A Deep Learning Framework Integrating the Spectral and Spatial Features for Image-Assisted Medical Diagnostics

SUSMITA GHOSH[1], **SWAGATAM DAS**[1],
AND RAMMOHAN MALLIPEDDI[2], **(Senior Member, IEEE)**
[1]Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata 700108, India
[2]Department of Artificial Intelligence, Kyungpook National University, Daegu 7027021, South Korea

Corresponding author: Rammohan Mallipeddi (mallipeddi.ram@gmail.com)

ABSTRACT The development of a computer-aided disease detection system to ease the long and arduous manual diagnostic process is an emerging research interest. Living through the recent outbreak of the COVID-19 virus, we propose a machine learning and computer vision algorithms-based automatic diagnostic solution for detecting the COVID-19 infection. Our proposed method applies to chest radiograph that uses readily available infrastructure. No studies in this direction have considered the spatial aspect of the medical images. This motivates us to investigate the role of spectral-domain information of medical images along with the spatial content towards improved disease detection ability. Successful integration of spatial and spectral features is demonstrated on the COVID-19 infection detection task. Our proposed method comprises three stages – Feature extraction, Dimensionality reduction via projection, and prediction. At first, images are transformed into spectral and spatio-spectral domains by using Discrete cosine transform (DCT) and Discrete Wavelet transform (DWT), two powerful image processing algorithms. Next, features from spatial, spectral, and spatio-spectral domains are projected into a lower dimension through the Convolutional Neural Network (CNN), and those three types of projected features are then fed to Multilayer Perceptron (MLP) for final prediction. The combination of the three types of features yielded superior performance than any of the features when used individually. This indicates the presence of complementary information in the spectral domain of the chest radiograph to characterize the considered medical condition. Moreover, saliency maps corresponding to classes representing different medical conditions demonstrate the reliability of the proposed method. The study is further extended to identify different medical conditions using diverse medical image datasets and shows the efficiency of leveraging the combined features. Altogether, the proposed method exhibits potential as a generalized and robust medical image-assisted diagnostic solution.

INDEX TERMS Medical imaging, diagnostic solution, COVID-19 detection, discrete cosine transform, discrete wavelet transform, deep learning, class imbalance, saliency map.

I. INTRODUCTION

The proper diagnosis of any medical condition plays an important role in effective treatment and also in the prevention of any infectious disease to spread out. Various Machine learning-based diagnostic solution has been proposed to ease such a process of manual diagnosis that requires domain

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

expertise and long training time [1]. The recent outbreak of Coronavirus disease 2019 (COVID-19) is the third significant Coronavirus outbreak in less than 20 years. In this context, a computer-based diagnostic solution with readily available infrastructure even in rural areas around the globe is the need of the hour. According to [2], a chest radiograph of a COVID-19 infected person exhibit 'patchy or diffuse reticular–nodular opacities and consolidation, with basal, peripheral and bilateral predominance'. Thus, the readily and

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

IEEE*Access*

widely available infrastructure for X-rays may be utilized for primary and immediate assessment for detecting COVID-19 infection.

We aim to develop an automatic model using machine learning algorithms that would aid the clinician as an adjunct tool for the diagnosis of COVID-19 infection. We pose this COVID-19 detection problem as a three-class classification paradigm where the classes are *Normal*, *Pneumonia* (non-COVID), and *COVID-19* by utilizing deep learning algorithms. The reason behind choosing *Pneumonia* as one of the classes is that Pneumonia and COVID-19 can be easily confused with one another. Both non-COVID-19 Pneumonia and COVID-19 may exhibit ground glass patterns in the lung due to lung infiltration and consolidation. Over the last decade, deep learning, a subfield of machine learning has gained popularity in assisting as a diagnostic aid. The successful application of deep learning can be found in diagnostic assessment of different biomedical conditions such as arrhythmia detection [3], skin cancer classification [4], breast cancer detection [5], brain disease classification [6], Pneumonia detection from chest X-ray images [7] and lung segmentation [8]. These studies on the application of deep learning algorithms on medical imaging data show the efficiency of a deep learning algorithm in expressing complex patterns that are even difficult to capture in untrained eyes. These studies motivate us to exploit deep learning algorithms for characterizing such intricate, differentiating patterns to identify the underlying medical condition.

Few recent studies have endeavored in detecting COVID-19 infection from the chest X-ray images with the help of deep learning algorithms. Wang and Wong [9] has proposed a Convolutional Neural Network (CNN) based architecture build using generative synthesis, referred to as COVID-net which is trained on 13,975 CXR images across 13,870 persons belongs to the categories of *Normal*, *Pneumonia*, and *COVID-19* class. They have reported achieving an overall test accuracy of 93.3% with sensitivity and precision to *COVID-19* class of 91.0% and 98.9% respectively. In another study [10], authors have presented a two-stage network to classify four classes, namely *Normal*, *Bacterial*, *Tuberculosis* (TB), and *Viral Pneumonia/COVID-19*. They first trained an extended fully convolutional
(FC)-DenseNet103 for image segmentation purposes, thereafter a patched-based CNN was trained by the segmented 403 lung images. The proposed method yields an accuracy of 88.9% with a specificity of 96.4% on the test data comprising 99 samples. Their study is further extended to three-class (*Normal*, *Pneumonia*, and *COVID-19*) classification which yielded an accuracy of 91.9%. Though, this study achieved high sensitivity for *COVID-19* class (100%), the low precision value (76.9%) for *COVID-19* class is not appropriate for any practical scenario. The DarkNet model was implemented using seven convolutional layers and various filterings on each layer for automatic detection of COVID-19 using the raw chest X-ray images [11]. The model aimed

at providing correct diagnostic predictions for binary classification (COVID vs. No-Findings) and multi-class classification (COVID vs. No-Findings vs. *Pneumonia*). This system yielded a classification accuracy of 98.08% for binary classes and 87.02% for multi-class cases for a dataset of a limited number of samples. The performances of various neural architectures for detecting COVID-19 infection from chest X-rays were evaluated in [12]. Their results indicated that deep neural networks aided with X-ray imaging could detect prominent biomarkers pertinent to COVID-19 infection, while the best accuracy, sensitivity, and specificity obtained were reported as 96.78%, 98.66%, and 96.46% respectively. Another study [13] claimed the superiority of COVID-CAPS, a modeling framework based on Capsule Networks with fewer trainable parameters over CNN-based models for COVID-19 detection. COVID-CAPS obtained an Accuracy of 95.7%, Sensitivity of 90%, Specificity of 95.8%, and Area Under the Curve (AUC) of 0.97 for binary classifications. Another study [14] proposed CNN model with end-to-end training process for classifying among the chest x-ray images of *Normal* and *COVID-19* classes. Integration of deep learning-based features extractor with Support Vector Machine (SVM) based classifier has achieved an accuracy of 92.6%. However, the binary classification performances were reported on very limited samples. Togaçar et. al. [15] employed the Fuzzy Color technique for preprocessing the chest X-ray images followed by an image stacking operation to eliminate the existing noises. The integration of deep learning architectures with SVM classifier led to an accuracy of 99.27% for three-class (*Normal*, *Pneumonia* and *COVID-19*) classification. However, the proposed method was validated on a dataset of a total of 458 chest radiograph images. Another similar study [16] showed that the fusion of five deep learning models via integration stacking achieved 99.08% accuracy on limited samples.

