Uncertainty Assisted Robust Tuberculosis Identification With Bayesian Convolutional Neural Networks

ZAIN UL ABIDEEN[1], **MUBEEN GHAFOOR**[1,2], **KAMRAN MUNIR**[2], **MADEEHA SAQIB**[3], **ATA ULLAH**[4], **TEHSEEN ZIA**[1], **SYED ALI TARIQ**[1], **GHUFRAN AHMED**[5], **AND ASMA ZAHRA**[1]

[1]Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad 44000, Pakistan
[2]FET - Computer Science and Creative Technologies, University of the West of England, Bristol BS16 1QY, U.K.
[3]Department of Computer Information Systems, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam 34212, Saudi Arabia
[4]Department of Computer Science, National University of Modern Languages (NUML), Islamabad 44000, Pakistan
[5]Department of Computer Science, National University of Computer and Emerging Sciences (NUCES), Karachi 54700, Pakistan

Corresponding author: Ata Ullah (aullah@numl.edu.pk)

ABSTRACT Tuberculosis (TB) is an infectious disease that can lead towards death if left untreated. TB detection involves extraction of complex TB manifestation features such as lung cavity, air space consolidation, endobronchial spread, and pleural effusions from chest x-rays (CXRs). Deep learning based approach named convolutional neural network (CNN) has the ability to learn complex features from CXR images. The main problem is that CNN does not consider uncertainty to classify CXRs using softmax layer. It lacks in presenting the true probability of CXRs by differentiating confusing cases during TB detection. This paper presents the solution for TB identification by using Bayesian-based convolutional neural network (B-CNN). It deals with the uncertain cases that have low discernibility among the TB and non-TB manifested CXRs. The proposed TB identification methodology based on B-CNN is evaluated on two TB benchmark datasets, i.e., Montgomery and Shenzhen. For training and testing of proposed scheme we have utilized Google Colab platform which provides NVidia Tesla K80 with 12 GB of VRAM, single core of 2.3 GHz Xeon Processor, 12 GB RAM and 320 GB of disk. B-CNN achieves 96.42% and 86.46% accuracy on both dataset, respectively as compared to the state-of-the-art machine learning and CNN approaches. Moreover, B-CNN validates its results by filtering the CXRs as confusion cases where the variance of B-CNN predicted outputs is more than a certain threshold. Results prove the supremacy of B-CNN for the identification of TB and non-TB sample CXRs as compared to counterparts in terms of accuracy, variance in the predicted probabilities and model uncertainty.

INDEX TERMS Tuberculosis identification, computer-aided diagnostics, medical image analysis, Bayesian convolutional neural networks, model uncertainty.

I. INTRODUCTION

Tuberculosis (TB) is a contagious disease that is designated among 10 highest causes of death and also the leading infectious disease above than human immunodeficiency Virus (HIV)/ acquired immune deficiency syndrome (AIDS) [1]. Each year millions of people get infected with tuberculosis. In 2017, approximately 1.3 million deaths were recorded for HIV-negative people. Additionally, the number of deaths

The associate editor coordinating the review of this manuscript and approving it for publication was Sohail Jabbar.

among HIV-positive were 3 million. The top estimation was that in total 10 million individuals suffered from TB infection disease in 2017. In which men were 5.8 million, women were 3.2 million and children were 1 million. According to the World Health Organization (WHO), the affected tuberculosis individuals were from all countries and age groups, i.e., over 90% were adults whereas, 9% of people were living with HIV. Furthermore, WHO 2018 TB report listed 30 high TB burden countries from which eight highest TB affected countries were: India (27%), China (9%), Indonesia (8%), Philippines (6%), Pakistan (5%), Nigeria (4%), Bangladesh (4%) and
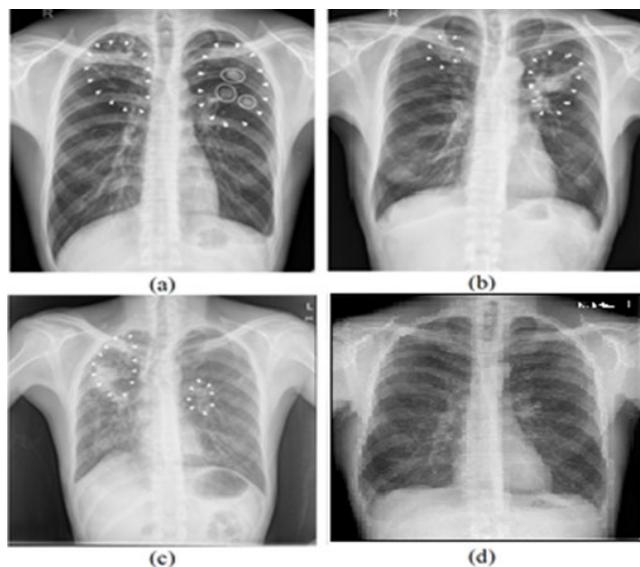
FIGURE 1. TB manifestation marked with white arrowheads whereas white circles show air space consolidation.

South Africa (3%). The other 22 countries were considered for 87% of the world's TB infected cases. Only 6% of global cases were in the European and American region (3% each) [1]. Moreover, WHO in its "End-TB Strategy" emphasizes on the timely and accurate diagnosis of TB in patients and recommends the use of chest radiography, or chest X-ray (CXR). While CXR is a commonly utilized tool for diagnosing pulmonary TB, the expertise of radiology interpretation is insufficient in TB dominant areas, which harm the efficacy of triaging and screening of TB [2]. For mass screening, an efficient and low-cost computer-aided solution can be vital for earlier identification of TB disease in developing countries [3]. In mass screening, precise disease identification and evaluation relies on the technologies that require image acquisition and image interpretation. The objective of these technologies is to overcome human interpretation issues such as restricted subjectivity, huge variations among human interpreters, high-cost and limited human resource and fatigue [4]. Such computer-aided solutions are likely to decrease the risk of false detections and facilitate mass screening efforts. These computer-aided solutions can highlight abnormalities and characterizes lung patterns to assist physicians and provide them with a second opinion. The identification of TB from CXRs is a challenging task that requires identification of patterns such as cavity, air space consolidation, endobronchial spread and pleural effusions [5].

Figure 1(a)-(c) shows sample CXR images where TB manifestation features are visible. White circles in Figure 1(a) represent air space consolidation. In Figures 1(b)-(c), spots and shaded area with ribcage depicts, endobronchial spread and the pleural effusions appearances, which are highlighted with the help of white arrowheads. In contrast, Figure 1(d) shows a normal CXR where lungs do not have any spots or shaded area other than the ribcage. As shown in Figure 1(a)-(c) TB manifested regions are overlapping that makes TB identification a challenging and complex task.

Researchers presented a number of mechanisms to differentiate between the CXRs of TB and non-TB patients. Machine learning (ML) and image analysis techniques are key components behind the recent developments in computer-aided image interpretation systems. Different researchers applied numerous techniques for feature-extraction including Gabor, histogram of oriented gradients (HOG) [6] and speeded up robust features (SURF) [7]. These features are then utilized for TB identification using ML classifiers, such as support vector machine (SVM) and regression tree (CART) [8], [9]. However, these approaches require handcrafted features for classification which do not signify all possible data exemplification.

Deep learning (DL) is an emerging development in ML and image analysis that has been designated as one of the ten breakthrough technologies of this decade. DL is a resurgence of convolutional neural network (CNN) with the capability of learning higher levels of abstraction which is crucial for improving the performance of data analysis algorithms [10]. DL based CNNs are evolving as the foremost ML approach for computer vision [11]. CNN based extracted representations are useful for recognizing and localizing objects in natural images [12]. However, due to the lack of labeled training data CNN model leads towards overfitting [13]. Furthermore, the CNN utilizes the softmax [14] in the last layer to classify the CXR as TB and non-TB. Though, the Softmax tends to classify CXR with a higher probability which is incorrectly interpreted as model confidence. In Bayesian CNN (B-CNN) the dropout between weighted layers is incorporated which can be interpreted as an approximation Bayesian model. The B-CNN for TB diagnostic provides inference ability with confidence.

The main problem is that uncertainty factor is not considered for TB diagnosis in most of the schemes including [15]–[17]. In case of higher values of uncertainty, the decision should be verified by the radiologists. Due to low explicitly and challenging inference of standard TB screening procedures, level of uncertainty may increase [18]. To address this problem, we proposed a novel dropout based Bayesian convolutional neural network which is utilized to robustly identify TB manifested CXRs. The main contribution of our work are as follows;

1) Explored extensive literature to analyze radiograph screen schemes based on machine learning and deep learning algorithms. Through literature study we have identified the problem.

2) Deep learning based TB diagnosis scheme is presented for effective mass screening of CXRs. In the proposed scheme three CNN architectures inspired by VGG-19 are presented to diagnose TB manifestations.

3) Identification of the softmax function utilized at the end of CNNs can only classify the CXR as TB and non-TB with higher confidence which is error prone to the borderline TB cases.

4) B-CNN based model is presented to identify TB by considering the uncertainty factor that effects results.

It assigns a higher uncertainty/confusion for erroneous predictions to make better decisions and avoid false positive cases.

Rest of the paper is organized as follows: Sections 2 elaborates the related work and Section 3 explains the DL preliminaries. The proposed TB identification methodology based on B-CNN is presented in Section 4. Furthermore, the results are presented in Section 5 followed by the conclusion.

