UncertaintyFuseNet: Robust Uncertainty-aware Hierarchical Feature Fusion with Ensemble Monte Carlo Dropout for COVID-19 Detection

Moloud Abdar, Soorena Salari, Sina Qahremani, Hak-Keung Lam, *Fellow, IEEE*, Fakhri Karray*, *Fellow, IEEE*, Sadiq Hussain, Abbas Khosravi, *Senior Member, IEEE,* U. Rajendra Acharya, *Senior Member, IEEE,* Saeid Nahavandi, *Fellow, IEEE*

***Abstract*—The COVID-19 (Coronavirus disease 2019) has infected more than 151 million people and caused approximately 3.17 million deaths around the world up to the present. The rapid spread of COVID-19 is continuing to threaten human's life and health. Therefore, the development of computer-aided detection (CAD) systems based on machine and deep learning methods which are able to accurately differentiate COVID-19 from other diseases using chest computed tomography (CT) and X-Ray datasets is essential and of immediate priority. Different from most of the previous studies which used either one of CT or X-ray images, we employed both data types with sufficient samples in implementation. On the other hand, due to the extreme sensitivity of this pervasive virus, model uncertainty should be considered, while most previous studies have overlooked it. Therefore, we propose a novel powerful fusion model named $UncertaintyFuseNet$ that consists of an uncertainty module: Ensemble Monte Carlo (EMC) dropout. The obtained results prove the effectiveness of our proposed fusion for COVID-19 detection using CT scan and X-Ray datasets. Also, our proposed $UncertaintyFuseNet$ model is significantly robust to noise and performs well with the previously unseen data. The source codes and models of this study are available at: https://github.com/moloud1987/UncertaintyFuseNet-for-COVID-19-Classification.**

***Index Terms*—COVID-19, Deep learning, Early fusion, Feature fusion, Uncertainty quantification.**

I. Introduction

The 2019 novel coronavirus (COVID-19) has been spreading astonishingly across the globe since its inception in Wuhan, China in December 2019 [1]. Overall, the COVID-19 worldwide pandemic has caused a consecutive series of catastrophic losses, from the sense of being healthy and safe to the social connections/bonds to the financial security. These all make the COVID-19 not only an epidemiological disaster, but also a psychological and emotional one. The uncertainties and grappling with the loss of normalcy in this pandemic provoke severe anxiety, stress and sadness among people.

Respiratory transmission of the disease from person to person occurred due to swift spread of the pandemic. While most of the COVID-19 cases showed milder symptoms, while the symptoms of the remaining cases are unfortunately illness or critical. The health-care systems in many countries including developed ones seem to have arrived to the point of collapse as the number of cases has been increasing drastically. With regard to the COVID-19 diagnosis, reverse transcription polymerase chain reaction (RT-PCR) is one of the gold standards for COVID-19 detection. However, RT-PCR yields low sensitivity. Hence, many COVID-19 cases will not be recognized by this test and the patients may not get the proper treatments. These unrecognized patients pose a threat to the larger population due to highly infectious nature of the virus. Chest X-ray (CXR) and Computed Tomography (CT) are widely utilized to locate the prominent pneumonia pattern of chest. These imaging modalities along with artificial intelligence tools may be applied to diagnose COVID-19 patients in more accurate, fast and cost-effective manner. Failure to prompt detection and treatment of COVID-19 cases may increase the mortality rate. Hence, the detection of COVID-19 cases using deep learning models with CXR and CT images have huge potential in healthcare applications.

Deep learning models have the widespread applicability not only in medical imaging paradigm but also in many other paradigms in recent years [2]–[5]. For COVID-19 detection and diagnosis, these models have been extensively applied. It is critical to discriminate COVID-19 from other forms of pneumonias and flus. Farooq et al. [6] introduced the open source code and an open access dataset with a CNN framework for distinguishing COVID-19 from analogous pneumonia cohorts by using chest X-ray (CXR) images. The authors designed COVIDResNet by utilizing a pre-trained ResNet-50 framework for improving model performance and reducing training time. An automatic and accurate identification

M. Abdar, A. Khosravi and S. Nahavandi are with the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia (e-mails: m.abdar1987@gmail.com, mabdar@deakin.edu.au, abbas.khosravi@deakin.edu.au & saeid.nahavandi@deakin.edu.au).

S. Salari and S. Qahremani are with the Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran (e-mails: soorena.salari374@gmail.com & sinaqahremani@gmail.com).

H.K. Lam is with the Centre for Robotics Research, Department of Engineering, King's College London, London, United Kingdom (e-mail: hak-keung.lam@kcl.ac.uk).

F. Karray is with the Centre for Pattern Analysis and Machine Intelligence, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada (e-mail: karray@uwaterloo.ca).

S. Hussain is with the System Administrator, Dibrugarh University, Dibrugarh, India (e-mail: sadiq@dibru.ac.in).

U. R. Acharya is with the Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Clementi, Singapore (e-mail: aru@np.edu.sg).

* Corresponding author: Fakhri Karray, karray@uwaterloo.ca

of COVID-19 using CT images help radiologists to screen patients in a better way. The authors in [7] proposed a fully automated system for COVID-19 detection by using chest CT. They dubbed their deep learning model as COVNet which investigated visual features from chest CT images. Moreover, Hall et al. [8] presented a new deep learning model termed as COVIDX-Net to aid radiologists for COVID-19 detection from CXR images. They explored seven deep learning architectures including DenseNet, VGG-19 and MobileNet v2.0, etc. In another study, Abbas et al. [9] devised Decompose, Transfer, and Compose (DeTraC) for COVID-19 classification using CXR images. A class decomposition approach was utilized to engage in irregularities in the image dataset by scrutinizing the class boundaries.
Segmentation also plays a key role in COVID-19 quantification measured by CT images. The authors in [10] proposed a novel deep learning method for segmentation of COVID-19 infection regions automatically. Aggregated Residual Transformations were employed to learn a robust and expressive feature representation and applied the soft attention technique to improve the potential of the system to distinguish several symptoms of the COVID-19. However, we noticed that there are still some open issues in the current proposed machine and deep learning methods for COVID-19 detection. For this reason, optimizing existing methods should be a priority in COVID-19 detection and classification. Ensemble and fusion-based methods [11] have shown outstanding performance for further improvement of different medical applications. In the following we provide more information about information fusion methods and how they can help to imprve the performance of deep learning methods.