The body of literature related to COVID-19 infection identification from chest X-rays shows the efficient application of deep learning algorithms. Among the above-mentioned studies, few studies have achieved excellent performances, yet, validation on a larger dataset is necessary. All the mentioned studies have investigated the spatial domain characteristics of chest x-rays. To the best of our knowledge, no studies on COVID-19 from chest X-rays have reported investigating the spectral characteristics of the same. This motivates us to investigate the unexplored spectral aspect of chest x-ray towards the COVID-19 infection detection. We aim to validate the hypothesis of the presence of complementary information in the spectral and spatial domain of chest x-ray that will improve disease detection ability. We employ two popular tools for transforming images in spectral and spectral as well as spatial domains namely, Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) to study the spatio-spectral characteristics of the chest X-ray images. The idea behind the usage of DCT and DWT transformation is to capture any information which is complementary to spatial information and might aid in discriminating among the three

classes considered in this study. The DCT decomposes the image into several spectral sub-bands with cosine function as a basis function, whereas, DWT has the advantage of assimilating both spatial and spectral domains simultaneously. Hence, chest radiograph images are studied in three different domains concurrently, namely, pixel, DCT, and DWT to bring about the potential of each of the fields in characterizing the considered medical conditions. The patterns that characterize different medical conditions are complex and intricate. State-of-the-art CNN shows notable efficiency to model complex patterns in the domain of image classification. Hence, we employ ResNet50, a state-of-the-art CNN architecture to extract features from the pixel, DCT, and DWT domain images. The features extracted in these three domains are integrated and the final class prediction is performed using a multilayer perceptron (MLP) network. The detailed description of our proposed method is given in Section II. Medical imaging datasets often have a limited number of samples for comparatively uncommon disease classes. This class imbalance usually directs to inferior performances corresponding to the minority class. Relevant techniques such as class weight, unbiased validation performance metrics, has been employed to prevent such undesired outcome and are discussed in Section II in detail.

Furthermore, we extend our experiment to six other medical imaging datasets. These datasets incorporate a variety in types of imaging techniques, number as well as types of diseases to be detected, and class imbalance. The purpose of this extended study is to validate the hypothesis of the presence of complementary information in the spectral and spatial domain of medical images. This study overall indicates the generalization ability of the proposed method towards a diagnostic solution using medical images.

The contributions of this study are listed below.

- We propose an automatic computer vision and machine learning-based diagnostic solution for medical images developed on the complementary knowledge of the spatial and spectral domain.
- The proposed method is validated by carrying out experiments on eight diverse medical imaging datasets suggesting its robustness and capability of generalization.
- The classification performances along with saliency maps demonstrate the fusion of spatial, spectral, and spatio-spectral domain features enhances the disease detection capability of the classification model.
- Analyzing the classification performance on the COVID-19 dataset reveals that the performance of the proposed method is unbiased to age and gender factors.

The rest of the paper is structured as follows. Along with the detailed description of the datasets, the methodology used in this study is presented in Section II. The experimental results and relevant discussion is given in Section III. Section IV concludes our study along with the possible future directions.

TABLE 1. Number of samples and patients belonging to *Normal*, *Pneumonia* and *COVID-19* class.

|	Number of Samples	Number of Patients
Normal	8851	8851
Pneumonia	6052	6034
COVID-19	573	394
Total	15476	15279

II. METHODS AND MATERIALS

A. DATASET DETAILS

We validate our hypothesis primarily on a dataset comprising 15476 chest x-ray images from 15279 persons which belong to any three of the categories representing three types of medical conditions namely —*Normal*, *Pneumonia* and *COVID-19*. Images belongs to *COVID-19* category has been taken from these four sources [17]–[20] whereas, samples of *Pneumonia* and *Normal* class been collected from these sources [19], [21]. This dataset is referred COVID-19 dataset. Collecting samples from more than one source leads to diversity in the nature of the images. Thus, the experimental results on this dataset are robust and close to a practical scenario. The number of samples and the number of patients that belong to each of the three classes are stated in Table.1. Among the 15279 patients considered in this study, the ages and gender of 15029 and 15112 patients are known respectively. The age distribution of the subjects in each category and gender distribution in each class is shown in Figure.1. The dataset suffers from the problem of class imbalance. The imbalance ratio (ρ) defined by the ratio of the number of samples in the majority class to that of the minority class is 15.45. The chest X-ray images are of different sizes, thus in the preprocessing step, we resized it to the size of $224 \times 224 \times 3$ where the third dimension represents the number of channels. The results reported in this study are yielded by five-fold cross-validation. Out of the five folds, one is kept for testing, another for validation purposes while the rest of the folds are used to train the model. While partitioning the data into five-folds, it is always ensured that samples from one patient are never groped into two or more folds to maintain the integrity of the reported result.

Additionally, seven medical imaging datasets of different modalities e.g. X-ray, histopathology, mammography, etc, and medical conditions are also used for the validation of the hypothesis. The diversity in the number of classes i.e. the medical conditions, image modalities, and the imbalance ratio makes the experiment more robust. A detailed description of each of the datasets used in this study is presented in Tab. 2. We have used a subset of these original CBIS-DDSM [22], DR [23] and BHI [24] datasets as they comprise a large number of samples. Each dataset is split into training and testing samples maintaining an 80:20 ratio, except for Chestxray1 [9] where 100 test samples are provided for each of the three classes.

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

IEEE *Access*

TABLE 2. Detailed description of medical imaging datasets used in this study.

Dataset	Type of data	Disease Type	Number of classes	Number of Training samples	Imbalance ratio
CBIS-DDSM [22]	Breast mammography images	Abnormality in breast tissue	2	8941	6.55
DR [23]	Eye images	Diabetic Retinopathy	2	13052	1.33
Colorectal Histology [25]	Histology images	Colorectal cancer tissue type	8	4000	1
ISIC18 [26]	Dermoscopic images of skin	Skin disease	7	8166	58.86
Chestxray1 [9]	Chest x-ray images	COVID-19	3	13898	16.84
Chestxray2 [27]	Chest x-ray images	Pneumonia	2	5232	2.88
BHI [24]	Breast Histopathology images	Invasive Ductal Carcinoma	2	12952	2.79
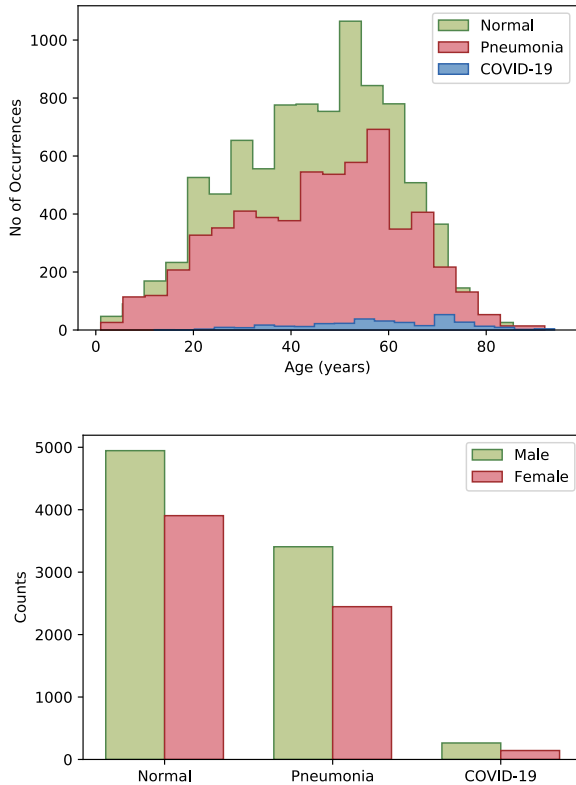
FIGURE 1. Age and gender distribution of persons belong to *Normal*, *Pneumonia* and *COVID-19* classes are presented in upper and lower panel of the figure respectively.