II. RELATED WORK

In this Section, we explore a number of TB diagnosis schemes based on machine-learning and deep learning. TB diagnosis relies primarily on radiological patterns. Early detection of TB can lead to its treatment and prevention of infection growth. However, lack of diagnostic resources such as skilled radiologists are major hindrance for effective and robust TB diagnostics. In comparison with human diagnoses, computer-aided tools can provide more significant results with less diagnostic errors and make it possible to perform efficient mass screening with fewer resources.

A. MACHINE LEARNING BASED TB DIAGNOSIS

This section describes TB diagnosis schemes based on machine learning (ML) algorithms. The machine learning approaches rely on hand crafted features selection [19] which may work for one scenario but, fail in another. Therefore, ML algorithms demonstrate not to be used full for mass TB screening. Chauhan *et al.* [6] extracted features from CXRs using Gabor filter, gist and variants of HOG like representing shape with a spatial pyramid kernel [20] and Histograms of oriented gradients for human detection [21]. The authors then used SVM [22] for classification and reported accuracy of 94.2% and 86.0% with gist and HOG features, respectively. Alfadhli et al. used SURF [7] for features calculation from varied window sizes used as input for SVM based classifier achieving 89% area under the curve (AUC) [15]. Hogeweg *et al.* presented a feature extraction and classification mechanism to compute local pixel characteristics. The method uses position features, texture features and features derived from the Hessian matrix. The hybrid active shape model pixel algorithm was used for classification. The receiver operator characteristic (ROC) based accuracy of 84.7% was achieved [8]. Jaeger et al. presented a TB diagnostic scheme. The authors have experimented two features, i.e., features of object detection as set A, and features that are based on image retrieval as set B. furthermore, the set A comprises on histogram for intensity, gradient, magnitude, shape-descriptors, curvature-descriptors, HoG and LBP. The set B comprises on descriptor such as Tamura-texture, edge and color, fuzzy-color, color-layout [23]. In [24], Yahiaoui et al. presented preliminary diagnosis of TB by using SVM for TB and non-TB manifestations. It involves private dataset for training and testing. The ML techniques for TB identification have achieved good results. However, these technique fails to address the uncertain cases due to features selection by human experts which are hand-crafted that may work for some certain scenarios

and fail for others situations. On the other hand, DL based techniques are most suitable as the features and representations are extracted automatically from image data by using the pattern learning and back-propagation.

B. DEEP LEARNING BASED TB DIAGNOSIS

This section presents TB diagnostics methods based on Deep learning (DL). The DL-based methods achieved significant results for computer aided diagnostics. Lopes and Valiati [25] proposed bags-of-features and ensemble-based method to extract the features from segmented CXRs. The authors have collected features by using multiple CNNs including Visual Geometry Group (VGG) [26], Residual neural network (ResNet) [27] and GoogleNet [16]. For the collected features, the authors trained SVM for classification. The classification achieved AUC 78.2%, 74.6%, and 75.3% on GoogleNet, ResNet, and VggNet, respectively. Lakhani et al. presented an ensemble of AlexNet and GoogleNet and attained AUC of 99%. It uses two non-publically available datasets with unbalanced classes which can effect generalized efficacy of TB identification [3]. In [9], Évora et al. developed Artificial neural networks (ANN) to diagnose drug-resistant TB (DR-TB) for their own collected dataset. The dataset consisted of 280 subjects and achieved a classification accuracy of 88.1%. Dataset is not publically available for experiments.

Alcantara et al. extracted features with GoogleNet and performed classification using SVM on their own collected dataset which comprised of 4701 images. It achieved 89.6% accuracy for binary classification of TB and 62.07% accuracy for multi-class classification of different TB manifestations. However, classes in dataset are uneven as Lymphadenopathy has 202 images whereas Infiltration comprises of 2252 images [28]. In [29], Melendez et al. explored three techniques including SVM, multi-instance learning (MIL) and active learning (AL) for TB diagnosis. It uses a dataset of 917 CXRs including 392 normal and 525 TB cases. In this scenario, SVM achieved highest AUC of 90%. Vajda *et al.* [30] presented TB diagnosis using segmentation based on deep learning. Authors achieved 95.6% accuracy with AUC of 99% on Shenzhen dataset. The method performs segmentation on lung images. After segmentation, authors have extracted shape descriptor histogram of lung shape descriptor. For classification of TB manifests, authors have utilized a simple neural network in [31] that achieved 90.3% accuracy and 96.4% AUC. It used transfer-learning mechanism with ImageNet where 10848 CXRs were used for training purpose. In [32], Islam et al. have presented a methodology which is based on AlexNet, VGG and ResNet models. ResNet achieved higher accuracy in comparison to AlexNet and VGG, i.e., 88% and 86.2% respectively. However, the ResNet achieved 91% AUC, which is lower than the other state-of-the-art. Authors also presented an ensemble comprises of six CNN models and achieved 90% accuracy with AUC of 0.94. In [23], a CNN-based TB diagnose scheme is presented that uses mimic AexNet network architecture. For training, Shenzen datasets were used.

Although, the scheme achieves accuracy of 84.4% but probability of classification from DL techniques generally inferred as model confidence that is not true for all scenarios. Conventional DL methods for identification and regression are not enough capable of detecting the model uncertainty.

In our proposed work, we focus on recent model uncertainty integration tools with DL algorithm to cope with uncertain CXRs. In uncertain CXRs, it is difficult to identify whether TB manifestations are present or not in confusing cases. Softmax tends to misclassify TB from CXRs and represents model confidence either very high or very low which is not true. The next section explains the detailed technical background of DL approaches used for TB identification.

III. PRELIMINARIES OF DEEP LEARNING

DL techniques can learn the TB manifestation patterns and features automatically from raw CXR data used for classification and localization task. These DL techniques are capable of learning features at different abstraction levels by piling non-linear layers that allow the DL techniques to attain better simplification for difficult computer vision tasks such as image localization, segmentation, and classification.

A. CONVOLUTIONAL NEURAL NETWORKS

CNN is a DL based approach that became an emerging technique for image classification due to its significant achievements. The success achieved is because of the precise and robust assumptions of CNN for the natural image data including associations and pixel locality [26]. Secondly, CNN optimizes the tasks by significantly less number of parameters as compared to feed-forward networks [12]. For classification tasks, layers preserve features and patterns that are essential for identification and throwaway irrelevant variations. Basic architecture of CNN consists of following layers.

1) CONVOLUTION LAYER

The units of convolution layer are linked to their respective units of the local patch coming from the preceding layer with a filter. The units are activated by using the feature map, which is computed through applying Rectified Linear Unit (ReLU) [33] above the locally weighted sums.

2) POOLING LAYER

This layer works by merging features and patterns semantically into a solitary feature map. The pooling layer computes the maximum or average of input features from the previous layer and uses them as an output feature map.

3) FULLY-CONNECTED LAYER

In the fully-connected layer, every unit is connected to previous layer units and making a mesh. Usually, before fully-connected layer two or three stacks of the convolution and pooling layers are placed to extract features.

4) SOFTMAX LAYER

The purpose of the softmax layer is to converts the features into probabilities which belong to each output class. The total number of units in the softmax layer is equal to the total output classes. The softmax function is defined by equation (1) where Softmax (a_i) and a_i represent the feature and the probability of ith class respectively. The e_i^a is the non-normalized probability measurement and $\sum_{j=1}^{m} e_j^a$ is used for normalizing the distribution of probability over m output classes [14]:

$$Softmax\ (a_i) = \frac{e^{a_i}}{\sum_{j=1}^{m} e^{a_j}} \qquad (1)$$

ReLU can be used as an activation function for displaying non-linearity as well as achieving faster learning convergence [33]. Learning phase deals with weight optimization of the units to minimize erroneous classifications. Optimizer such as stochastic gradient descent is typically used to gradients for all the weights which are computed with the back-propagation algorithm.

B. BAYESIAN CONVOLUTIONAL NEURAL NETWORKS

To cope with low visibility between CXRs, a model is needed that should be proficient enough to represent the prediction uncertainty. The state-of-the-art techniques such as [34], [35] and [36], are kernel-based methods where pairs of images are checked for similarity measurement. The measured similarity is fed to the classifier as input such as in SVM. However, the proposed methodology focuses on utilizing the efficacy of CNN and highlight the utility of Bayesian uncertainties approximation [37]. The preceding probability distributions of B-CNN are defined for a set of parameters, such as $\omega = \{Wht_1, \ldots, Wht_L\}$ where Wht_L are the weights of Lth layer. A probability model can be described by a standard Gaussian assumption over ω. Probability of class y equals to c given input $\mathbb{x}$ is defined by using softmax [14] function. It includes parameter weights ω for feature set f as given in Equation (2).

$$p\ (y = c\ |\ \mathbb{x}, \omega) = softmax\ \left(f^\omega\ (\mathbb{x})\right) \qquad (2)$$

The inference in B-CNN model is implemented by using the initial stochastic regularization method such as dropout [38], [13]. The Dropout is added after every fully-connected and convolution layers in B-CNN model. During the testing phase and approximate posterior sampling, dropout is also employed. This is equal to carrying out an estimated variational inference which aims to find a manageable distribution $q_\theta^*(\omega)$ using a dataset for training $\mathcal{D}_{train}$. It is accomplished by reducing Kullback-Leibler (KL) divergence with model posterior $p\ (\omega\ |\ \mathcal{D}_{train})$. The dropout layer helps to maintain uncertainty in weights during prediction through relegating the estimated posterior via employing integration of Monte Carlo as given in equation (3) [39] where $q_\theta^*(\omega)$ is referred to as dropout distribution [40] and $\hat{\omega}$ represents the sample weights from the distribution. Equation (4) averages output probability with respect to $\hat{\omega}$ for the T number of stochastic forward passes. The vital characteristics of B-CNN for this work is its efficiency to deal with small datasets [37] and possessing of uncertainty information to deal with the uncertain

TABLE 1. List of notations.