A. Information fusion

Information fusion is initiated from data fusion and can be termed as multi-sensor information fusion [12], feature fusion for combining different features [13], various biological sources [14], [15], medical signals [16] or medical image fusion [17]–[19]. The data fusion methods were largely used for military applications. Their purpose was to integrate or correlate data of several sensors in different or same type to achieve better results than that yielded by a single sensor. Gradually, data fusion models are converted into information fusion. Information fusion does not rely on only multi-sensor data, but its area of research and applications have drastically altered. The rapid growth of network technology has made it possible for the information fusion technology to change from centralized single node information fusion in to distributed information fusion. There are numerous studies on big data fusion and image fusion methods.
Modern medicine nowadays depends on amalgamation of data and information from manifold sources that includes structured imaging data, laboratory data, unstructured narrative data, and even observational or audio data in some occasions [20]. Substantial clinical context is required for medical image interpretation to facilitate diagnostic decision [21]. The imaging data and its significant impact is not only limited to radiology but also on many other image-based medical specialties such as dermatology, ophthalmology and pathology [22]. Unstructured and structured clinical data from the electronic heath record (EHR) is crucial for clinically relevant medical image interpretation [23]. Clinically relevant models rely on automated diagnosis and classification systems that use both clinical data from EHR and medical imaging data. Multimodal learning models employ various imaging data with other data types (fusion) have in various applications such as video classification, autonomous driving and medical data analysis. The current medical imaging paradigm showcases a drift where both pixel and EHR data are utilized in fusion-domain for tackling complicated tasks that cannot be resolved by single modality. A wide variety of fusion techniques have been applied to a numerous of machine and deep learning techniques. Machine and deep learning models have shown their efficacy in many well-known benchmark datasets in the last several years. This facilitates an increasing interest in several domains and non-traditional issues each of which has their specific prerequisites. In medicine, customized predictions carry significant meaning as incorrect decisions are associated with severe costs due to associated ethical concerns and risk to human life [24]. Prevailing deep neural networks (DNNs) in medicine which use one or an ensemble of models focus on enhancing the accuracy of probabilistic predictions. Model uncertainty that is inherent in fitting DNNs is not well addressed by the models while they capture only data uncertainty. For example, when in an intensive care unit (ICU) setting, mortality of a patient is predicted, the state-of-the-art (SOTA) methods may be able to yield high AUC-ROC, but will be unable to discriminate between those for whom the model is certain about its probabilistic prediction and patients for whom the model is fairly uncertain. Hence, there is a need for examining the utilization of model uncertainty specifically in the context of predictive medicine. Model uncertainty achieved several methodological advances in recent years—including function priors, deep ensembles and efficient alternatives, Monte Carlo (MC) dropout and reparameterization-based variational Bayesian neural networks (BNNs). Several clinical care problems that naturally transpire in predictive medicine are addressed by the DNNs integrated with model uncertainty techniques.

B. Uncertainty quantification (UQ)

Many machine and deep learning models have been developed not only for CXR and CT images but also for many other medical applications and yielded great accuracies with limited number of images [5]. However, DNNs require a large number of data to fine tune trainable parameters. Limited number of images lead to epistemic uncertainty. Trust is an issue for such models deployed with less number of training samples. Out of distribution (OoD) samples or discrimination between the training and testing samples will make such models fail in real world applications. Lack of confidence in unknown or new cases are not reported by these models. Such information is essential for the development of reliable medical diagnostic tool. These unknown samples which are hard to predict have more practical values. It is very essential to estimate uncertainties with an extra insight to their point estimates. This additional vision aims to enhance the overall trustworthiness in the systems such that clinicians or users will come to know when and where they can trust predictions made by these models. The flawed decisions made by these models could be fatal for the patients at risk. Hence, proper uncertainty estimations are essential to improve the efficacy of the model and apply to medical domain with trust and reliability.

Proper UQ is crucial to enhance interpretability and trust in machine and deep learning models, which is significantly important in various healthcare applications [25], [26]. Trustworthy uncertainty estimations can facilitate informing clinical decision making, and more importantly, prepare clinicians with appropriate feedback on when to relinquish automatically obtained results. As discussed above, COVID-19 has had many negative effects on all aspects of human life around the world so far. Most importantly, COVID-19 disease has caused thousands of deaths worldwide. In this regard, this study, therefore, attempts to propose a simple accurate and robust deep learning model for COVID-19 disease detection called $UncertaintyFuseNet$. Due to the high susceptibility of this disease and in order to increase trust in the obtained results, we include uncertainty quantification method.

C. Main Contributions

The major contributions of this study are listed as follows:

- Proposed a novel feature-based fusion model for accurate detection of COVID-19 cases called $UncertaintyFuseNet$.
- Presented quantify uncertainty in our proposed fusion model using recently ensemble MC (EMC) dropout.
- Developed fusion model demonstrates strong robustness to data contaminations (data noise).
- Generated model yielded excellent results in terms of unknown data detection.

The rest of this study is organized as follows. Section II summarizes a few relevant studies. Section III formulates the proposed methodology. The main experiments of this study is discussed in Section IV. Section V presents the achieved outcomes and provides a comprehensive comparison with previous studies. Finally, the paper concludes in Section VI.

II. Literature Review

In this section, we will briefly review a few recent studies conducted on information fusion on disease identification, COVID-19 detection/segmentation as well as the importance of UQ in medical image analysis.

