Fig.8. By visualizing the Grad-CAM of the samples taken from three transfer learnt network which are working as the backend of the proposed receptive field aware neural network, it can be concluded that the region of interests are consistent among the networks.

E. PERFORMANCE COMPARISON BETWEEN DIFFERENT STATE-OF-THE-ART APPROACHES

The performance evaluation of previously proposed Tuberculosis classification model applied in Montgomery County (MC) and Shenzhen dataset is depicted in Table.7. Previously, Jaeger *et al.*, Govindarajan and Swaminathan, and Chandra *et al.* used Hand-Engineered features based approach in this particlar problem [14], [36], [37]. Abedin *et al.* and Pasa *et al.* exploited deep learning approach to handle the Tuberculosis classification problem [42], [43] and found significant performance.

Comparing all the models with our proposed one, it is found from the Table.7 that our HE and DL ensemble model performs significantly better than the other models on both of the datasets. We have used different cross-validation techniques to compare our results with recent methods. Govindarajan and Swaminathan [14], Chandra *et al.* [37] used 10 fold cross-validation in Montgomery County(MC) dataset and got 87.8%, and 95.6% accuracy respectively. On the other hand, Abideen *et al.* [42], Pasa *et al.* [43] used 5 fold cross-validation on MC dataset and got 96.42%, and 79% accuracy respectively whereas our proposed method scored 97%, and 98.4% accuracy on MC dataset when 5 fold, and 10 fold cross-validation were used respectively. In Shenzhen dataset, Chandra *et al.* [37] used 10 fold cross-validation whereas Abideen *et al.* [42], Pasa *et al.* [43] used 5 fold cross-validation. They acquired 99.4%, 86.46%, and 84.4% accuracy respectively. When 5 fold, and 10 fold cross-validation were used on Shenzhen dataset our proposed method scored 97.8%, and 99.7% accuracy respectively.

IV. CONCLUSION

Tuberculosis(TB) is the second-largest cause of patient morbidity and mortality, and so early detection of TB is very crucial to prevent death. For diagnosing TB, the most common imaging tool is a chest x-ray. However, as the number of patients is increasing day by day, the computer aid diagnosis system is very advantageous for rapid detection of TB. Machine learning and deep learning both are very obliging in this regard. As both of them have some advantages of their own, in this paper, we proposed a method that ensemble hand-engineered features with deep learning features to overcome error probability. From our investigation, we have found that our proposed method provides better accuracy than employing deep learning or machine learning individually

for Tuberculosis detection purposes. To get a more generalized outcome, rotation-invariant augmentation technique is applied which makes the model a more generalized one. After doing extensive simulation, a high level of accuracy is found on two standard datasets which eventually proves the performance of the proposed method.

REFERENCES

[1] *Global Tuberculosis Report*, World Health Org., Geneva, Switzerland, 2019.

[2] *Global Tuberculosis Control*, World Health Org., Geneva, Switzerland, 2018.

[3] *The Global Plan to Stop TB 2011–2015*, Stop TB Partnership, World Health Org., Geneva, Switzerland, 2011.

[4] C. Archer, A. R. Levy, and M. McGregor, "Value of routine preoperative chest X-rays: A meta-analysis," *Can. J. Anaesthesia*, vol. 40, no. 11, pp. 1022–1027, Nov. 1993.

[5] M. Traub, M. Stevenson, S. McEvoy, G. Briggs, S. K. Lo, S. Leibman, and T. Joseph, "The use of chest computed tomography versus chest X-ray in patients with major blunt trauma," *Injury*, vol. 38, no. 1, pp. 43–47, Jan. 2007.

[6] J. Melendez, B. van Ginneken, P. Maduskar, R. H. H. M. Philipsen, K. Reither, M. Breuninger, I. M. O. Adetifa, R. Maane, H. Ayles, and C. I. Sanchez, "A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 179–192, Jan. 2015.

[7] A. Kumar, Y.-Y. Wang, K.-C. Liu, I.-C. Tsai, C.-C. Huang, and N. Hung, "Distinguishing normal and pulmonary edema chest X-ray using Gabor filter and SVM," in *Proc. IEEE Int. Symp. Bioelectron. Bioinf. (IEEE ISBB)*, Apr. 2014, pp. 1–4.

[8] S. K. Wajid, A. Hussain, K. Huang, and W. Boulila, "Lung cancer detection using local energy-based shape histogram (LESH) feature extraction and cognitive machine learning techniques," in *Proc. IEEE 15th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCICC)*, Palo Alto, CA, USA, Aug. 2016, pp. 359–366.

[9] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 4978–4981.

[10] S. Jaeger, K. Palaniappan, C. S. Casas-Delucchi, and M. C. Cardoso, "Dual channel colocalization for cell cycle analysis using 3D confocal microscopy," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2580–2583.