B. DISCRETE COSINE TRANSFORM

The Discrete Cosine Transform (DCT) of an image is a real transformation that transforms the image from spatial domain to frequency domain by linear combinations of weighted basis functions pertinent to its frequency components [28]. DCT of an image X of dimension $N \times N$ is given by the following equation.

$$DCT_x(u, v) = \frac{2}{N} C(u)C(v) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} X(m, n)$$
$$\times cos[\frac{\pi(2m+1)u}{2N}]cos[\frac{\pi(2n+1)v}{2N}], \quad (1)$$

where, $X(m, n)$ denotes the pixel value X in (m, n) coordinate, $u = 0, \ldots, N-1$, $v = 0, \ldots, N-1$ and

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}} & if \ k = 0, \\ 1 & otherwise. \end{cases}$$

In this study, images are of dimension $224 \times 224 \times 3$. DCT is applied to the images by considering each channel separately

over the segmented patches of size 8×8. Thus, the dimension of the DCT images is the same as the input images.

C. DISCRETE WAVELET TRANSFORM

The discrete wavelet transform (DWT) uses multiresolution filter banks to perform the wavelet analysis [29]. It represents the signal in terms of the wavelet coefficients from which it is possible to reconstruct the original signal once again. The signal is represented in various frequency bands by the wavelet coefficients. There can be several ways to process these coefficients, thus, endowing DWT with attractive properties over linear filtering.

The DWT is applied to the image X of size $(N \times N)$ to achieve four decomposed subband images of $(N/2 \times N/2)$ dimension. The process includes application of a set of half-band low pass and high pass filters to the rows of the image and followed by the decimation by a factor of 2. The same procedure is applied to the two subband images obtained in the previous step but this time along the columns. Thus, it results in decomposed images in four different frequency bands which can be mathematically expressed as follows,

$$LL(x, y) = \sum_m \sum_n h_0(m - 2x)h_0(n - 2y)X(m, n), \quad (2)$$

$$HL(x, y) = \sum_m \sum_n h_0(m - 2x)h_1(n - 2y)X(m, n), \quad (3)$$

$$LH(x, y) = \sum_m \sum_n h_1(m - 2x)h_0(n - 2y)X(m, n), \quad (4)$$

$$HH(x, y) = \sum_m \sum_n h_1(m - 2x)h_1(n - 2y)X(m, n). \quad (5)$$

Here, h_0 and h_1 are half band low and high pass filters respectively. Thus, LL subband approximates the image i.e. low-frequency content of the image, whereas, three other bands i.e. LH, HL, and HH subbands contain the details i.e. the high-frequency content of the image. The four subband images of size $N/2 \times N/2$ are arranged in the manner as shown in Figure 2 to form one image of size $N \times N$. Similar to DCT, DWT is also applied individually on each of the channels of images and the transformed image dimension is also the same as the input image dimension.

D. PERFORMANCE METRICS

Let, C be the confusion matrix for n-class classification where element C_{ij} indicate the number of samples of ith class predicted as of jth class. The following metrics presented in terms of the elements of the confusion matrix are used

IEEEAccess·

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics
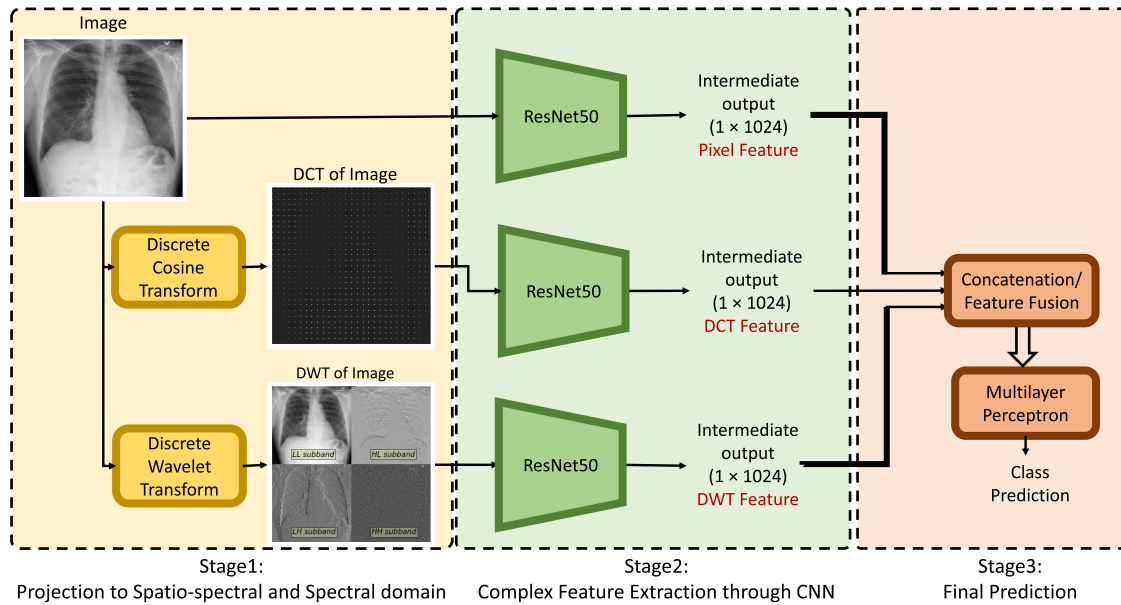

FIGURE 2. Schematic diagram of the classification framework.

to quantify the classification capability of machine learning models.

$$Sensitivity_i = \frac{C_{ii}}{\sum_j C_{ij}}, \tag{6}$$

$$Precision_i = \frac{C_{ii}}{\sum_j C_{ji}}, \tag{7}$$

$$F1\text{-}score_i = \frac{2 * Sensitivity_i * Precision_i}{(Sensitivity_i + Precision_i)}. \tag{8}$$

Along with the above-mentioned performance metrics, three other measurements are also used to quantify the performances of models — average class-specific accuracy/sensitivity (ACSA), average class-specific precision (ACSP), and an average class-specific F1-score (ACSF) respectively indicating sensitivity, precision, and F1-score averaged over all the classes, thus imparting equal priority to all the classes. Hence, these metrics quantify the model performances without being influenced by the existing class imbalance. The mean absolute deviation of the F1-score (MADF) is also reported along with ACSF to quantify the spread of individual class wise F1-scores.

$$MADF = \frac{1}{n} \sum_i |F1\text{-}score_i - \frac{1}{n} \sum_i F1\text{-}score_i|. \tag{9}$$

E. MACHINE LEARNING FRAMEWORK
The block diagram presented in Figure 2 shows an outline of the complete framework for the proposed diagnostic system. The proposed system consists of three stages — projection by using DCT and DWT, complex feature extraction, and prediction. In the first stage, the images are projected to spectral as well as spatio-spectral domain by the application

of DCT and DWT respectively as described in Section II-B and Section II-C respectively. Hence, at the end of this stage, we get three types of images of equal dimension in three different domains — pixel, DCT, and DWT (shown in Figure 2).