Sr.	Notation	Description
1.	x	Input defined by using softmax
2.	D_train	Training dataset
3.	$q\_\theta^{*\,(\omega)}$	Dropout distribution
4.	T	Number of stochastic forward passes
5.	$I = \{I_{1..}I_N\}$	Set of CXRs
6.	$Y = \{y_{1..}Y_N\}$	Set of parallel labels
7.	i	Convolutional layer
8.	$t \times t$	Size of convolution layer
9.	b_i	Variational parameters
10.	M_i	Bernoulli distribution
11.	p_i	Dropout probabilities
12.	υ	Variance
13.	$k^{((r))}$	Extraction of features component
14.	$z^{((r))}$	Extracts features
15.	bias$^{(r)}$	Biases at rth layer

cases [40].

$$p\left(y = c \mid x, \mathcal{D}_{train}\right) = \int p\left(y = c \mid x, \omega\right)$$

$$p\left(\omega \mid \mathcal{D}_{train}\right) d\omega \approx \int p\left(y = c \mid x, \omega\right) q_\theta^*\left(\omega\right) d\omega \quad (3)$$

$$\approx \frac{1}{T}\sum_{t=1}^{T} p\left(y = c \mid x, \hat{\omega}_t\right) \quad (4)$$

By exploring the literature, we have identified that the state-of-the-art schemes are CNN-based and uses softmax for classification of TB and non-TB. The use of Softmax at the end of CNN architecture make the inference either 0 or 1 which means the CNN will always confidently predict between 'Yes' or 'No'. Due to softmax classifier, the state-of-the-art schemes either cannot identify reliably or inference is not correctly predicted. To address this issue we have deployed a Bayesian base CNN architecture which robustly infer the TB and non-TB manifestations.

IV. PROPOSED METHODOLOGY

In this Section, a robust TB diagnostic scheme based on Bayesian is presented. We have explored three baseline-CNN architectures that are trained on two benchmark datasets named Montgomery and Shenzhen datasets. Moreover, we have deployed Bayesian based CNN (B-CNN) architecture to overcome the softmax inference issue. We have calculated variance from prediction results of B-CNN on Shenzhen dataset to validate the robustness of proposed B-CNN. List of notations is presented in TABLE 1.

By applying B-CNN, TB identification task can be described as follows; consider a set of CXRs denoted as $I = \{I_1, I_2, \ldots, I_N\}$, where each CXR is described as CXR $I \in [0; 1]^{m \times n}$ (m is the height and n is the width of all the CXRs denoted by N number of CXRs) as a set of parallel labels $Y = \{y_1, y_2, \ldots, y_N\}$ where all the labels corresponds to a binary classification result $U \in [0; 1]$. The function f learns to map the input CXR I to equivalent labels Y. The output label presented by U_{out} which is equal to the ground-truth label represented by U_{gt}. The B-CNN architecture of proposed TB identification methodology is based on dropout [39], which is utilized for variational

inference in B-CNN [37]. Authors in [37], have described a relation between variational inference and dropout for B-CNN by Bernoulli distributions and CNN weights. This similar approach is used to represent uncertainties for the B-CNN while CXRs classification. In this proposed study, the network weights of B-CNN are utilized to learn the posterior distribution. The CXR training data I and labels Y are given as $p\left(\omega \mid I, Y\right)$. In general, this distribution is not tractable; therefore, the distribution over the weights is requisite for approximation [37]. The variational inference is employed for approximating the weights [37]. This process helps to improve the distribution approximated over weights $q(\omega)$, through reducing the KL divergence among $q\left(\omega\right)$ and $p(\omega \mid I, Y)$ as given by equation (5) where $q\left(\omega_i\right)$ is defined for each convolutional layer i of $t \times t$ size, having j units. Furthermore, $b_{i,j} \sim Bernoulli\left(p_i\right)$ for $j = 1, 2, \ldots, t_i$ and $\omega_i = M_i diag\left(b_i\right)$. In this scenario, b_i and M_i are vectors for variables distribution with variational parameters and Bernoulli distribution, respectively. Although, the optimization for dropout probabilities p_i, are at a standard value of 0.5 [13].

$$KL\left(q(\omega)\|p\left(\omega \mid I, Y\right)\right) \quad (5)$$

In [37], the authors describe that, in order to reduce KL divergence, they minimized the cross-entropy loss function. Hence, using stochastic gradient descent for network learning acquires a network weights distribution. The B-CNN model is trained by adding dropout for the CXRs classification. To obtain the class probabilities posterior distribution over the weights, the dropout is also added at testing-phase. The variance υ and mean of the samples are utilized as uncertainty and confidence, for the CXRs classification respectively. A simple heuristic function for confusion is *if* $\upsilon \geq \tau$ whereas in case of no-confusion *if* $\upsilon < \tau$ defines the confusion while prediction where τ is threshold. If the variance of B-CNN prediction is above the τ, it would mean that classifier is uncertain or confused about TB existence in the CXR and the CXR should be further analyzed by an expert radiologist for the final decision.

Proposed CNN architecture for CXRs classification is shown in Figure 2. Primarily, it comprises of three components including extraction of features, selection of features, and prediction. All component contains a sequence of procedure which comprises on layer functionality. The extraction of features component at stage $k^{(r)}$ extracts features $Z^{(r)}$ as given by the equation (6) where $*$ operator denotes convolution, $\omega^{(r)}$ is weights and $bias^{(r)}$ is biases at the rth layer. $Z^{(r-1)}$ is the input CXR I for $r = 1$ where $Z(0) = I$ as $r - 1^{th}$ hidden layer activation for $r > 1$.

$$Z^{(r)} = k^{(r)}\left(Z^{(r-1)}; \omega^{(r)}, bias^{(r)}\right) normalize$$

$$\left(pool\left(relu\left(\omega^{(r)} * Z^{(r-1)} + bias^{(r)}\right)\right)\right) \quad (6)$$

Moreover, it includes the computation of operations such as non-linear transformation, convolution, local normalization and max-pooling [12]. The feature selection component
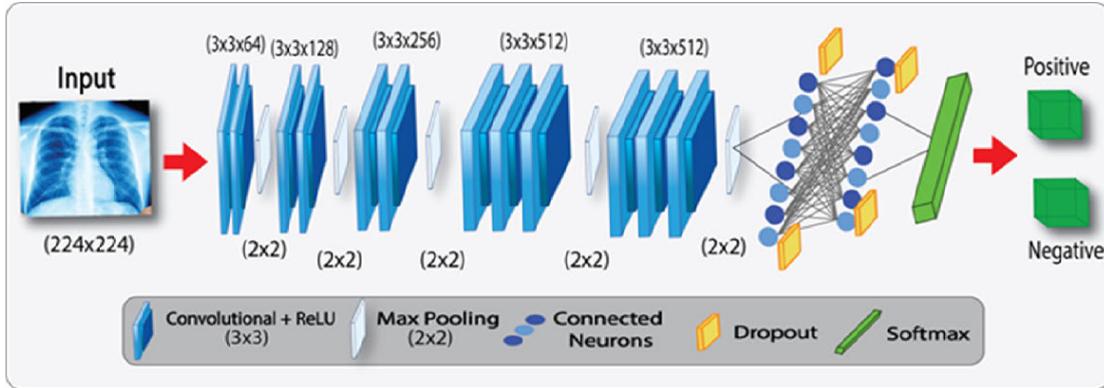
FIGURE 2. Schematic diagram of the proposed TB identification.

TABLE 2. Description of the baseline CNN architectures used to explore the efficacy of CNN.

	Architecture − 1		Architecture - 2		Architecture - 3	
Layer #	Layer type - size	Activation Function	Layer type − size	Activation Function	Layer type - size	Activation Function
1	Convol-1 − 3x3 x64	ReLU	Convol-1 - 3x3 x 64	ReLU	Convol-1 - 3x3 x64	ReLU
2	Convol-2 − 3x3 x64	ReLU	Convol-2 - 3x3 x 64	ReLU	Convol-2 - 3x3 x64	ReLU
3	Pooling-1 − 2x2	-	Pooling-1 - 2x2	-	Pooling-1 - 2x2	-
4	Convol-3 − 3x3x128	ReLU	Convol-3 - 3x3 x128	ReLU	Convol-3 - 3x3 x128	ReLU
5	Convol-4 − 3x3x128	ReLU	Convol-4 - 3x3x128	ReLU	Convol-4 - 3x3x128	ReLU
6	Pooling-2 − 2x2	-	Pooling-2 - 2x2	-	Pooling-2 - 2x2	-
7	Convol-5 − 3x3x256	ReLU	Convol-5 - 3x3x256	ReLU	Convol-5 - 3x3x256	ReLU
8	Convol-6 − 3x3x256	ReLU	Convol-6 - 3x3x256	ReLU	Convol-6 - 3x3x256	ReLU
9	Pooling-3 − 2x2	-	Pooling-3 - 2x2	-	Convol-7 - 3x3x256	ReLU
10	Convol-7 − 3x3x512	ReLU	Convol-7 - 3x3x512	ReLU	Pooling-3 - 2x2	-
11	Convol-8 − 3x3x512	ReLU	Convol-8 - 3x3x512	ReLU	Convol-8 - 3x3x512	ReLU
12	Convol-9 − 3x3x512	ReLU	Convol-9 - 3x3x512	ReLU	Convol-9 - 3x3x512	ReLU
13	Pooling-4 − 2x2	-	Pooling-4 - 2x2	-	Convol-10 - 3x3x512	ReLU
14	Convol-10 − 3x3x512	ReLU	Convol-10 - 3x3x512	ReLU	Pooling-4 - 2x2	-
15	Convol-11 − 3x3x512	ReLU	Convol-11 - 3x3x512	ReLU	Convol-11 - 3x3x512	ReLU
16	Pooling-5 − 2x2	-	Convol-12 - 3x3x512	ReLU	Convol-12 - 3x3x512	ReLU
17	FC-1 − 4096	ReLU+Drop	Pooling-5 - 2x2	-	Convol-13 - 3x3x512	ReLU
18	FC-2 − 4096	ReLU+Drop	FC-1 - 4096	ReLU+Drop	Pooling-5 - 2x2	-
19	FC-3	Softmax	FC-2 - 4096	ReLU+Drop	FC-1 - 4096	ReLU+Drop
20	--	--	FC- 3	Softmax	FC-2 - 4096	ReLU+Drop
					FC-3	Softmax