A. Information Fusion in Medical Systems

In this sub-section, we briefly summarized a few recent information fusion-based methods developed for the accurate diagnosis of diseases. Smart health care systems utilizing body sensor data for pattern analysis and a wide range of human-computer interaction have attracted wide attention from academics. Uddin et al. [27] employed deep Recurrent Neural Network (RNN) to design a body sensor-based framework for the identification of behaviour. They fused data from various body sensors like magnetometer, accelerometer, electrocardiography (ECG), etc. In another research, Chen et al. [15] provided a medical artificial intelligence approach based on data width evolution and self-learning to facilitate skin ailment health service gathering the necessity of real time, individualization and extendibility. Firstly, remote medical data and in the closed-loop flow of medical user data in large quantities were obtained in this research. In the next step, to improve the learning ability of cloud remote analysis model and to lessen the load of edge node, an information entropy-based data set filter method was formulated.

Quellec et al. [28] proposed a new content-based heterogeneous information retrieval approach to browse particularly medical data and support CADx (Computer Aided Diagnosis) systems. Incomplete documents comprising of semantic information and several images could be retrieved using the developed technique. Their approach could handle complex data types such as video. They also devised two new information fusion techniques to integrate degrees of match for ranking the reference documents. They evaluated their technique on two medical databases and yielded promising results. Wu et al. [29] introduced a personalized interactive simulator which modelled COVID-19 pandemic via multi-source information fusion. It is vital in terms of outbreak prediction and for its dynamics and effects estimation for a novel pandemic like COVID-19. There were three challenges in this– complexity in programming, roughness in models and uncertainty in data. Moreover, Muzammal et al. [30] proposed an ensemble approach based on data fusion in a fog computing setting using medical data from Body Sensor Network (BSNs). A collection of sensors was used to obtain daily activity data that were fused to used and fed as an input to an ensemble model for the classification of heart diseases. Their empirical results yielded an accuracy of 98% using 8 features and 40 estimators.

B. COVID-19 Classification/Segmentation

It is crucial to recognize COVID-19 infected cases quickly to better manage and prevent the pandemic from further spreading. Ardakani et al. [31] analyzed 108 COVID-19 pandemic patients, viral pneumonia and other atypical patients using CT scan images. They applied ten CNN approaches to discriminate COVID-19 group from non-COVID-19 cohorts. Xception and ResNet-101 demonstrated superior performance in all networks with an AUC value of 0.994 for both cases. Deep learning model can assist the clinicians and radiologists utilizing CXR scans for the detection of COVID-19. Khan et al. [32] introduced CoroNet, a deep CNN, for such automated detection. Xception architecture was used for pre-training and two publically

available X-ray datasets were used for the classification of normal, pneumonia and COVID-19 X-ray radiographs. Their model yielded an accuracy of 95% for 3-class (normal vs Pneumonia vs COVID) classification. In addition, their model demonstrated an overall accuracy of 89.6% for 4-class (Pneumonia bacterial vs pneumonia viral vs normal vs COVID) cases. Coronet exhibited promising outcome with minimal pre-processing of data. Chimmula et al. [33] used modern deep learning methods to design a COVID-19 forecasting model (the Long short-term memory (LSTM) method) for the outbreak in Canada based on publically available Canadian health authority and John Hopkins University datasets. They scrutinized the vital features to forecast the probable stopping time and trends of the pandemic.
Afshar et al. [34] devised an approach based on Capsule networks (COVID-CAPS) and produced efficient results with smaller X-Ray datasets. Their framework exhibited better performance than the existing CNN-based models. COVID-CAPS showed AUC value of 0.97, specificity of 95.8%, sensitivity of 90%, and accuracy of 95.7% while dealing with less number of parameters than its counterparts. Transfer learning and pre-training was used to enhance diagnostic nature of the framework further and tested with a new X-Ray dataset. The utilization of artificial intelligence (AI) to exploit CXR images for patient triage and diagnosis of COVID-19 is of supreme importance. Lack of systematic collection of CXR data set for training of deep learning strategies hindered the proper diagnosis. Oh et al. [35] presented a patch-based CNN technique with less number of trainable parameters for COVID-19 diagnosis to address the issue. Their statistical analysis of potential imaging biomarkers of CXR images was the inspiration for the proposed task. Punn et al. [36] proposed the weighted class loss function and random oversampling methods for transfer learning in different SOTA deep learning methods. They used posteroanterior CXR images for multiclass classification: (pneumonia, COVID-19, and normal cases) and binary classification (COVID-19 and normal cases). Empirical results demonstrated that each model was scenario dependent; NASNetLarge showed better scores in comparison to its counterparts.

C. Uncertainty Quantification in Medical Image Analysis

There are numerous studies conducted on Uncertainty Quantification (UQ) for medical image analysis using machine and deep learning methods. In this section, we briefly introduce a few recent studies which applied UQ methods for medical image analysis. Deep CNNs seldom facilitate uncertainty estimations regarding medical image segmentation e.g., image-based (aleatoric) and model (epistemic) uncertainties despite delivering the SOTA performance. Wang et al. [37] examined different types of uncertainties related with 3D and 2D medical image segmentation tasks at both structural and pixel levels. Moreover, they introduced test-time augmentation-based aleatoric uncertainty to measure the effect of various transformations of the input image on the output segmentation. MC simulation with prior distributions of parameters estimated a distribution of prediction in an image acquisition model with noise and image transformations was provided to formulate test-time augmentation.
There are two commensal and traditional tasks *i.e.*, direct ventricle function index estimation and bi-ventricle segmentation are assigned to tackle ventricle quantification issue. Luo et al. [38] introduced a unified bi-ventricle quantification approach based on commensal correlation between the direct area estimation and bi-ventricle segmentation. They also devised a new deep commensal network (DCN) to combine these two commensal tasks into a unified framework based on the proposed commensal correlation loss. The DCN yielded fast convergence and end-to-end optimization as well as uncertainty estimation with one-time inference. Colorectal cancer which is one of the prime reasons of cancer-related fatalities around the globe and its key precursor is colorectal polyps. Convolutional Neural Networks (CNNs) based decision support systems for segmentation and detection of colorectal polyps showed outstanding performance. In another research, Ghoshal et al. [39] used drop-weights based Bayesian CNN (BCNN) to measure uncertainty in deep learning methods to enhance the diagnostic performance. They demonstrated that accuracy of the prediction was highly correlated with the uncertainty in prediction.