The objective of the second stage is to extract the features from output images of the previous stage. The underlying pattern that distinguishes one medical condition from another is intricate in nature. Convolutional Neural Network (CNN) can extract complex features from input images. ResNet50 is one of the widely used CNN architectures in the domain of image categorization. Thus, we choose to employ ResNet50 to serve the purpose of extracting complex features. Moreover, the reasonable number of trainable parameters of ResNet50 makes it suitable for medical image classification tasks where there are limited training samples available. Three separate ResNet50 models are trained for three different types of images. The output feature is extracted from an intermediate layer of trained ResNet50 models which results in complex features of dimension 1024. For training each of the ResNet50 models, the categorical cross-entropy loss function is optimized with the Adam optimization algorithm with a learning rate of 0.0001 for 300 epochs with a batch size of 32. The model with the best validation ACSA (defined in Section II-D) is chosen for feature mapping. ACSA is chosen for this purpose as this measure is unbiased to class imbalance present in the training data.

The third stage comprises feature level fusion followed by final prediction using a classification head. Three types of feature vectors, each of dimension 1024 are combined in all seven possible combinations by the concatenation operation. The dimension of the concatenated feature vector is presented in Table 3. The resulting feature vector is fed to

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

IEEE*Access*

TABLE 3. All possible combination of pixel, DCT and DWT features and corresponding integrated feature vector's dimension.

Combination Type	Dimension of Feature Vector
pixel	1×1024
DCT	1×1024
DWT	1×1024
pixel+DCT	1×2048
pixel+DWT	1×2048
DCT+DWT	1×2048
pixel+DCT+DT	1×3072

a classification head for final prediction. The Multilayer Perceptron (MLP) is chosen as the classification head because of its efficiency to model complicated nonlinear relationship between input and output vectors. The MLP network comprises one input layer, two hidden layers with 256 and 64 nodes, and one output layer. The leaky ReLU activation function is applied after each layer except the last layer which uses softmax as the activation function. To train the MLP network, the mean square error loss function is optimized with the Adam optimization algorithm with a learning rate of 0.0002 for 300 epochs with a batch size of 32. The model with the best training ACSA is evaluated with the test data.

The effect of class imbalance present in the datasets is handled by incorporating the penalty factor of class weights computed as $N/(c * N_i)$ to the loss function while training the ResNet50 model and MLP model, where c is the number of classes, N and N_i is the number of total samples and the number of samples that belong to i^{th} class respectively.

All the experiments (Python scripts) are executed using Keras with TensorFlow as backend on a computer with Intel core i5 processor running at 2.40 GHz using 16 GB of RAM and NVIDIA GeForce RTX 2060 GPU with 6 GB RAM.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present the results obtained by applying our proposed method to the medical imaging datasets.

A. RESULT ON COVID-19 DATASET

Due to the severity of the COVID-19 outbreak, the primary focus is paid on the elaborated study that is conducted COVID-19 dataset (described in Section 1). The result obtained from the above-mentioned experiment (Section 2) is presented in Table 4. The classification performances are quantified by class-wise sensitivity, precision, and F1-score. Additionally, we also present sensitivity, precision, and F1-score averaged over the different classes to measure overall performance. They are mentioned as ACSA, ACSP, and ACSF, and these metrics are defined in Section II-D. Table 4 compares the performance of the seven feature combinations as listed in Table 3. By comparing the performances of three types of features (pixel, DCT, and DWT) when used individually, pixel and DWT feature yields better results than DCT features. However, the fusion of any two types of features shows greater performance than any of the features while used solely for classification. Furthermore, the best classification performance is achieved when all three types of features i.e. pixel, DCT, and DWT are combined altogether (pixel+DCT+DWT). The corresponding value of ASCA, ACSP, ACSF are 93.78%, 91.30% and 92.42% respectively. It is important to note that the above-mentioned feature combination also yields the best sensitivity to the COVID-19 class which is 95.28% as we aim to detect COVID-19 infection with sensitivity as high as possible. These observations suggest the presence of complementary information in pixel, DCT, and DWT features. To validate the above-said statement from a statistical perspective, the Wilcoxon rank-sum test is performed on the ACSA, ACSP, and ACSF values. Table 5 summarises the result obtained from tests, where

TABLE 4. The classification performances all possible combination of pixel, DCT and DWT features in detection of *Normal*, *Pneumonia* and *COVID-19* classes are presented. Performances are quantified by sensitivity, precision. Average class specific sensitivity, specificity and F1-score (ACSA, ACSP, and ACSF) are also reported. All the performance metrics are reported on 5-fold cross-validation. Comparing among different feature combination, best value of each metrics are marked in bold.

Feature Combinations	Sensitivity(%)			ACSA	Precision(%)			ACSP	F1-Score(%)			ACSF
	Normal	Pneumonia	COVID-19		Normal	Pneumonia	COVID-19		Normal	Pneumonia	COVID-19	
Pixel	94.55 (±0.94)	90.26 (±0.65)	92.60 (±03.59)	92.47 (±01.49)	93.89 (±0.53)	92.04 (±01.25)	84.78 (±03.1)	90.24 (±01.4)	94.21 (±0.55)	91.14 (±0.75)	88.44 (±02.79)	91.26 (±01.30)
DCT	93.15 (±0.70)	87.21 (±01.72)	80.93 (±04.58)	87.10 (±01.21)	91.66 (±0.69)	89.40 (±0.96)	81.18 (±04.86)	87.41 (±01.6)	92.39 (±0.42)	88.27 (±0.68)	80.82 (±01.81)	87.16 (±0.76)
DWT	95.06 (±0.43)	89.75 (±0.98)	92.15 (±02.38)	92.32 (±0.88)	93.51 (±0.52)	92.79 (±0.56)	84.93 (±05.20)	90.41 (±01.72)	94.28 (±0.39)	91.24 (±0.63)	88.30 (±03.12)	91.27 (±01.14)
Pixel+DCT	95.18 (±0.56)	90.27 (±0.96)	93.25 (±03.41)	92.90 (±01.19)	93.93 (±0.56)	92.91 (±0.74)	85.34 (±03.65)	90.72 (±01.44)	94.55 (±0.49)	91.56 (±0.81)	89.02 (±02.41)	91.71 (±01.06)
Pixel+DWT	95.15 (±0.99)	**90.91** (±0.64)	94.86 (±02.75)	93.64 (±0.94)	**94.30** (±0.44)	93.04 (±01.23)	**86.39** (±04.91)	91.24 (±01.86)	94.72 (±0.58)	91.96 (±0.85)	90.32 (±02.47)	92.33 (±01.08)
DCT+DWT	95.31 (±0.55)	89.72 (±01.38)	93.03 (±01.89)	92.69 (±0.63)	93.51 (±0.82)	93.04 (±0.71)	86.45 (±03.21)	91.00 (±01.29)	94.40 (±0.52)	91.34 (±0.87)	89.58 (±02.25)	91.77 (±0.93)
Pixel+DCT+DWT	**95.49** (±0.85)	90.58 (±0.80)	**95.28** (±2.52)	**93.78** (±0.78)	94.13 (±0.49)	**93.51** (±1.08)	86.26 (±4.12)	**91.30** (±1.62)	**94.80** (±0.53)	**92.01** (±0.85)	**90.45** (±1.74)	**92.42** (±0.86)

IEEE *Access*

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

TABLE 5. Statistics and p-value obtain in Wilcoxon rank-sum test with performance of pixel+DCT+DWT feature combination is compared with that of the rest of the feature combinations. The statistics indicating superiority of pixel+DCT+DWT feature combinations are marked with *.