* Conv.: Convolutional, D: Dropout, FC: Fully Connected

at stage $f^{(r)}$ is expressed as given in equation (7) where (.) specifies dot product as it involves dot product operation. Furthermore, $Z^{(r-1)}$ is $r-1^{th}$ activation for the hidden layer. Precisely, it includes dot product operation tailed by non-linear transformation.

$$Z^{(r)} = f^{(r)} \left(Z^{(r-1)}; \omega^{(r)}, bias^{(r)} \right)$$
$$= \left(\text{relu} \left(\omega^{(r)} . Z^{(r-1)} + bias^{(r)} \right) \right) \quad (7)$$

Lastly, prediction component involves the softmax [14] function which provides the probability p for each neuron indicating the possible class as output C. It can be formulated as

equation (8).

$$p(C \mid I; \omega, bias) = softmax \left(\omega^{(r)} . Z^{(r-1)} + bias^{(r)} \right) \quad (8)$$

To construct TB identification based on B-CNN model, the three components are stacked as in equation (9):

$$p(C \mid I; \omega, bias) = softmax$$
$$= \left(\text{f}^{(5)} \left(\text{f}^{(4)} \left(\text{k}^{(3)} \left(\text{k}^{(2)} \left(\text{k}^{(1)} (I) \right) \right) \right) \right) \right) \quad (9)$$

Features are extracted and selected by CNN components. Additionally, ReLU is utilized as the nonlinearity activation function [40]. Next, we explore the three baseline CNN architectures as illustrated in TABLE 2. These are used to compare the efficiency of CNN.
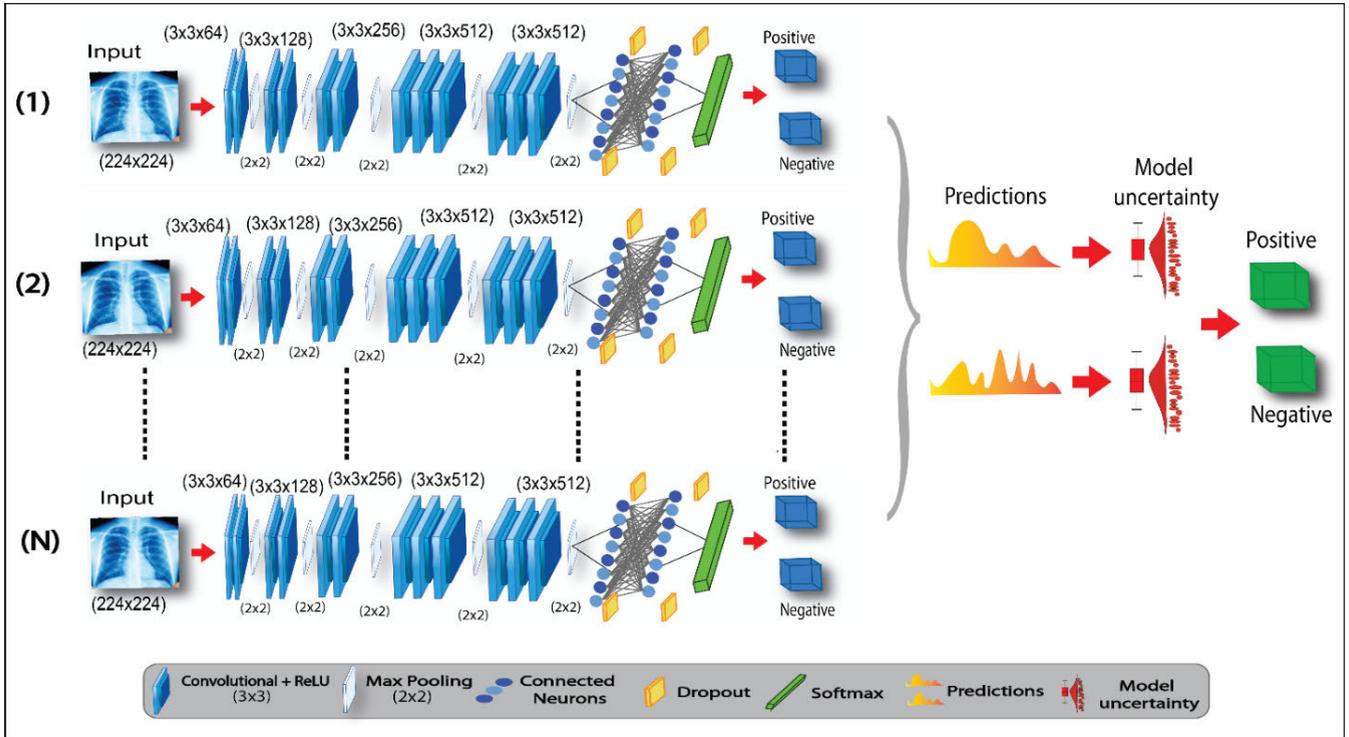
FIGURE 3. Illustration of the proposed TB identification based on B-CNN Testing method.
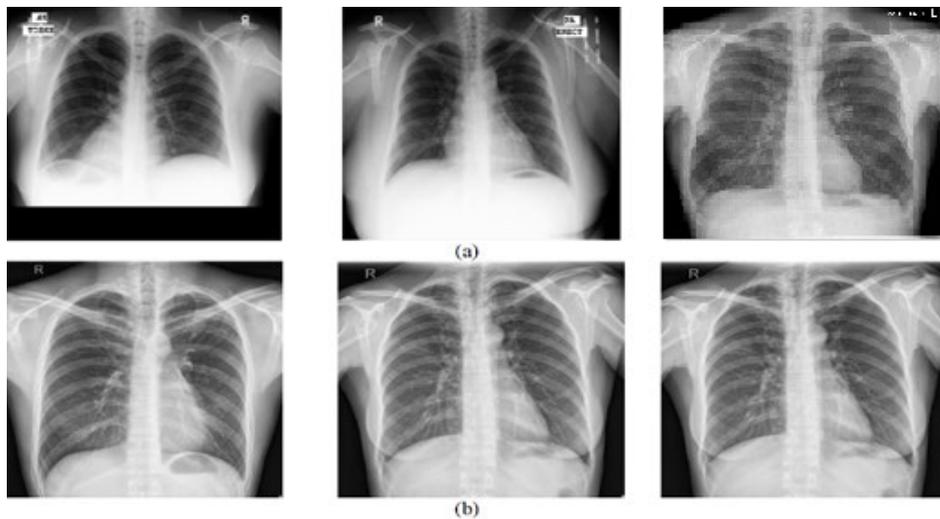

FIGURE 4. Sample CXRs from both datasets. (a) Montgomery. (b) Shenzen.

Figure 3 illustrates overall functionality of the proposed TB identification methodology based on B-CNN model. The model comprises of $1, 2, \ldots, N$ layers where each layer involves baseline-CNN architecture-2. The input CXR is processed by all these layers of B-CNN to further perform simultaneous prediction. In this model, we enable the dropout at the time of prediction which in result produces different probabilities for classification. However, the magnitude of variation between the probabilities of CXR inference shows either the model is predicting with high/low confidence which