D. Research Gaps

Comprehensive literature review identified several important research gaps in COVID-19 detection/segmentation. We list a few most important ones as follows:

- There are no sufficient COVID-19 data to develop accurate and robust machine and deep learning methods, which hinders This can impact the performance of such methods.
- There are very few studies that have used both types (CT scan and X-Ray) of images simultaneously.
- To the best of our knowledge, there is no study that has examined the uncertainty of model in predicting the COVID-19 cases.
- Moreover, we found that there were very few studies considering robustness and unknown data detection for COVID-19 classification.
- The impressive effect of different fusion methods have received less attention in the COVID-19 studies. It can be noted that these methods are very effective both to improve the performance and also to deal with the uncertainty with ML and DL models (as a kind of ensemble).

III. Proposed Methodology

This section has two main sub-sections: basic deep learning methods in sub-section III-A and our proposed fusion model: $UncertaintyFuseNet$ in sub-section III-B. It may be noted that we also applied two traditional machine learning algorithms (*i.e.*, $Random\ Forest$ (RF) and

Decision Tree (DT, max-depth=50 and n-estimators=200)) and compared their performances with the applied deep learning methods.

A. Basic Deep Learning Methods

In this sub-section, we provide more details regarding two basic deep learning methods: deep 1 (Simple CNN) and deep 2 (Multi-headed). Figs. 1 and 2 show deep 1 (Simple CNN) and deep 2 (Multi-headed), respectively. The first deep (Simple CNN) includes three convolutional layers followed by MC dropout in the feature extraction layer. The extracted features are then given to the classification layer, including three dense layers and MC dropout. More details of deep 1 (Simple CNN) can be found in Fig. 1. In the second deep learning, *i.e.*, Multi-headed, three main heads (as feature extractors) are included. The extracted features in each branch are then given to the fusion layers followed by the classification layer as illustrated in Fig. 2.

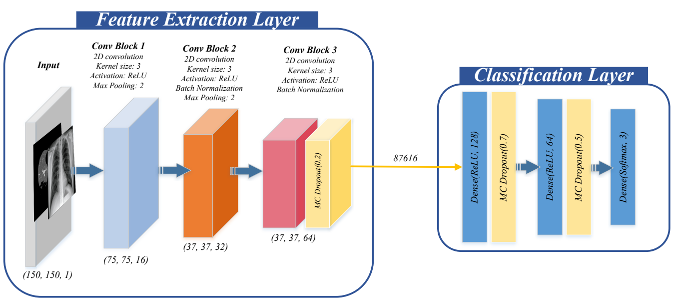


Fig. 1: A general overview of the applied deep learning 1 (Simple CNN).

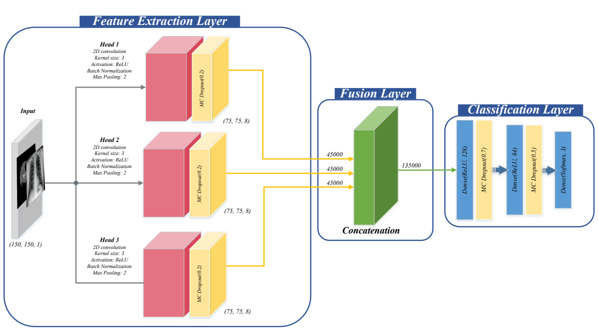


Fig. 2: A general overview of the applied deep 2 (Multi-headed).

B. Proposed Fusion Model: *UncertaintyFuseNet*

Feature fusion is an approach used to combine features (different information) of the same sample (input) extracted by various methods. Assume $\Omega = \{\xi \mid \xi \in \mathbb{R}^N\}$ be a training sample (image) space of m labeled samples (images). Given $A = \{x \mid x \in mathbbR^p\}$, $B = \{y \mid y \in \mathbb{R}^q\}$, ..., and $Z = \{n \mid n \in \mathbb{R}^k\}$, where x, y, ..., n are the feature vectors of the same input ξ sample extracted by various deep learning methods, respectively. Therefore, the total feature fusion vector space D_{vs}^{ff} obtained from different sources can be calculated as follows:

$$D_{vs}^{ff} = Concat[A, B, ..., Z], \tag{1}$$

In this study, after preprocessing the data, we feed our data set to the model. Our model consists of two major branches: The first branch has five convolutional blocks. Each block is made up of two tandem convolutional layers followed by batch normalization and max-pooling layers. Also, the fourth and fifth blocks have dropout layers in their outputs. The second branch is a VGG16 transfer learning network whose output will be used in the fusion layer. After two branches, the model is followed by a fusion layer that concatenates the third, fourth, and fifth convolutional layers' outputs with VGG16's output.

Finally, we used fully connected layers to process the fusion features and classify the data. In this part, we have used four dense layers with 512, 128, 64, and 3 neurons with ReLU activation function, respectively. The first three dense layers' output have a dropout in their outcomes with a rate equal to 0.7, 0.5, and 0.3, respectively.
The stated model is not simplistic. Indeed, to boost the model's power in dealing with data and extracting high-quality features, we have employed a novel feature fusion approach obtained from different sources:

- We selected the third convolutional block's output as a fusion source to have a holistic perspective about the data distribution. These features help the model to consider the unprocessed and raw information and use them in the prediction.
- We included the final and penultimate convolutional blocks' outputs in the feature fusion layer to have more accurate information. This feature gives a detailed view of the dataset to model and helps the model to process advanced classification features.
- As pre-trained networks have been widely used in pneumonia detection in recent studies and can create high-quality and generalizable features, we used the output of VGG16 in the fusion layer.