	ACSA		ACSP		ACSF	
	Statistics	p-value	Statistics	p-value	Statistics	p-value
Pixel	-5.04*	<0.05	-2.53*	<0.05	-4.37*	<0.05
DCT	-8.61*	<0.05	-2.82*	<0.05	-8.62*	<0.05
DWT	-7.05*	<0.05	-0.25	0.80	-4.82*	<0.05
Pixel-DCT	-4.55*	<0.05	-1.09	0.27	-2.67*	<0.05
Pixel-DWT	-0.12	.90	-8.50*	<0.05	-0.68	0.5
DCT-DWT	-6.38*	<0.05	-0.62	0.53	-3.03*	<0.05

the null hypothesis is that the performance of three features combined together (pixel+DCT+DWT) is equivalent to that of other combinations of the features. The null hypothesis is rejected when the p-value is less than 0.05, confirming that the samples belong to different distributions whereas a larger p-value suggests that the two distributions are similar. Furthermore, if test-statistics comes out to be negative with a p-value less than 0.05, we infer that the performance of pixel+DCT+DWT is superior to that of the other feature combination considered in the test. Table 5 shows that the test-statistics is never positive indicating performance of pixel+DCT+DWT is superior (in most of the cases, marked by *) or comparable (in a few cases) with the performance of the rest of the feature combinations. Thus, the outcome indicates that both DCT and DWT features capture information complementary to that of pixel feature that enhances the discriminating capability among the three classes considered in this study. Hence, our hypothesis of the existence of complementary information in the spatial and spectral domain of medical images is validated for the COVID-19 dataset.

The number of the hidden layers and the number of nodes in hidden layers of the MLP network are similar for all the combination of features types. However, the number of nodes in the input layers varies with the type of feature integration. For example, the number of nodes in the input layer of MLP is 1024 for pixel, DCT, DWT features, 2048 for Pixel+DCT, pixel+DWT, and DCT+DWT feature

combination, and 3072 for pixel+DCT+DWT feature combination. Consequently, the number of trainable parameters also increases with the number of nodes in the input layers. Thus, the question may arise that whether the superior performance of the combined pixel+DCT+DWT features is due to the integration of the features or the increased number of trainable parameters of the model. In search for the answer, we performed another experiment, where the considered architecture of the model is the same as it is used for pixel+DCT+DWT features. We construct the new feature vector for pixel by concatenating the pixel feature thrice so that the dimension of the new pixel feature vector becomes 1×3072 vector. A similar operation is applied to DCT and DWT features also. The performances of these new feature vectors are evaluated on the above-mentioned MLP individually. Table 6 compares these performances with pixel+DCT+DWT features (last row of Table 4). The test-statistics obtained from Wilcoxon rank-sum tests confirm the superiority of the performance of combined features concerning that of single features where the model complexity is kept unchanged.

1) INTERPRETING CLASSIFICATION MODEL

In this section, we attempt to understand how information from DCT and DWT domains contributes to enhancing the discriminative potential of the model. For this purpose, the saliency map i.e. the gradient of the class activation function concerning the input images is visualized. Thus, the saliency map for a particular class quantifies the amount of change in classification score caused by the small change in image pixel [30] indicating the decisive regions in the image. In Figure 3(a), the saliency map produced by the proposed model is shown for one image and its corresponding DCT and DWT images from each class. It confirms that all three types of images contribute to decoding the classes. Moreover, the saliency map of our model for each class validates its reliability as the highlighted regions from each of the images lie in the lung and its surrounding area. It is also noticed that the subband of DWT that represents the higher frequency component of the images does not contribute significantly towards disease detection.

TABLE 6. The performances of the new pixel, DCT, and DWT features evaluated on the MLP model that has similar architecture as used in the case of piexel+DCT+DWT features. The new pixel, DCT, and DWT feature vectors are constructed by concatenating each type of feature three times.

Feature	Sensitivity(%)			ACSA	Precision(%)			ACSP	F1-Score(%)			ACSF
	Normal	Pneumonia	COVID-19		Normal	Pneumonia	COVID-19		Normal	Pneumonia	COVID-19	
Pixel	94.84 (±1.22)	90.10 (±0.54)	93.65 (±3.80)	92.86 (±1.54)	93.96 (±0.23)	92.66 (±1.81)	81.85 (±4.17)	89.49 (±1.69)	94.39 (±0.61)	91.35 (±0.80)	87.30 (±3.28)	91.02 (±1.47)
DCT	93.08 (±0.72)	86.91 (±1.69)	83.64 (±3.55)	87.88 (±0.84)	91.72 (±0.58)	89.65 (±1.08)	77.69 (±7.11)	86.35 (±2.34)	92.38 (±0.41)	88.24 (±0.66)	80.25 (±2.92)	86.96 (±1.22)
DWT	94.86 (±0.81)	90.01 (±1.23)	92.42 (±2.82)	92.43 (±1.06)	93.77 (±0.65)	92.62 (±1.01)	83.31 (±5.76)	89.90 (±1.9)	94.30 (±0.42)	91.28 (±0.68)	87.53 (±3.81)	91.04 (±1.37)
Pixel+DCT+DWT	**95.49** (±0.85)	**90.58** (±0.80)	**95.28** (±2.52)	**93.78** (±0.78)	94.13 (±0.49)	**93.51** (±1.08)	**86.26** (±4.12)	**91.30** (±1.62)	94.80 (±0.53)	**92.01** (±0.85)	**90.45** (±1.74)	**92.42** (±0.86)

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

IEEE*Access*

TABLE 7. Comparison of performance of proposed method with COVID-net ([9]).