shows the model uncertainty. If the variation is high then, the model uncertainty is also high which demonstrates that the model has low confidence while predicting CXR. In case of low uncertainty, the mean of predicted probabilities from B-CNN model is considered as final prediction of CXR. In Figure 4, CXRs are passed to N number of CNN models that generate N outputs. Mean of these outputs reflect confidence of B-CNN model with final classification result. However, each CNN produces different outputs due to addition of dropout at testing-phase. The uncertainty or confusion
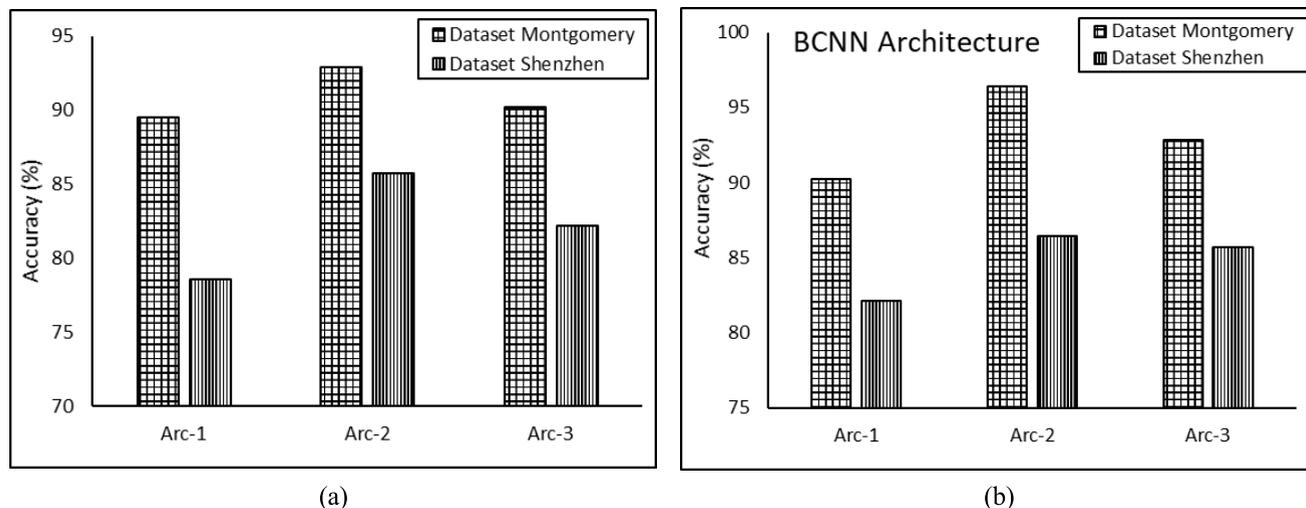
FIGURE 5. Comparison of (a) baseline- CNN architecture and (b) B-CNN for Montgomery and Shenzhen datasets.
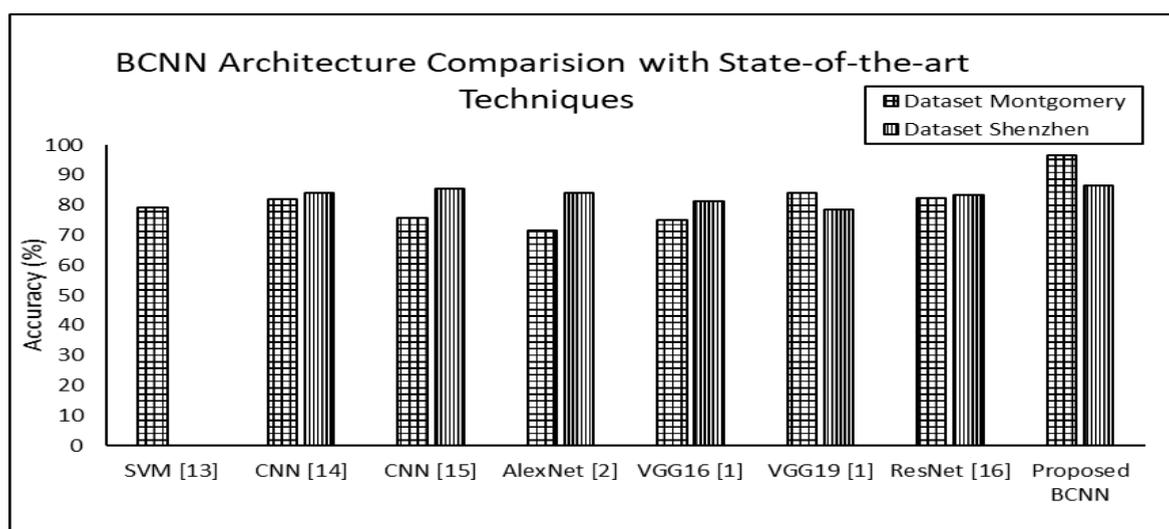

FIGURE 6. Comparison of baseline- CNN architecture for Montgomery and Shenzhen datasets.

of B-CNN model is estimated by extracting variance values from the output of *N* number of CNN models.

V. RESULTS AND DISCUSSION

In this section, we explore experimental setup and B-CNN model based on baseline-CNN architectures. We analyze baseline-CNN architectures conversion into B-CNN and their prediction probability. In the last, we have discussed in detail important part of model uncertainty and its role to pronounce the model confidence. In this study, two benchmark datasets including Montgomery and Shenzhen [23] are used for evaluating the efficacy of proposed TB identification methodology. Shenzhen dataset consists of 662 grayscale CXRs where 336 are positive TB CXRs and 326 are negative TB CXRs. The Montgomery dataset consists of 138 grayscale CXRs, where 58 CXRs are TB positive, and 80 CXRs are non-TB. Figure 4(a) and Figure 4(b) shows sample images from Montgomery and Shenzhen, respectively. For training,

80% of the both datasets are used, while the other 20% of each dataset is used for the testing. The experiments are performed on three different baseline CNN architectures as shown in Table 1. All the training and the testing experiments are performed on the platform provided by Google Colab with 16 GB of RAM and NVidia Tesla K80 GPU with ∼12 GB VRAM [41].

A. ACCURACY OF TB IDENTIFICATION

The identification accuracy is used to gauge the performance of proposed TB identification methodology based on B-CNN in comparison with other techniques from literature including: SVM [15], AlexNet [12], VGG16 [26], VGG19 [26], ResNet [42], and CNN based methodologies proposed by Lopes and Valiati [25] and Sivaramakrishnan *et al.* [17]. For the experimentation and comparison, three baseline CNN architectures are proposed, which are presented in Table 2 including Architecture-1, Architecture-2, and Architecture-3.
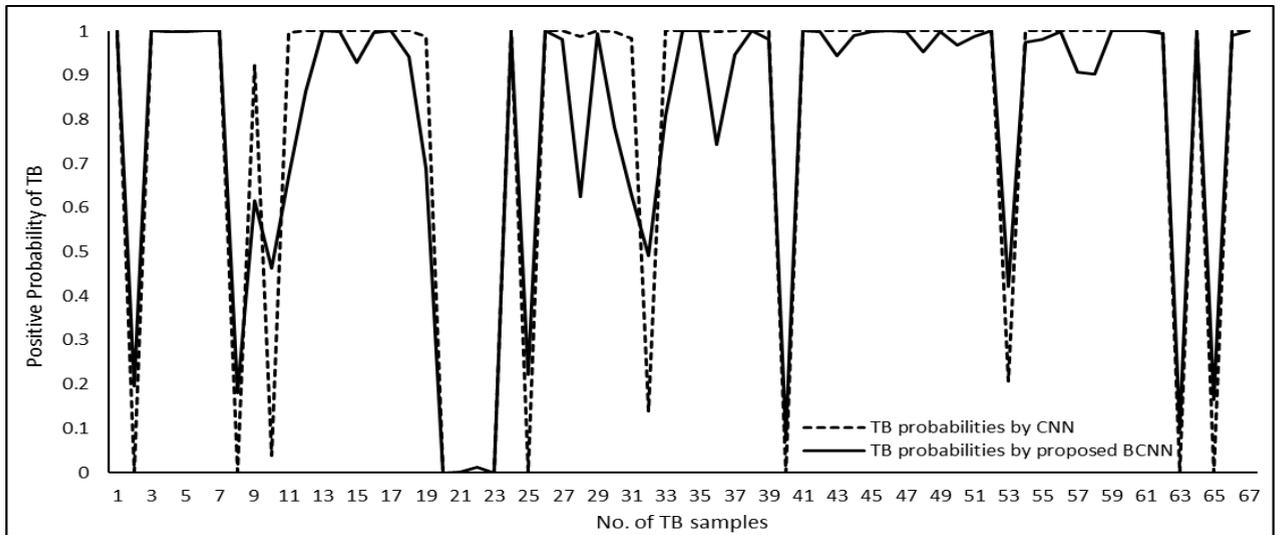
(a)
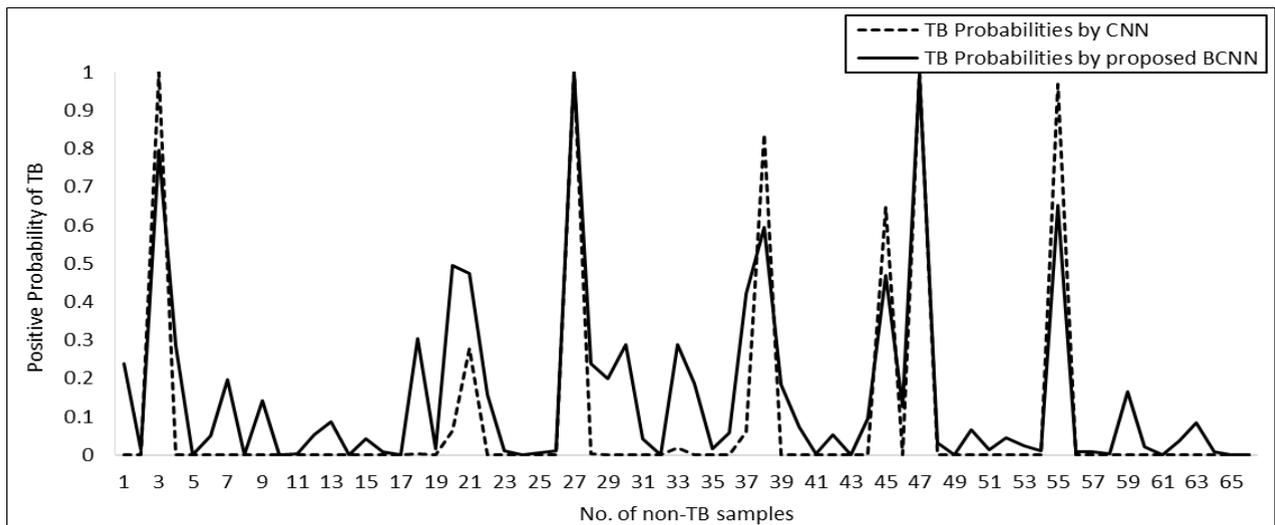
(b)

FIGURE 7. (a) Positive probability of TB in TB manifested CXRs samples. (b) Positive probability of TB in non-TB manifested CXRs samples.