The pseudo-code of the proposed *UncertaintyFuseNet* model for COVID-19 detection is shown by **Algorithm 1** and its general view in Fig. 3.

After training our models, to generate final prediction, we have averaged the predicted softmax probabilities in N random forward paths of data x through *UncertaintyFuseNet* and stochastic sampling dropout mask $\mathbf{w}_t$ for each path:

$$\hat{\mathbf{y}}_t = \text{Softmax}\left(\mathit{UncertaintyFuseNet}(\mathbf{x}; \mathbf{w}_t)\right), \tag{2}$$

$$\hat{\mathbf{y}}_* = \frac{1}{N}\sum_{0}^{N} \hat{\mathbf{y}}_t. \tag{3}$$

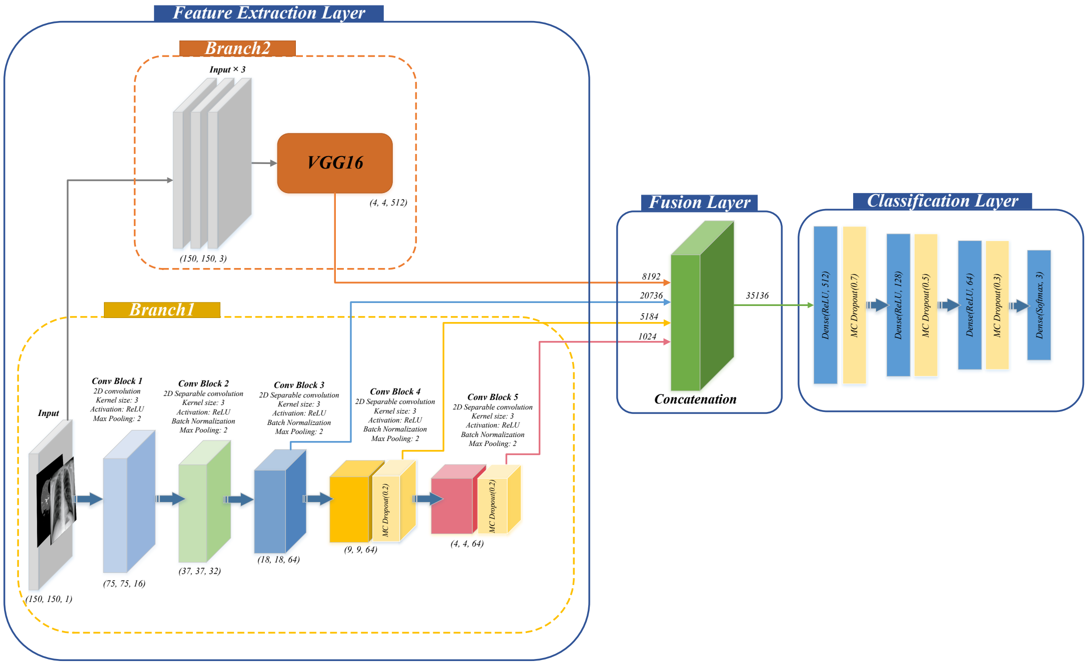


Fig. 3: A general overview of the proposed $UncertaintyFuseNet$ model inspired by a novel hierarchical feature fusion and MC dropout.

Algorithm 1: Pseudo-code of the proposed fusion model ($UncertaintyFuseNet$)

Input: A gray-scale image.
Output: COVID-19 classification with higher certainty.
Feature Extraction Layer: Branch1:
Conv1 ← *First Convolutional Block* ← *Input Image*
Conv2 ← *Second Convolutional Block* ← *Conv1*
Conv3 ← *Third Convolutional Block* ← *Conv2*
Conv4 ← *MCDropout(rate=0.2)* ← *Fourth Convolutional Block* ← *Conv3*
Conv5 ← *MCDropout(rate=0.2)* ← *Fifth Convolutional Block* ← *Conv4* **Branch2:**
VGG features ← *VGG16 Block* ← *Input Image*
Fusion Layer:
X ← Concatenation of *(Conv3, Conv4, Conv5, VGG features)*
Classification Layer:
X ← *Dense(X, units=512, activation=ReLU)*
X ← *MCDropout(X, rate=0.7)*
X ← *Dense(X, units=128, activation=ReLU)*
X ← *MCDropout(X, rate=0.5)*
X ← *Dense(X, units=64, activation=ReLU)*
X ← *MCDropout(X, rate=0.3)*
Output ← *Dense(X, units=3, activation=softmax)*

We then used model ensembling and acquired predictions from the N trained model with various weight distributions and initialized weights by this strategy. Hence, we obtained a meaningful improvement in our achieved results.

$$Predicted\_Class = \text{Argmax}(\hat{\mathbf{y}}_*). \quad (4)$$

The pseudo-code of the applied ensemble MC dropout in our proposed $UncertaintyFuseNet$ model for COVID-19 detection is given in **Algorithm 2**.

Algorithm 2: Ensemble MC dropout

Predictions = 0
for $k = 1$, $k{+}{+}$, while $k < i$ **do**
  Probability ← *UncertaintyFuseNet* ← *Input Image*
  Predictions ← *Predictions* + *Probability*
Predictions ← *Mean(Predictions)*
Predicted Class ← *Argmax(Predictions)*

IV. Experiments

In this section, we provide more information about the datasets in IV-A, the obtained results using the proposed methods are illustrated in IV-B, robustness against noise is presented in IV-C, and unknown detection is shown in IV-D.