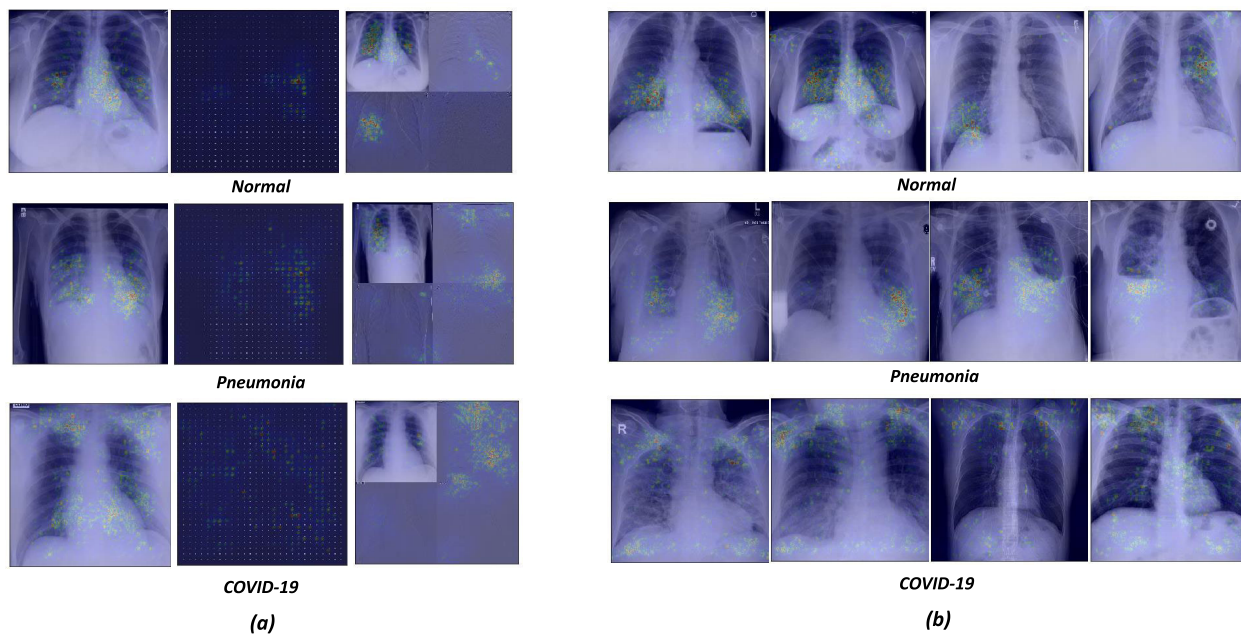|	Sensitivity(%)			ACSA	Precision(%)			ACSP	F1-Score(%)			ACSF
	Normal	Pneumonia	COVID-19		Normal	Pneumonia	COVID-19		Normal	Pneumonia	COVID-19	
COVID-net	**95.00**	**94.00**	91.00	93.33	90.50	91.30	**98.90**	93.57	92.70	**92.63**	94.79	93.37
Proposed	94.80	91.80	**94.30**	**93.63**	**91.60**	**92.55**	96.92	**93.69**	**93.17**	92.17	**95.59**	93.64

FIGURE 3. (a) Saliency maps corresponding to *Normal*, *Pneumonia* and *COVID-19* classes are presented for chest X-ray image(left panel), DCT image (middle panel) and DWT image (right panel). The highlighted regions play important role in characterizing the three classes. (b) Saliency maps of chest X-ray images sampled down from three classes are presented.

Moreover, we inspect the saliency map of chest X-ray images to learn if there is any apparent pattern corresponding to each class and correlate them with existing literature in the medical domain. Four such maps from each class are presented in Figure 3(b). While, saliency map of *Normal* class shows a wide variety in the area of chest X-ray that is highlighted, in case of *Pneumonia* class, the influential pixel cluster around the lower lobes of the both or either lung. On the other hand, along with the other lung regions, the upper lobe (bilateral) area of the lung is found to be persistently dominating in the case of *COVID-19* class. The bilateral Ground-glass opacities (GGO) have been reported in COVID-19 chest x-ray whereas the unilateral and central distribution of GGO has been found in chest x-ray of Pneumonia patients [11]. This characteristics are consistent with the patterns shown the saliency maps of *COVID-19* and *Pneumonia* classes.

2) EFFECT OF AGE AND GENDER ON CLASSIFICATION PERFORMANCE

The ages of the patients considered in this study varied from as low as 1 year to as high as 94 years Figure 1. To analyze if there is any biasedness of age on the classification performance, we carry out an analysis by clustering samples into five age groups – 0-20 years, 20-40 years, 40-60 years, 60-80 years, and 80-100 years. For each class, we investigated how the performance of the classifiers varies over different age groups. Figure 4(b) depicts the fraction of samples of a particular class classified as *Normal*, *Pneumonia*, and *COVID-19* classes. It is observed that over different age groups, the proposed method yields a similar trend suggesting that the age factor does not influence our classification results.

Furthermore, another similar analysis is executed to investigate the effect of gender of the patients on the classification result. The outcome of the analysis is shown in Figure 4(b) which demonstrates that the gender of the patients does not behave as a factor in classification performance. All the classification performances, reported in this section are yielded using the combined features (i.e. pixel, DCT, and DWT).

3) COMPARISON WITH COVID-NET

We compare our result with the state-of-the-art experimental result reported in [9]. For this purpose, we apply our methodology to the same dataset as mentioned in [9]. The comparison in terms of sensitivity, precision, and F1-score is
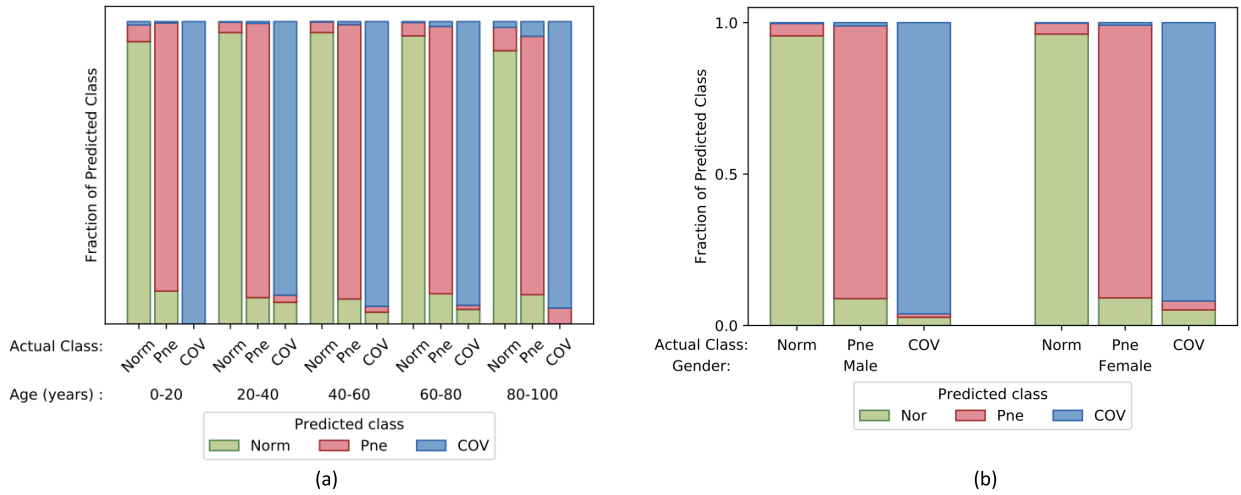
IEEE *Access*

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

FIGURE 4. Age wise and gender wise classification performance is shown in (a) and (b) respectively. *Norm*, *Pne*, and *COVD-* represent *Normal*, *Pneumonia*, and *COVID-19* respectively. The height of the three stacked up bars indicate the fraction of a particular class that belongs to three classes. For instance, the green, red and blue bars in the leftmost bar of panel (a) indicate the fraction of *Normal* samples (of 0-20 years age group) that is classified as *Normal*, *Pneumonia* and *COVID-19* class.

TABLE 8. The performances of pixel, DCT, DWT, and pixel+DCT+DWT in identifying different medical condition for different datasets as listed in Table 2.