[11] S. Jaeger, K. Palaniappan, C. S. Casas-Delucchi, and M. C. Cardoso, "Classification of cell cycle phases in 3D confocal microscopy using PCNA and chromocenter features," in *Proc. 7th Indian Conf. Comput. Vis., Graph. Image Process. (ICVGIP)*, 2010, pp. 412–418.

[12] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y.-X. Wang, P.-X. Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 233–245, Feb. 2014.

[13] L. Hogeweg, C. I. Sánchez, P. Maduskar, R. Philipsen, A. Story, R. Dawson, G. Theron, K. Dheda, L. Peters-Bax, and B. Van Ginneken, "Automatic detection of tuberculosis in chest radiographs using a combination of textural, focal, and shape abnormality analysis," *IEEE Trans. Med. Imag.*, vol. 34, no. 12, pp. 2429–2442, Dec. 2015.

[14] S. Govindarajan and R. Swaminathan, "Analysis of tuberculosis in chest radiographs for computerized diagnosis using bag of keypoint features," *J. Med. Syst.*, vol. 43, no. 4, p. 87, Apr. 2019.

[15] J. A. Dunnmon, D. Yi, C. P. Langlotz, C. Ré, D. L. Rubin, and M. P. Lungren, "Assessment of convolutional neural networks for automated classification of chest radiographs," *Radiology*, vol. 290, no. 2, pp. 537–544, Feb. 2019.

[16] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Aug. 2017.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[18] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest X-ray classification," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Dec. 2019.

[19] S. Xu, H. Wu, and R. Bie, "CXNet-m1: Anomaly detection on chest X-rays with image-based deep learning," *IEEE Access*, vol. 7, pp. 4466–4477, 2018.

[20] R. Hooda, A. A. Mittal, and S. Sofat, "Automated TB classification using ensemble of deep architectures," *Multimedia Tools Appl.*, vol. 78, no. 22, pp. 31532–331515, Nov. 2019.

[21] S. Bakheet, "An SVM framework for malignant melanoma detection based on optimized HOG features," *Computation*, vol. 5, no. 4, p. 4, Jan. 2017.

[22] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *Proc. Int. Workshop Automat. Face Gesture Recognit.*, vol. 12, 1995, pp. 296–301.

[23] S. Shojaeilangari, W. Y. Yau, J. Li, and E. K. Teoh, "Feature extraction through binary pattern of phase congruency for facial expression recognition," in *Proc. 12th Int. Conf. Control Automat. Robot. Vis. (ICARCV)*, Dec. 2012, pp. 166–170.

[24] B. Yang, T. Zhang, C. Gu, and X. Guan, "A novel face recognition method based on IWLD and IWBC," *Multimedia Tools Appl.*, vol. 75, no. 12, pp. 6979–7002, 2016.

[25] P. Fischer and T. Brox, "Image descriptors based on curvature histograms," in *Proc. German Conf. Pattern Recognit.*, Cham, Switzerland, 2014, pp. 239–249.

[26] S. Gu, P. Koehl, J. Hass, and N. Amenta, "Surface-histogram: A new shape descriptor for protein-protein docking," *Proteins, Struct., Function, Bioinf.*, vol. 80, no. 1, pp. 221–238, Jan. 2012.

[27] M. S. Islam and S. Auwatanamongkol, "Facial expression recognition using local arc pattern," *Trends Appl. Sci. Res.*, vol. 9, no. 2, p. 113, 2014.

[28] T. Ahsan, T. Jabid, and U.-P. Chong, "Facial expression recognition using local transitional pattern on Gabor filtered facial images," *IETE Tech. Rev.*, vol. 30, no. 1, pp. 47–52, Sep. 2013.

[29] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 2, no. 7, pp. 1160–1169, 1985.

[30] A. Hafiane, G. Seetharaman, and B. Zavidovique, "Median binary pattern for textures classification," in *Proc. Int. Conf. Image Anal. Recognit.* Berlin, Germany: Springer, Aug. 2007, pp. 387–398.

[31] A. Bhattacharrya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.

[32] J. Greene, "Feature subset selection using thornton's separability index and its applicability to a number of sparse proximity-based classifiers," in *Proc. Annu. Symp. Pattern Recognit. Assoc. South Africa*, 2001, pp. 1–5.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[35] D. W. Hosmer, Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, vol. 398. Hoboken, NJ, USA: Wiley, 2013.

[36] M. Ding, S. Antani, S. Jaeger, Z. Xue, S. Candemir, M. Kohli, and G. Thoma, "Local-global classifier fusion for screening chest radiographs," *Proc. SPIE*, vol. 10138, Mar. 2017, Art. no. 101380.