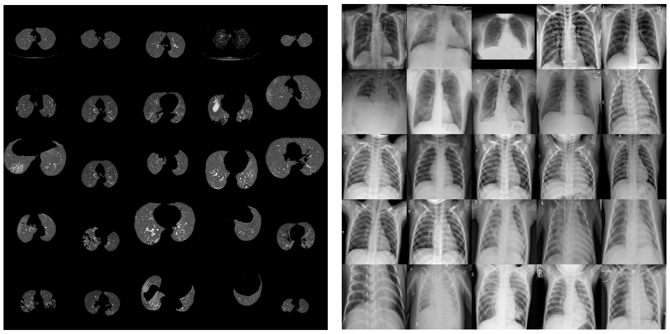
A. Datasets

In this study, we used two datasets: CT scan [1] and X-Ray [2]. More information about each dataset is presented in Table I. In addition, few random samples of both CT scan and X-Ray datasets used in this study is showed in Fig. 4. The CT scan dataset has three main classes: non-informative CT (NiCT), positive CT (pCT), and negative CT (nCT). Moreover, the X-Ray dataset also has three main classes: COVID-19, Normal, and Pneumonia.

[1]Sources: https://www.kaggle.com/azaemon/preprocessed-ct-scans-for-covid19,http://ictcf.biocuckoo.cn/
[2]Sources: https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia

TABLE I: The COVID-19 datasets used in this study.

Dataset	# of Samples	# of Classes
CT scan	19685	3
X-Ray	6432	3



(a) CT scan (b) X-ray

Fig. 4: Some random samples from both CT scan and X-Ray datasets are used in this study.

B. Experimental Results

In this section, the experimental results are presented and discussed. Since we considered the impact of UQ methods, the experiments are conducted without applying the UQ method for CT scan and X-Ray datasets. In the first step, we used five different algorithms, including Random Forest (RF), Decision Tree (DT, max-depth=50, and n-estimators=200), Deep 1 (Simple CNN), Deep 2 (Multi-headed), and our proposed model (Fusion model) to evaluate the performance.

*1) **COVID-19 classification without considering uncertainty**:* First, we investigated the performance of five classifiers(RF, DT, simple CNN, multi-headed and our proposed fusion model) without considering uncertainty: using two COVID-19 datasets (CT scan and X-Ray). The obtained results are presented in Tables and II and III for CT scan and X-Ray datasets, respectively. It may be noted from Table II that, our proposed fusion model outperformed the other applied methods using CT scan dataset with a test accuracy of 99.136%, followed by Simple CNN with a test accuracy of 98.763%. The results also indicate that DT has obtained the weakest performance with CT scan dataset among the five algorithms used. Figs. 5 and 6 show the confusion matrix and ROC using different CT scan dataset without quantifying uncertainty, respectively.

To demonstrate the effectiveness of our proposed fusion model, the same methods tested on CT scan dataset are applied to the X-Ray dataset too. We noticed that the results obtained using X-Ray dataset are different from the CT scan dataset. It may be noted from Table III that, our proposed fusion model performed far better than the other applied methods with an accuracy of 97.127%, followed by multi-headed model with an accuracy of 94.953%. Using the CT scan dataset, it can be seen that, DT performed worse on automated COVID-19 detection using X-Ray dataset with an accuracy of 84.006%. Figs. 7 and 8 indicate the confusion matrix and ROC obtained using three deep learning methods with X-Ray dataset and without quantifying uncertainty, respectively.

TABLE II: Comprehensive comparison of results obtained using different CT scan dataset methods for COVID-19 detection without considering uncertainty (%).

Method	Precision	Recall	F-Measure	Accuracy
RF	97.111	97.070	97.091	97.070
DT	93.049	93.040	93.045	93.040
Deep 1 (Simple CNN)	98.787	98.763	98.775	98.763
Deep 2 (Multi-headed)	98.599	98.577	98.588	98.577
Proposed (Fusion model)	**99.137**	**99.136**	**99.136**	**99.136**

TABLE III: Comprehensive comparison of results obtained using different methods for COVID-19 detection using X-Ray dataset without considering uncertainty (%).

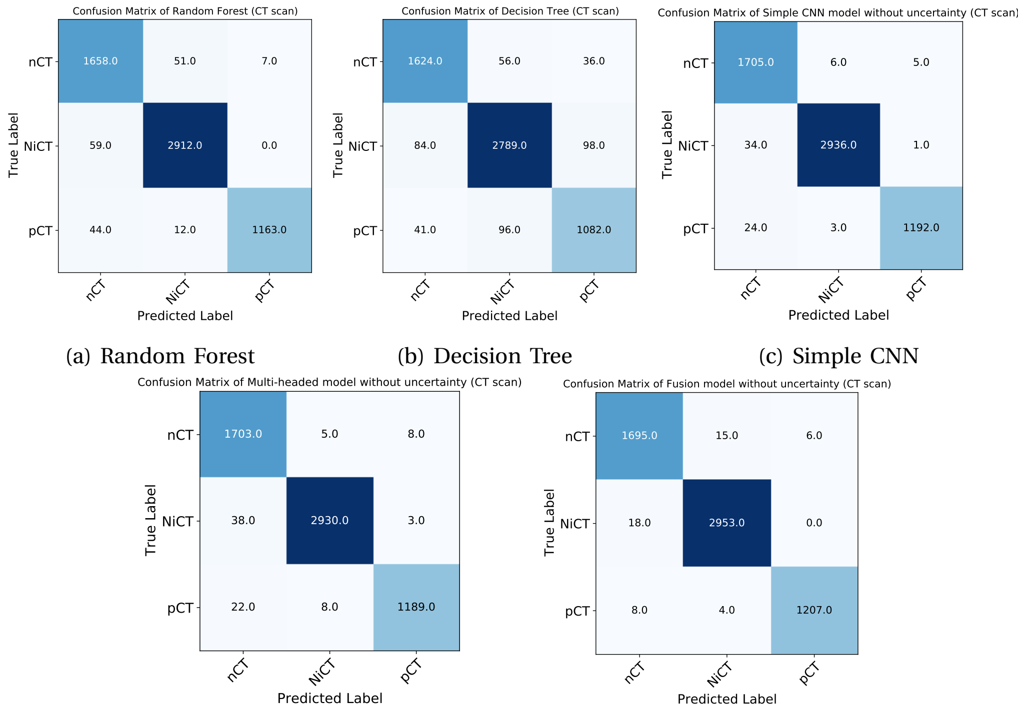
Method	Precision	Recall	F-Measure	Accuracy
RF	91.532	91.381	91.456	91.381
DT	83.828	84.006	83.917	84.006
Deep 1 (Simple CNN)	93.847	93.167	93.506	93.167
Deep 2 (Multi-headed)	95.041	94.953	94.997	94.953
Proposed (Fusion model)	**97.121**	**97.127**	**97.124**	**97.127**