Dataset	Pixel				DCT				DWT				Combined				
	ACSA	ASCP	ASCF	MADF	ACSA	ASCP	ASCF	MADF	ACSA	ASCP	ASCF	MADF	Combination Type	ACSA	ASCP	ASCF	MADF
BHI	89.32 (±0.53)	90.44 (±0.44)	89.85 (±0.28)	5.09 (±0.18)	87.15 (±0.51)	87.00 (±0.34)	87.07 (±0.21)	6.33 (±0.18)	85.52 (±0.45)	86.47 (±0.57)	85.97 (±0.33)	7.02 (±0.18)	pixel+DCT+DWT, SF	**89.86** (±0.3)	**90.98** (±0.2)	**90.40** (±0.17)	**4.81** (±0.1)
CBIS_DDSM	91.19 (±0.52)	**94.59** (±0.3)	92.79 (±0.26)	5.36 (±0.21)	77.58 (±1.55)	80.46 (±1.42)	78.82 (±0.43)	15.78 (±0.53)	86.60 (±0.89)	87.58 (±0.82)	87.05 (±0.26)	9.51 (±0.23)	pixel+DCT+DWT, FF	**91.98** (±1.24)	94.17 (±1.14)	**93.00** (±0.48)	**5.17** (±0.37)
CH	95.28 (±0.34)	95.42 (±0.25)	95.32 (±0.32)	3.34 (±0.2)	92.36 (±0.34)	92.38 (±0.42)	92.33 (±0.36)	5.07 (±0.42)	94.05 (±0.19)	94.03 (±0.19)	94.02 (±0.2)	4.04 (±0.28)	pixel+DWT, FF	**95.63** (±0.24)	**95.67** (±0.21)	**95.62** (±0.23)	**2.96** (±0.24)
CHESTXRAY1	89.97 (±0.46)	90.31 (±0.42)	90.00 (±0.45)	1.04 (±0.21)	87.33 (±0.47)	88.30 (±0.38)	87.34 (±0.49)	**0.82** (±0.24)	91.23 (±0.45)	91.37 (±0.46)	90.02 (±0.45)	91.95 (±0.22)	pixel+DCT+DWT, FF	**93.63** (±0.43)	**93.69** (±0.42)	**93.64** (±0.43)	1.44 (±0.34)
CHESTXRAY2	71.62 (±1.86)	86.31 (±0.58)	72.87 (±2.17)	12.44 (±1.37)	68.90 (±1.28)	85.41 (±0.37)	69.58 (±1.44)	14.57 (±0.89)	74.46 (±0.4)	88.00 (±1.51)	76.18 (±1.35)	10.49 (±0.93)	pixel+DWT, SF	74.68 (±1.28)	**88.08** (±0.37)	76.42 (±1.44)	**10.35** (±0.89)
DR	69.62 (±0.41)	69.42 (±0.35)	69.49 (±0.36)	4.24 (±0.3)	60.17 (±0.48)	60.28 (±0.35)	60.14 (±0.45)	6.67 (±1.51)	65.02 (±0.77)	64.79 (±0.58)	64.73 (±0.67)	4.19 (±1.3)	pixel+DCT, FF	**70.08** (±0.68)	**69.76** (±0.63)	**69.84** (±0.64)	**3.80** (±0.38)
ISIC18	75.89 (±0.63)	76.40 (±1.53)	75.86 (±0.87)	**9.49** (±0.44)	62.70 (±0.97)	66.97 (±0.89)	64.51 (±0.76)	13.93 (±0.82)	63.44 (±1.99)	65.81 (±1.36)	64.03 (±1.21)	15.17 (±0.74)	pixel+DCT+DWT, FF	**78.53** (±1.38)	**77.55** (±1.35)	**77.69** (±0.79)	9.69 (±0.75)

presented in table 7. The main aim of this study is to detect the COVID-19 class with as high sensitivity as possible and also maintain reasonably high precision at the same time. Though the value of the average class-specific accuracy, precision, and F1-score obtained in the proposed method is comparable with the result presented in [9], however, the higher sensitivity to *COVID-19* class of the trained model makes our proposed method more suitable for practical implementations.

B. RESULT ON ADDITIONAL DATASETS

We validate our proposed method on other medical imaging datasets as described in section 1. These datasets comprise different modalities, medical conditions as well as class imbalance ratios. Since the number of classes varies in the range from two to eight, instead of showing class-specific metrics, we present average class-specific metrics. Additionally, the mean absolute deviation of the F1-score (MADF) is used to quantify the spread of the individual class-wise F1-score. The lower the value of MADF indicates lower dispersion among the F1-score of each class. Table 8 presents the

classification performance for seven datasets listed in Table 2 for Pixel, DCT and DWT features individually. The efficiency of these features is also tested by performing feature level fusion (FF) and also score level fusion (SF). The best result obtained in feature level fusion or score level fusion is reported in Table 8. It is observed that the combination of two or three types of features among pixel, DCT, and DWT has surpassed the result obtained while using the each of features separately. Moreover, the lower MADF with higher ASCF of combined features also indicates the superiority in handling class imbalance. Thus, the result obtained in this experiment supports our hypothesis of the presence of complementary information in spatial and spectral domains in medical images that assist in decoding the underlying medical condition.

IV. CONCLUSION AND FUTURE DIRECTIONS

In this study, we present a novel medical image-based diagnostic solution for the detection of various underlying medical conditions. Along with the spatial information of the medical images, we exploited the less explored spectral-domain information of the medical images.

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

IEEE*Access*

We aspire to achieve an enhancement in disease detection performance by integrating the features from spatial, spectral, and spatio-spectral space. Our proposed system comprises three stages — conversion of a spatial image into the spectral and spatio-spectral domains, feature mapping from higher to lower dimension using CNN architecture, and final classification using MLP. The potential and the robustness of the proposed method are demonstrated by using eight diverse medical imaging datasets of different modalities, diseases to be identified, and class imbalance ratio.

The results suggest that the spatial, spectral and spatio-spectral domain features are solely capable of representing distinguishing characteristics of the underlying medical conditions. However, integration of those three types of features results in a significant increment in classification performance, suggesting that all three types of features possess complementary information. The saliency maps also validate the integrity of our proposed model. A detailed study on COVID-19 shows that the performance of our method is unbiased to gender and age factor. In comparison to the method proposed in [9] for COVID-19 detection, our method achieved significantly higher sensitivity to *COVID-19* class and also marginal improvement in average class-specific accuracy, precision, and F1-score. Altogether, our study demonstrates a novel approach in identifying various diseases using the medical images that can assist the healthcare worker in the primary screening process.

This study can be extended to other types of available medical images for generalization purposes. The implementation of other deep learning algorithms for feature extraction purposes can be explored towards the enhancement of the performance. Three separate ResNet50 networks are trained for extracting features from pixel, DCT, and DWT images. As a future research direction of this study, we consider overcoming this limitation. We have used an iterative training approach for updating the parameters of both CNN and MLP networks. The adaptation of non-iterative learning mechanism as used in Neural Networks with Random Weights (NNRW) [31]–[33], semi-random learning mechanism as used in Bidirectional stochastic configuration network (BSCN) [34] for updating the weights of the models should reduce the training time significantly. It will be interesting to study how the performance of the proposed method will be influenced by such non-iterative training. Future research direction may consider the interesting aspect of non-iterative training.

REFERENCES

[1] K. Suzuki, "Overview of deep learning in medical imaging," *Radiol. Phys. Technol.*, vol. 10, no. 3, pp. 257–273, 2017.

[2] D. Cozzi, M. Albanesi, E. Cavigli, C. Moroni, A. Bindi, S. Luvarà, S. Lucarini, S. Busoni, L. N. Mazzoni, and V. Miele, "Chest X-ray in new coronavirus disease 2019 (COVID-19) infection: Findings and correlation with clinical outcome," *La Radiologia Medica*, vol. 125, no. 8, pp. 730–737, Aug. 2020.