The baseline CNN architectures are trained and tested separately for both datasets including Montgomery and Shenzhen [23]. Moreover, the results from Table 3 clearly depicts that all of the proposed CNN architectures outperformed the state-of-the-art DL approaches from the literature in terms of identification accuracies in case of Montgomery. For both datasets, the Architecture-2 shows highest identification accuracy as compared to state-of-the-art DL approaches along with Architecture-1 and Architecture 3.

Figure 5(a) elucidates a graphical representation of achieved accuracy for all three proposed baseline-CNN architectures. The baseline-CNN architectures are trained and tested for both Montgomery and Shenzhen datasets. It can be observed that the proposed baseline-CNNs including Arc-1, Arc-2 and Arc-3 correctly predicted TB manifestation and achieved 89.47%, 92.85% and 90.22% test-time prediction accuracy respectively for Montgomery dataset. For Shenzhen

dataset, the baseline-CNN Arc-1, Arc-2 and Arc-3 achieved 78.57%, 78.57% and 82.14% test-time prediction accuracy respectively. The test-time accuracy results clearly depict that proposed baseline-CNN arc-2 outperforms other two proposed baseline CNNs.

Figure 5(b) demonstrates results for Bayesian–based CNNs architectures. The test-time prediction accuracy of all three architectures, i.e., Arc-1, Arc-2 and Arc-3 is increased after incorporating Bayesian. The improved prediction accuracy of all architectures is 90.22%, 96.42%, 92.85% respectively for Montgomery. The accuracy of all three B-CNN for Shenzhen dataset is also improved such as Arch-1 achieved 82.14%, Arch-2 achieved 86.46% and, Arch-3 achieved 85.71% test- time prediction accuracies respectively. The results show that Arch-2 B-CNN dominates over other two architecture of B-CNN.
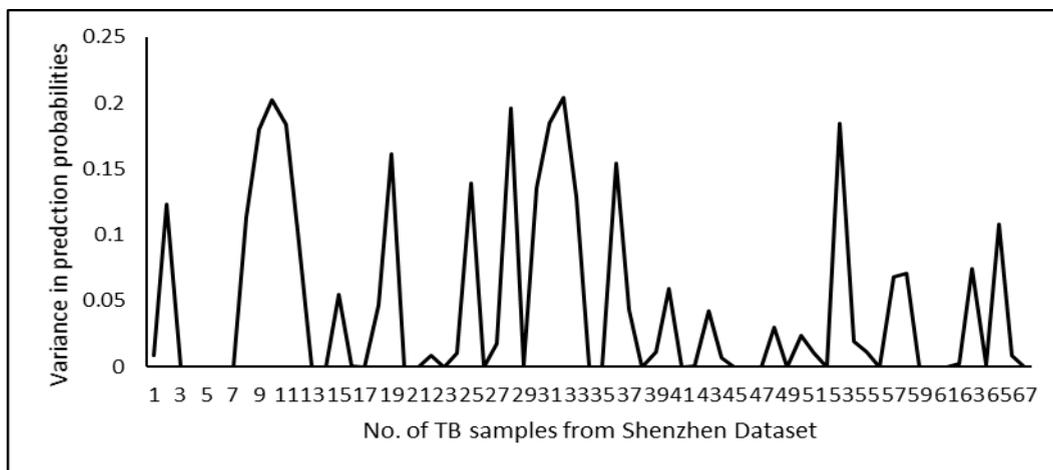
(a)
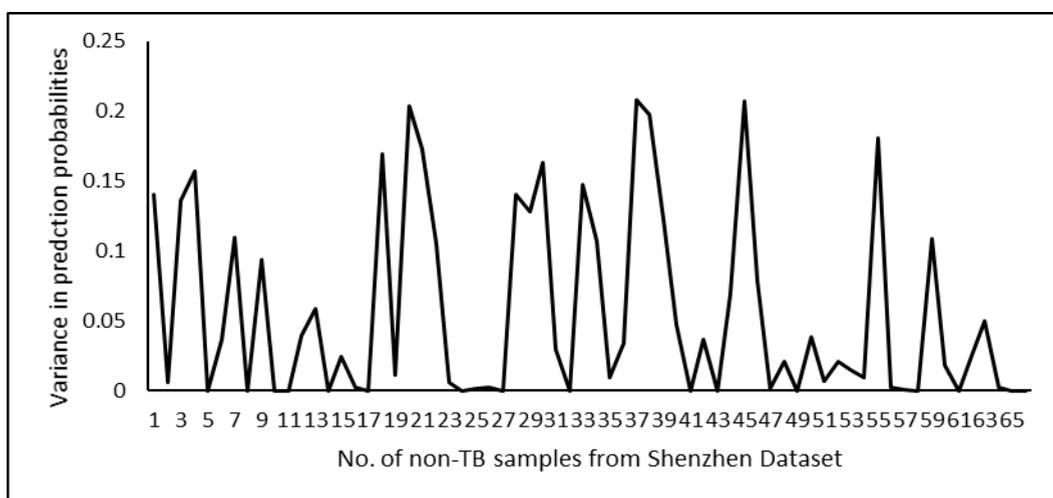
(b)

FIGURE 8. (a) Uncertainty captured while predicting test samples from Shenzhen in TB samples. (b) The captured uncertainty in non-TB samples.

To evaluate the robustness of our proposed scheme we have compared the results of Acr-2 B-CNN with other state-of-the-art schemes. From Figure 6 it can be observed that the proposed Arch-2 B-CNN model accurately predicted the TB manifestations with test-time prediction accuracy 96.42% for Montgomery and 86.46 % for Shenzhen respectively. Whereas for the same experimental settings the other state-of-the-art schemes such as SVM, pertained–CNN1, pertained–CNN2, AlexNet, VGG16, VGG19 and ResNet achieved 79.1%, 82%, 75.8%, 71.42%, 75%, 78.57%, 82.14% respectively for Montgomery dataset. The prediction accuracy results for Shenzhen dataset are 84%, 85.5%, 84.21%, 81.20%, 84.21%, 83.45% respectively for pertained–CNN1, pertained–CNN2, AlexNet, VGG16, VGG19 and ResNet. Result show that proposed Acr-2 B-CNN model performs better as compared to state-of-the-art schemes.

Furthermore, Figure 7 (a-b) shows the predictive probabilities of softmax and confidence of B-CNN respectively for TB and Non-TB cases from the Shenzhen dataset. It can be observed from both graphs that softmax tends to be at the extreme in most of the cases which shows

TABLE 3. Result summary for both benchmark datasets.

Method		Identification Accuracy (%)	
		dataset Montgomery	dataset Shenzhen
SVM [15]		79.1	-
CNN [25]		82	84
CNN [17]		75.8	85.5
AlexNet [12]		71.42	84.21
VGG16 [26]		75	81.20
VGG19 [26]		78.57	84.21
ResNet [45]		82.14	83.45
Baseline – CNN	**Arc-1**	89.47	78.57
	Arc-2	92.85	85.71
	Arc-3	90.22	82.14
	Arc-1	90.22	82.14
Proposed B-	**Arc-2**	**96.42**	**86.46**
	Arc-3	92.85	85.71