[37] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, and S. S. Netam, "Automatic detection of tuberculosis related abnormalities in chest X-ray images using hierarchical feature extraction scheme," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113514.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[39] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[40] K. C. Santosh, S. Candemir, S. Jaeger, A. Karargyris, S. Antani, G. R. Thoma, and L. Folio, "Automatically detecting rotation in chest radiographs using principal rib-orientation measure for quality control," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 29, no. 2, Mar. 2015, Art. no. 1557001.

[41] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[42] Z. U. Abideen, M. Ghafoor, K. Munir, M. Saqib, A. Ullah, T. Zia, S. A. Tariq, G. Ahmed, and A. Zahra, "Uncertainty assisted robust tuberculosis identification with Bayesian convolutional neural networks," *IEEE Access*, vol. 8, pp. 22812–22825, 2020.

[43] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019.

[44] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 385–400.

WEI-PING ZHU (Senior Member, IEEE) received the B.E. and M.E. degrees in electrical engineering from Nanjing University of Posts and Telecommunications, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1991. He was a Postdoctoral Fellow, from 1991 to 1992, and a Research Associate, from 1996 to 1998, with the Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada. From 1993 to 1996, he was an Associate Professor with the Department of Information Engineering, Nanjing University of Posts and Telecommunications. From 1998 to 2001, he worked with Hi-Tech Companies, Ottawa, Canada, including with Nortel Networks and SR Telecom Inc. Since July 2001, he has been a full-time Faculty Member with the Department of Electrical and Computer Engineering, Concordia University, where he is currently a Full Professor. His research interests include digital signal processing fundamentals, speech and statistical signal processing, signal processing for wireless communication with a particular focus on MIMO systems, and cooperative communication. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: FUNDAMENTAL THEORY AND APPLICATIONS, from 2001 to 2003, *Circuits, Systems, and Signal Processing*, from 2006 to 2009, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: TRANSACTIONS BRIEFS, from 2011 to 2015.

A. F. M. SAIF (Member, IEEE) received the B.Sc. degree from the Department of Electrical and Electronic Engineering (EEE), Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2018. He is currently doing his research under the supervision of Dr. Celia Shahnaz as a Professor with the Department of EEE, BUET. His research interests include machine learning and deep learning, bio-medical signal processing, and image processing.

TAMJID IMTIAZ (Member, IEEE) received the B.Sc. degree from the Department of Electrical and Electronic Engineering (EEE), Bangladesh University of Engineering and Technology (BUET), where he is currently pursuing the M.Sc. degree. He is also working as a Lecturer with the Department of EEE, Primeasia University. His research interests include biomedical signal processing, image processing, machine learning, deep learning, and computer vision. He is a former Treasurer of the IEEE Signal Processing Society (Bangladessh Chapter) and a Current Associate Member of the Bangladesh Computer Society.

M. OMAIR AHMAD (Life Fellow, IEEE) received the B.Eng. degree in electrical engineering from Sir George Williams University, Montreal, QC, Canada, and the Ph.D. degree in electrical engineering from Concordia University, Montreal. From 1978 to 1979, he was a Faculty Member with New York University College, Buffalo, NY, USA. In 1979, he joined the Faculty of Concordia University, as an Assistant Professor in computer science. Subsequently, he joined the Department of Electrical and Computer Engineering, Concordia University, where he was the Chair of the Department, from 2002 to 2005, and is currently a Professor. He was a Founding Researcher of Micronet, a Canadian Network of Centers of Excellence, from 1990 to 2004. He is also the Concordia University Research Chair (Tier I) in multimedia signal processing. He has authored in the area of signal processing and holds four patents. His current research interests include the areas of image and speech processing, biomedical signal processing, watermarking, biometrics, video signal processing and object detection and tracking, deep learning techniques in signal processing, and fast signal transforms and algorithms. In 1988, he was a member of the Admission and Advancement Committee of the IEEE. He was a recipient of numerous honors and awards, including the Wighton Fellowship from the SandfordFleming Foundation, an induction to Provosts Circle of Distinction for Career Achievements, and the Award of Excellence in Doctoral Supervision from the Faculty of Engineering and Computer Science, Concordia University.

CELIA SHAHNAZ (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Concordia University, Montreal, QC, Canada, in 2000 and 2002, respectively, and the Ph.D. degree in electrical and computer engineering, in 2009. She is currently working as a Professor with the Department of Electrical and Electronic Engineering (EEE), BUET. She is a fellow of IEB and has published more than 120 international journals and conference papers. She was a recipient of the Canadian Commonwealth Scholarship and Fellowship for pursuing the Ph.D. study, Canada, in 2004. Her research interests include the areas of speech analysis, speech enhancement, multimodal emotion recognition, digital watermarking, audio-visual recognition for biometric security, multimedia communication, control systems, robotics, signal processing, pattern recognition, and machine learling for audio, video, biomedical, and power signals. She is a member of the Technical Committee, Image, Video, and Multimedia (IVM) and Asia Pacific Signal and Information Processing Association (APSIPA).

● ● ●