[3] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.

[4] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017.

[5] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Automated invasive ductal carcinoma detection based using transfer learning with whole-slide images," *Pattern Recognit. Lett.*, vol. 133, pp. 232–239, May 2020.

[6] M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, and U. R. Acharya, "Convolutional neural networks for multi-class brain disease detection using MRI images," *Comput. Med. Imag. Graph.*, vol. 78, Dec. 2019, Art. no. 101673.

[7] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, A. Bagul, C. P. Langlotz, and B. N. Patel, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLOS Med.*, vol. 15, no. 11, Nov. 2018, Art. no. e1002686.

[8] J. C. Souza, J. O. B. Diniz, J. L. Ferreira, G. L. F. da Silva, A. C. Silva, and A. C. de Paiva, "An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks," *Comput. Meth. Programs Biomed.*, vol. 177, pp. 285–296, Aug. 2019.

[9] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020.

[10] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.

[11] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103792.

[12] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Jun. 2020.

[13] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," 2020, *arXiv:2004.02696*.

[14] A. M. Ismael and A. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 114054.

[15] M. Toğaçar, B. Ergen, and Z. Cömert, "COVID-19 detection using deep learning models to exploit social mimic optimization and structured chest X-ray images using fuzzy color and stacking approaches," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103805.

[16] A. Gupta, S. Gupta, and R. Katarya, "InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using chest X-ray," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106859.

[17] A. Chung. (2020). *Figure1-COVID-Chestxray-Dataset*. Accessed: Jul. 14, 2020. [Online]. Available: https://github.com/agchung/Actualmed-COVID-chestxray-dataset

[18] A. Chung. (2020). *Actualmed-COVID-Chestxray-Dataset*. Accessed: Jul. 14, 2020. [Online]. Available: https://github.com/agchung/Figure1-COVID-chestxray-dataset

[19] J. Paul Cohen, P. Morrison, L. Dao, K. Roth, T. Q Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*.

[20] RSNA. (2020). *Covid-19 Radiography Database*. Accessed: Jul. 14, 2020. [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

[21] RSNA. (2018). *RSNA Pneumonia Detection Challenge*. Accessed: Jul. 14, 2020. [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data

[22] R. Sawyer-Lee, F. Gimenez, A. Hoogi, and D. Rubin. (2016). *Curated Breast Imaging Subset of DDSM*. [Online]. Available: https://wiki.cancerimagingarchive.net/x/lZNXAQ

[23] *Diabetic Retinopathy Detection | Kaggle*. Accessed: Aug. 11, 2021. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/overview

[24] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Informat.*, vol. 7, no. 1, p. 29, 2016.

[25] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Sci. Rep.*, vol. 6, no. 1, pp. 1–11, Sep. 2016.

IEEE *Access*

S. Ghosh *et al.*: Deep Learning Framework Integrating Spectral and Spatial Features for Image-Assisted Medical Diagnostics

[26] N. Codella, V. Rotemberg, P. Tschandl, M. Emre Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019, *arXiv:1902.03368*.

[27] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, and J. Dong, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[28] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-100, no. 1, pp. 90–93, Jan. 1974.

[29] M. Shensa, "The discrete wavelet transform: Wedding the a trous and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.

[30] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, "SmoothGrad: Removing noise by adding noise," 2017, *arXiv:1706.03825*.

[31] W. Cao, X. Wang, Z. Ming, and J. Gao, "A review on neural networks with random weights," *Neurocomputing*, vol. 275, pp. 278–287, Jan. 2018.

[32] W. Cao, P. Yang, Z. Ming, S. Cai, and J. Zhang, "An improved fuzziness based random vector functional link network for liver disease detection," in *Proc. IEEE 6th Int. Conf. Big Data Secur. Cloud (BigDataSecurity), Int. Conf. High Perform. Smart Comput. (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS)*, May 2020, pp. 42–48.

[33] W. Cao, L. Hu, J. Gao, X. Wang, and Z. Ming, "A study on the relationship between the rank of input data and the performance of random weight neural network," *Neural Comput. Appl.*, vol. 32, pp. 1–12, Jan. 2020.

[34] W. Cao, Z. Xie, J. Li, Z. Xu, Z. Ming, and X. Wang, "Bidirectional stochastic configuration network for regression problems," *Neural Netw.*, vol. 140, pp. 237–246, Aug. 2021.

SWAGATAM DAS received the B.E. and M.E. degrees in telecommunication engineering (control engineering specialization) and the Ph.D. degree from Jadavpur University, India, in 2003, 2005, and 2009, respectively. He is currently an Associate Professor and the Head of the Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, India. He has published more than 300 research articles in peer-reviewed journals and international conferences. He has more than 24,500 Google Scholar citations and an H-index of 74 till date. He has been associated with the international program committees and organizing committees of several reputed international conferences, including NeurIPS, AAAI, AISTATS, ACM Multimedia, BMVC, IEEE CEC, and GECCO. His research interests include evolutionary computing and machine learning. He was a recipient of the 2012 Young Engineer Award from the Indian National Academy of Engineering (INAE) and the 2015 Thomson Reuters Research Excellence India Citation Award as the Highest Cited Researcher from India in Engineering and Computer Science category, from 2010 to 2014. He is the Founding Co-Editor-In-Chief of *Swarm and Evolutionary Computation* and an international journal from Elsevier. He is also an Editorial Board Member of *Information Fusion* (Elsevier), *Progress in Artificial Intelligence* (Springer), *Applied Soft Computing* (Elsevier), *Engineering Applications of Artificial Intelligence* (Elsevier), and *Artificial Intelligence Review* (Springer). He has served/serving as an Associate Editor for the IEEE Transactions on Cybernetics, *Pattern Recognition* (Elsevier), *Neurocomputing* (Elsevier), *Information Sciences* (Elsevier), and IEEE Transactions on Systems, Man, and Cybernetics: Systems. He has acted as a Guest Editor for special issues in journals like IEEE Transactions on Evolutionary Computation and IEEE Transactions on Systems, Man, and Cybernetics, Part C.

RAMMOHAN MALLIPEDDI (Senior Member, IEEE) received the master's and Ph.D. degrees in computer control and automation from the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, in 2007 and 2010, respectively. He is currently an Associate Professor with the Department of Artificial Intelligence, School of Electronics Engineering, Kyungpook National University, Daegu, South Korea. He has coauthored articles published in IEEE Transactions on Evolutionary Computation (TEVC). His research interests include evolutionary computing, artificial intelligence, image processing, digital signal processing, robotics, and control engineering. He is currently serving as an Associate Editor for *Swarm and Evolutionary Computation*, an international journal from Elsevier, and a regular reviewer for journals, including IEEE Transactions on Evolutionary Computation (TEVC) and IEEE Transactions on Cybernetics (TCYB).

• • •

SUSMITA GHOSH received the B.E. degree in electronics and communication engineering from The University of Burdwan and the M.S. degree in biomedical signal processing from the Indian Institute of Technology Kharagpur. She is currently pursuing the Ph.D. degree with the Indian Statistical Institute, Kolkata. Her research interests include deep learning in medical imaging, biomedical signal, and image processing.