its unrealistic behavior. On the contrary, B-CNN provides realistically looking confidence since various examples may vary in confidence regarding their association with classes.
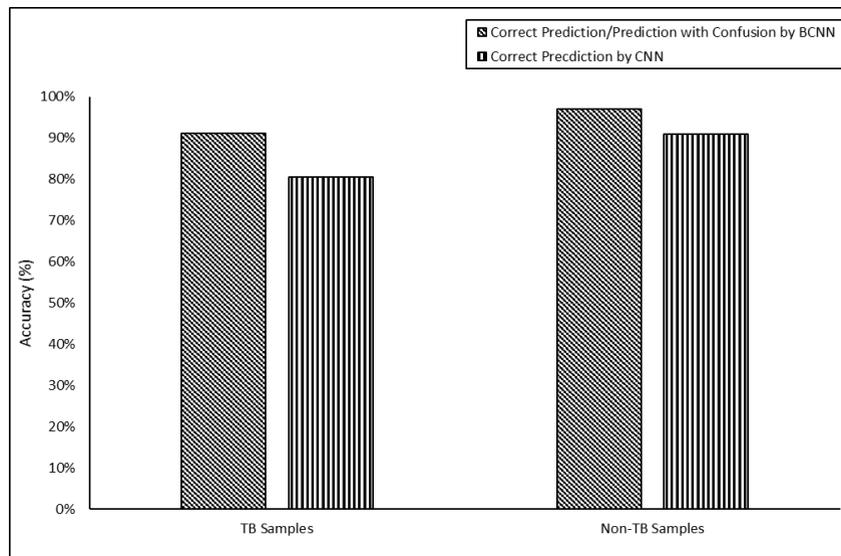
FIGURE 9. CNN and B-CNN comparison result after threshold on variance while predicting TB and non-TB samples.
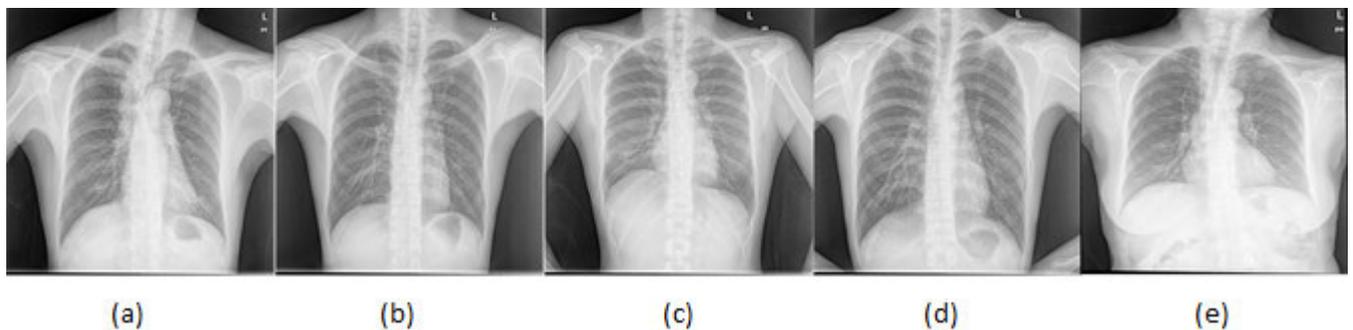

FIGURE 10. Comparison of sample CXRs from Shenzhen dataset predicted by CNN and the proposed methodology based on B-CNN.

Figure 8 (a) and (b) show variance in probability of TB and non-TB samples by B-CNN. After applying a threshold of 0.1 for TB and non-TB samples from Shenzhen dataset the variance is up to 10.4% of the false negatives (wrong prediction of TB CXRs) and 4.5% of the false positives (wrong prediction of non-TB CXRs) are declared as confusion cases from Shenzhen dataset, respectively. The CXRs which are declared as confusion cases are wrongly predicted by CNN architectures. However, the B-CNN model has either correctly classified those CXRs or has assign them as confusion cases that gives a chance to an expert radiologist for making the final decision. In Figure 9, we further investigated the utility of B-CNN. we have compared the model variance to filter confusion cases after applying threshold on all the TB and non-TB samples from Shenzhen dataset. The prediction accuracy achieved by CNN for TB-samples 80.6%, whereas B-CNN achieved prediction accuracy of 91.04%. For non-TB sample, baseline-CNN achieved prediction accuracy of 90.91% in-comparison to B-CCN achieved prediction accuracy of 96.97%. Proposed TB identification methodology based on B-CNN not only

improves the accuracy of TB identification from the CXRs but also validates its prediction. Furthermore, B-CNN can correctly classify up to 93.9% that is either the CXR belongs to TB or non-TB status or is a marginal confusion case that needs to be examined by expert radiologists. Results show that CNN Arc-2 achieves 85.7% accuracy whereas proposed B-CNN achieves 93.9% accuracy using Shenzhen dataset. The effectiveness of B-CNN over baseline CNN models can be further analyzed by Figure 10 that shows the TB prediction results on Shenzhen dataset. Figure 10 (a) - (c) display sample non-TB CXRs. Figure 10 (d) and (e) demonstrate sample TB.

The CXR in Figure 10 (a) is incorrectly predicted as positive for TB by baseline CNN model. However, the proposed TB identification methodology based on B-CNN model correctly predicted it as a non-TB sample. The baseline CNN architectures and B-CNN model wrongly predicted the CXRs in Figure 10(b) to Figure 10 (e); however, the B-CNN model lowered the prediction confidence and showed confusion on those CXRs in comparison to baseline CNN architectures. Moreover, the variance in the prediction of CXRs in

Figure 10 (b) to Figure 10 (e) is more than 0.1. Thus by applying a threshold on the variance of predicted CXRs the wrongly predicted CXRs fall into the confusion cases. This proposed TB identification methodology based on B-CNN is significant and novel as it characterizes the wrongly predicted CXRs as confusion cases which can be further analyzed by an expert radiologist. The diagnostic of contagious diseases is crucial through computer-aided solutions as mentioned in the future work of DSW Ting *et al.* [18], there is a need to validate the results of diagnostic. In our work, the proposed B-CNN based TB identification validates its results by utilizing the uncertainty. It shows higher uncertainty where the model faces confusion or wrongly predicts marginal cases among the TB and non-TB manifestations. However, the state-of-the-art DL approaches tend to wrongly predict the borderline TB and non-TB manifestation cases with higher confidence which leads to erroneous prediction. It shows that rather than making classification decisions on model confidence only, the inclusion of uncertainty can improve the decision process.

VI. CONCLUSION
TB identification is a challenging task due to the occurrence of complex and obscured patterns present in CXRs. Numerous TB identification approaches are proposed in the literature. However, the accuracy of TB identification is still low. Recently, DL based approaches have presented significant results for TB identification by discovering complex features from large datasets. However, conventional deep learning based models are not capable to suggest uncertainty for output class prediction. Proposed B-CNN exploits the model uncertainty and Bayesian confidence to improve the accuracy of TB identification as well as validation of the results. The dropout is utilized at the training and testing phase to acquire the posterior probability distribution of the classes. The variance in probabilities is utilized as uncertainty and the mean of the posterior probability distribution of the classes is used to make the final output classification decision. The proposed methodology based on B-CNN has experimented on two TB benchmark datasets: Montgomery and Shenzhen. The results have demonstrated that the proposed methodology achieved highly significant results in terms of TB identification accuracy, i.e., 96.42% and 86.46% for both datasets (i.e., Montgomery and Shenzhen) in comparison with the state-of-the-art DL approaches in the literature. Furthermore, by utilizing the variance of proposed TB identification methodology based on B-CNN a threshold of 0.1 is applied to filter out the confusion cases of the Shenzhen dataset. The applied threshold helped to capture the 10.4% of false negative and the 4.5% of false positive as confusion cases. These confusion cases improve the classification in terms of validation up to 93.9% that is either the CXR belongs to TB or non-TB status or is a marginal confusion case that needs to be examined by an expert radiologist. This proposed work shows compelling results and provides an improved foundation for future work. However, the real challenge is to determine that, how such DL systems will accurately fit with

the workflow for diagnostic of thorax diseases and clinical screening settings. In future, we shall focus on potential validation that will provide a robust approach to deal with multiple thorax diseases from CXRs. Furthermore, BCNN models shall be utilized to eliminate uncertainty factor in natural images classification.

REFERENCES
[1] *WHO | Global Tuberculosis Report 2018*, World Health Org., Geneva, Switzerland, 2018.
[2] *Chest Radiography in Tuberculosis*, WHO Library Cataloguing-in-Publication Data, World Health Org., Geneva, Switzerland, 2016, p. 44.
[3] 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.
[4] M. J. Cummings and N. W. Schluger, "The diagnosis of pulmonary tuberculosis: Established and emerging approaches for clinicians in high-income and low-income settings," *Clin. Pulmonary Med.*, vol. 25, no. 5, pp. 170–176, Sep. 2018.
[5] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[6] A. Chauhan, D. Chauhan, and C. Rout, "Role of gist and PHOG features in computer-aided diagnosis of tuberculosis without segmentation," *PLoS ONE*, vol. 9, no. 11, Nov. 2014, Art. no. e112980.
[7] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.
[8] L. Hogeweg, C. I. Sanchez, 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. Imaging*, vol. 34, no. 12, pp. 2429–2442, Dec. 2015.
[9] L. Évora, J. Seixas, and A. Kritski, "Neural network models for supporting drug and multidrug resistant tuberculosis screening diagnosis," *Neurocomputing*, vol. 265, pp. 116–126, Nov. 2017.
[10] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[11] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, "Recent advances in convolutional neural network acceleration," *Neurocomputing*, vol. 323, pp. 37–51, Jan. 2019.
[12] J. Günther, P. M. Pilarski, G. Helfrich, H. Shen, and K. Diepold, "First steps towards an intelligent laser welding architecture using deep neural networks and reinforcement learning," *Procedia Technol.*, vol. 15, pp. 474–483, 2014.
[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[14] N. M. Nasrabadi, "Pattern recognition and machine learning," *J. Electron. Imag.*, vol. 16, no. 4, Jan. 2007, Art. no. 049901.
[15] F. H. O. Alfadhli, A. A. Mand, M. S. Sayeed, K. S. Sim, and M. Al-Shabi, "Classification of tuberculosis with SURF spatial pyramid features," in *Proc. Int. Conf. Robot., Autom. Sci. (ICORAS)*, Mar. 2018, pp. 1–5.
[16] 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, Dec. 2019, Art. no. 6268 .
[17] R. Sivaramakrishnan, "Comparing deep learning models for population screening using chest radiography," in *Proc. Med. Imag., Comput.-Aided Diagnosis*, vol. 10575, 2018, p. 49.
[18] D. S. W. Ting, T.-E. Tan, and C. C. T. Lim, "Development and validation of a deep learning system for detection of active pulmonary tuberculosis on chest radiographs: Clinical and technical considerations," *Clin. Infectious Diseases*, vol. 69, no. 5, pp. 748–750, Aug. 2019.