*2) **COVID-19 classification with considering uncertainty**:* The obtained results of the previous sub-section (IV-B1) are promising hence, our proposed fusion model can be used in the clinical domain. We also believe that providing intelligent (*i.e.* ML and DL) models to deal with this serious virus is urgently needed. But at the same time, we firmly believe in the transparency of the results obtained by such intelligent-based models with certainty. To accomplish this, we applied the uncertainty quantification method called EMC dropout with the uncertainty of the deep learning models, *i.e.*, deep 1 (simple CNN), deep 2 (multi-headed), and our proposed fusion model. Hence, these three deep learning models are fed with both CT scan and X-Ray datasets.

Table IV and Figs. 9 and 10 show the obtained results of three applied deep learning methods on COVID-19 classification using CT scan dataset with uncertainty. It may be noted from Table IV that our fusion model has obtained more promising performance than other two models with an accuracy of 99.085% followed by simple CNN with the accuracy of 98.831% using CT scan dataset. Comparison of results obtained using deep learning models without and with UQ method reveal that our proposed fusion model with UQ method has yielded slightly poorer performance than the model without UQ. However, simple CNN with UQ performed better as compared to when it is applied without UQ method.

TABLE IV: Comprehensive comparison of results obtained using different CT scan methods for COVID-19 detection t with uncertainty (%).

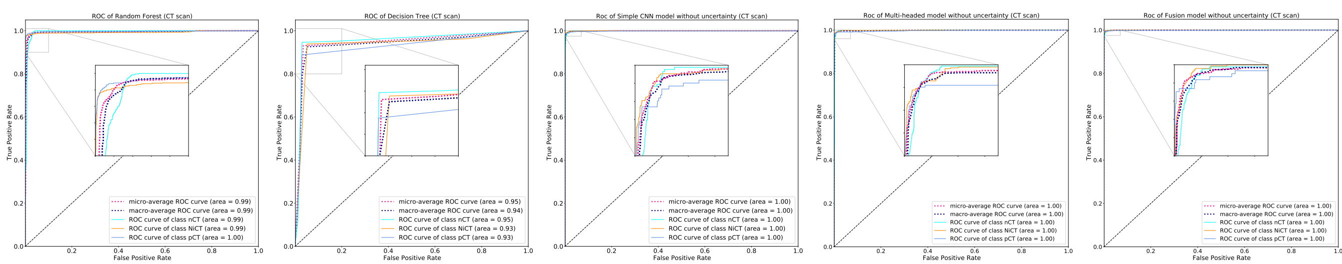
Method	Precision	Recall	F-Measure	Accuracy
Deep 1 (Simple CNN)	98.831	98.854	98.843	98.831
Deep 2 (Multi-headed)	98.493	98.523	98.508	98.493
Proposed (Fusion model)	**99.085**	**99.085**	**99.085**	**99.085**



(a) Random Forest (b) Decision Tree (c) Simple CNN

(d) Multi-headed (e) Fusion Model

Fig. 5: Confusion matrices obtained using different CT scan datasets methods without quantifying uncertainty.



(a) Random Forest (b) Decision Tree (c) Simple CNN (d) Multi-headed (e) Fusion Model

Fig. 6: ROC obtained using different CT scan methods without quantifying uncertainty.

Similar to the previous procedures, we evaluated the performance of three applied deep learning models with uncertainty on X-Ray dataset. The obtained results, confusion matrix, and ROC of the applied deep learning methods on X-Ray dataset with uncertainty are presented by Table V, Figs. 11 and, 12, respectively. Our proposed fusion model achieved the best performance for COVID-19 detection using X-Ray dataset with an accuracy of 96.350% compared to the simple CNN (accuracy of 95.263%). The obtained results in Tables IV and V show that our proposed fusion model performed outstandingly in COVID-19 detection using both CT scan and X-Ray image datasets.

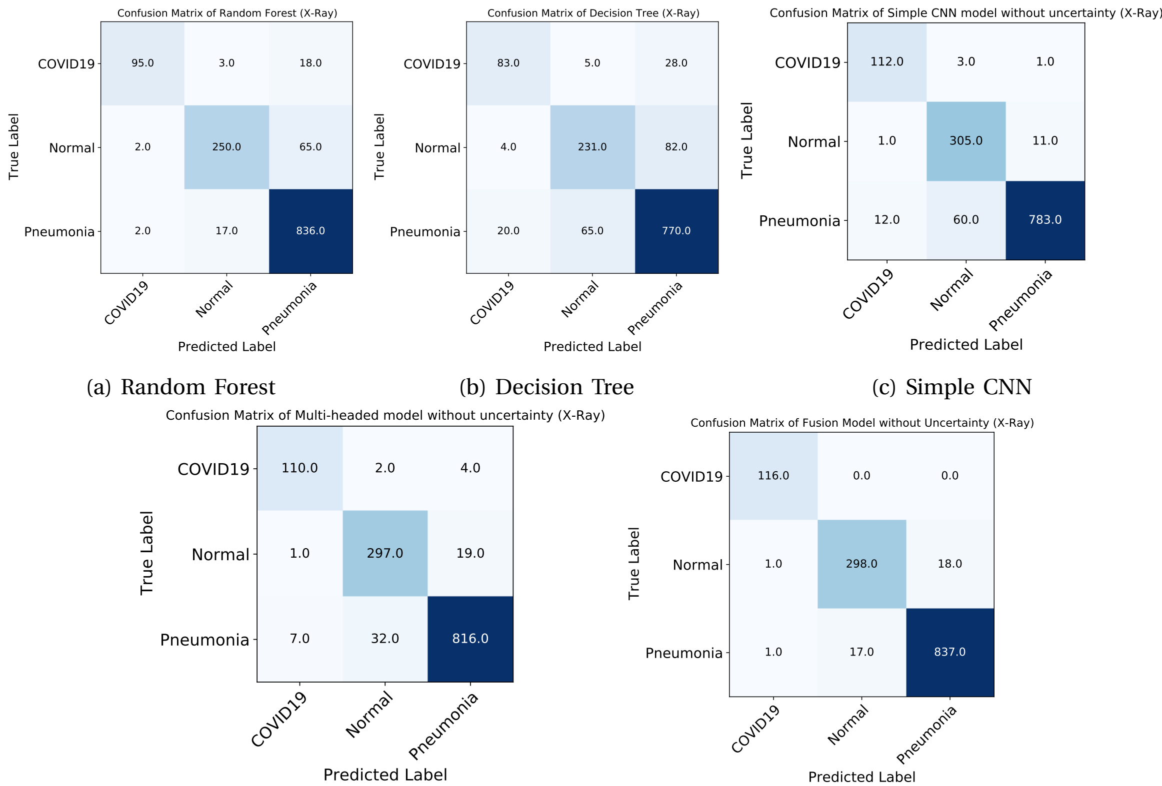
TABLE V: Comprehensive comparison of results obtained using different methods for COVID-19 detection using X-Ray dataset with uncertainty (%).