[19] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018.

[20] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proc. 6th ACM Int. Conf. Image Video Retr.*, 2007, pp. 401–408.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2005, pp. 886–893.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[23] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases.," *Quant. Imag. Med. Surg.*, vol. 4, no. 6, pp. 475–477, Dec. 2014.

[24] S. Ivanova, K. Ivanov, E. Petkova, S. Gueorgiev, and D. Kiradzhiyska, "Methods for detection of the misuse of 'anti-oestrogens and aromatase inhibitors' in sport," *Biomed. Res.*, vol. 28, no. 16, pp. 7157–7166, 2017.

[25] U. Lopes and J. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Comput. Biol. Med.*, vol. 89, pp. 135–143, Oct. 2017.

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

[27] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," 2016, *arXiv:1603.08029*. [Online]. Available: https://arxiv.org/abs/1603.08029

[28] M. F. Alcantara, Y. Cao, C. Liu, B. Liu, M. Brunette, N. Zhang, T. Sun, P. Zhang, Q. Chen, Y. Li, C. Morocho Albarracin, J. Peinado, E. Sanchez Garavito, L. Lecca Garcia, and W. H. Curioso, "Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Perú," *Smart Health*, vols. 1–2, pp. 66–76, Jun. 2017.

[29] J. Melendez, B. Van Ginneken, P. Maduskar, R. H. H. M. Philipsen, H. Ayles, and C. I. Sanchez, "On combining multiple-instance learning and active learning for computer-aided detection of tuberculosis," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1013–1024, Apr. 2016.

[30] S. Vajda, A. Karargyris, S. Jaeger, K. C. Santosh, S. Candemir, Z. Xue, S. Antani, and G. Thoma, "Feature selection for automatic tuberculosis screening in frontal chest radiographs," *J. Med. Syst.*, vol. 42, no. 8, p. 146, Aug. 2018.

[31] S. Hwang, H.-E. Kim, J. Jeong, and H.-J. Kim, "A novel approach for tuberculosis screening based on deep convolutional neural networks," *Proc. SPIE*, vol. 9785, 2016, Art. no. 97852W.

[32] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest X-rays using deep convolutional neural networks," 2017, *arXiv:1705.09850*. [Online]. Available: https://arxiv.org/abs/1705.09850

[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[34] X. Zhu, J. Lafferty, and Z. Ghahramani, "Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Workshop Continuum Labeled Unlabeled Data Mach. Learning Data Mining (ICML)*, vol. 3, 2003, pp. 1–8.

[35] X. Li and Y. Guo, "Adaptive active learning for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 859–866.

[36] J. Wu, V. S. Sheng, J. Zhang, P. Zhao, and Z. Cui, "Multi-label active learning for image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5227–5231.

[37] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with Bernoulli approximate variational inference," 2015, *arXiv:1506.02158*. [Online]. Available: https://arxiv.org/abs/1506.02158

[38] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," Jul. 2012, *arXiv:1207.0580*. [Online]. Available: https://arxiv.org/abs/1207.0580

[39] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn.*, vol. 48, 2016, pp. 1050–1059.

[40] Y. Gal, "Uncertainty in deep learning," M.S. thesis, Univ. Cambridge, Cambridge, U.K., 2016. [Online]. Available: http://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf

[41] *Hello, Colaboratory—Colaboratory*. Accessed: Jan. 19, 2019. [Online]. Available: https://colab.research.google.com/notebooks/welcome.ipynb

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

ZAIN UL ABIDEEN received the M.C.S. degree from the National University of Modern Languages (NUML), Islamabad, Pakistan, in 2016, and scored Gold Medal in his degree. He is currently pursuing the M.S. (CS) degree in in the field of deep learning and digital image processing from COMSATS University, Islamabad, Pakistan. From 2016 to 2017, he was a Software Developer with Phlox Solutions, Islamabad. He is teaching as a Visiting Lecturer at NUML, Islamabad, since 2017. He has published several articles in international conferences. He is currently working as a Software Development Team Lead at Industrial Technology Cell ORIC, NUML, Islamabad. His areas of interests are deep learning, digital image processing, nature inspired, and meta-heurist algorithms.

MUBEEN GHAFOOR received the Ph.D. degree in image processing from Mohammad Ali Jinnah University, Pakistan. He is currently working as an Assistant Professor with the Computer Science Department, COMSATS Islamabad. He has vast research and industrial experience in the fields of data sciences, image processing, machine vision systems, biometrics systems, signal analysis, GPU-based hardware design, and software system designing.

KAMRAN MUNIR is currently an Associate Professor in data science with the Department of Computer Science and Creative Technologies. His research projects are in the areas of data science, big data and analytics, artificial intelligence, and virtual reality mainly funded by the European Commission (EC), Innovate U.K., and British Council. In the past, he has contributed to the various CERN (the European Organization for Nuclear Research) and EC funded projects e.g., CERN WISDOM, CERN CMS Production, EC Asia Link STAFF, EC Health-e-Child, EC neuGRID, and EC neuGRID4You (N4U) in which he led the Joint Research Area and the development of Data Atlas/Analysis Base, Big Data Integration and Information Services. He has published over 50 research articles. He is a regular PC member and an editor of various conferences and journals. His role also includes the leadership/production of Computer Science and Data/Information Science degree courses/modules in collaboration with the industry, such as big data, data science, cloud computing, and information practitioner. He also enjoys frequent collaborations with graduates, including the conduct of collaborative work with industry and has a number of successful M.Phil. and Ph.D. theses supervisions.

MADEEHA SAQIB is currently a Lecturer with the Department of Computer Information Systems, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. Her research interests include technology management, organizational management, and organizational communication.

ATA ULLAH received the B.S. (CS) and M.S. (CS) degrees from COMSATS Islamabad, Pakistan, in 2005 and 2007, respectively, and the Ph.D. (CS) degree from IIUI, Pakistan, in 2016. From 2007 to 2008, he was a Software Engineer at Streaming Networks, Islamabad. In 2008, he joined NUML, Islamabad, Pakistan, and worked as an Assistant Professor/Head Project Committee at the Department of Computer Science, until 2017. From November 2017 to 2018, he worked as a Research Fellow with the School of Computer and Communication Engineering, University of Science and Technology Beijing, China. He rejoined NUML, in September 2018. He has supervised 112 projects at undergraduate level and won One International and 45 National Level Software Competitions. He received the ICT funding for the development of projects. He has published several articles in ISI indexed impact factor journals and international conferences. He is also a reviewer and a guest editor for Journal and conference publications. His areas of interests are WSN, the IoT, cyber physical social thinking (CPST) space, health-services, NGN, and VoIP and their security solutions.

TEHSEEN ZIA received the Ph.D. degree from the Vienna University of Technology. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University of Information Technology, Islamabad. Before that, he was an Assistant Professor and a Lecturer with the University of Sargodha and a Researcher at the Vienna University of Technology, Austria. His research interests include machine learning, text mining, and surveillance and monitoring systems.

SYED ALI TARIQ received the B.S. degree in computer and information sciences from PIEAS, Pakistan, and the M.S. degree in computer science from Abasyn University, Islamabad. He is currently pursuing the Ph.D. degree with COMSATS University, Islamabad. He is also working with the Medical Imaging and Diagnostics Lab, COMSATS. His areas of interests include image processing, deep learning, biometric systems, and GPU-based parallel computing.
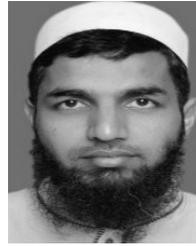
GHUFRAN AHMED received the Ph.D. degree from the Department of Computer Science, Mohammad Ali Jinnah University (renamed to Capital University of Science and Technology), Islamabad, in 2013. He has been serving as an Associate Professor with the Department of Computer Science, FAST – National University of Computer and Emerging Sciences (NUCES), since January 2020. Before joining FAST, he served as an Assistant Professor for the Department of Computer Science, COMSATS University, Islamabad, Pakistan. He held a postdoctoral position with the Department of Computer Science and Digital Technology, Faculty of Engineering and Environment, Northumbria University, Newcastle Upon Tyne, U.K., in 2016. During his Ph.D., he worked as a Visiting Research Scholar at the CReWMaN Lab, Department of Computer Science and Engineering, The University of Texas at Arlington, from 2008 to 2009. His areas of research are the IoT, wireless sensor networks, and wireless body area networks. He has chaired a number of IEEE conferences and workshops, such as iThing2019, UIC2019, IOP2019, and SCI2019. He is serving as an Editorial Board Member of *Ad Hoc & Sensor Wireless Networks* (AHSWN). He has also served as a Lead Guest Editor for special sections in *International Journal of Distributed Sensor Networks* (IJDSN), *Hindawi Journal of Sensors*, and *Hindawi Journal of Wireless Communication and Mobile Computing*. He is working as an Associate Editor of IEEE Access, an Academic Editor of *Hindawi Journal of Wireless Communications and Mobile Computing*, and *Hindawi Journal of Sensors*.

ASMA ZAHRA received the B.S. (CS) degree from SBKWU, Pakistan, in 2016. She is currently pursuing M.S. (CS) degree from COMSATS Islamabad, Pakistan. She is currently working as a Graphic Designing Team Lead with Information Technology Cell ORIC, NUML, Islamabad. She has published several articles in international conferences. Her areas of interests are image processing, deep learning, the Internet of Things, and smart city surveillance and monitoring systems.

• • •