Method	Precison	Recall	F-Measure	Accuracy
Deep 1 (Simple CNN)	95.263	95.354	95.309	95.263
Deep 2 (Multi-headed)	95.186	95.257	95.222	95.186
Proposed (Fusion model)	**96.350**	**96.370**	**96.360**	**96.350**

C. Robustness Against Noise

The individual visual system is significantly robust against a wide variety of natural noises and corruptions occurring in the nature such as snow, fog or rain [40]. However, the overall performance of various modern image and speech recognition systems is greatly degraded when evaluated using previously unseen noises and corruptions. This means that intelligent image and speech recognition systems are weak against noise and corruptions. Therefore, considering different ML and DL methods in robustness test can reveal the level of stability of the models against noise. In this study, the robustness of the applied deep learning models against noise is investigated.

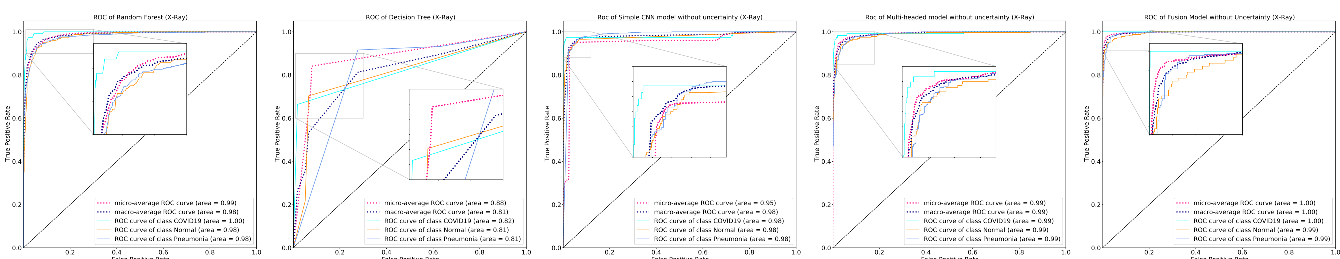
We added different values of noises to both CT scan and X-Ray datasets to evaluate the performance of simple



(a) Random Forest (b) Decision Tree (c) Simple CNN

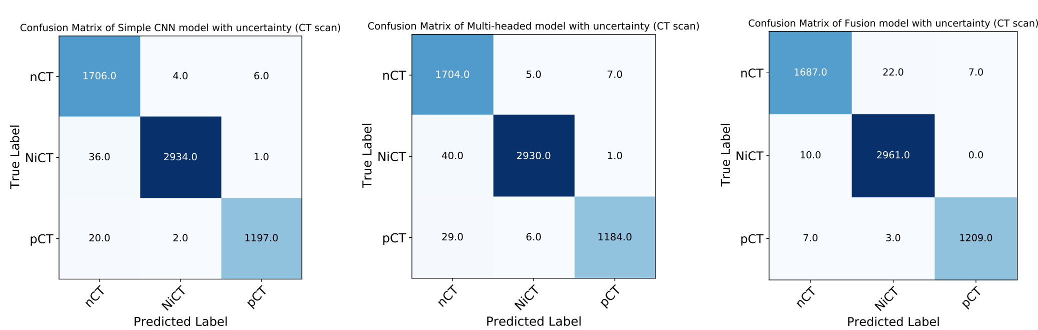
(d) Multi-headed (e) Fusion Model

Fig. 7: Confusion matrices obtained using different X-Ray dataset methods without quantifying uncertainty.



(a) Random Forest (b) Decision Tree (c) Simple CNN (d) Multi-headed (e) Fusion Model

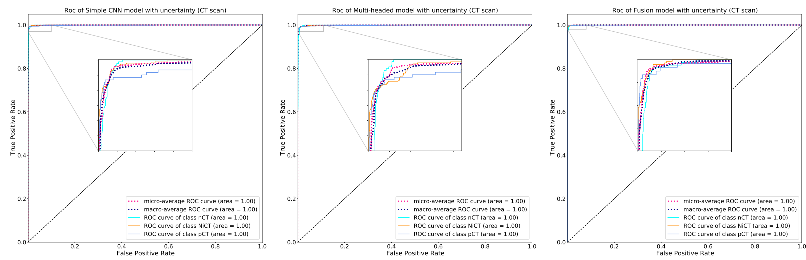
Fig. 8: ROC obtained using different X-Ray dataset methods without quantifying uncertainty.



(a) Simple CNN (b) Multi-headed (c) Fusion Model

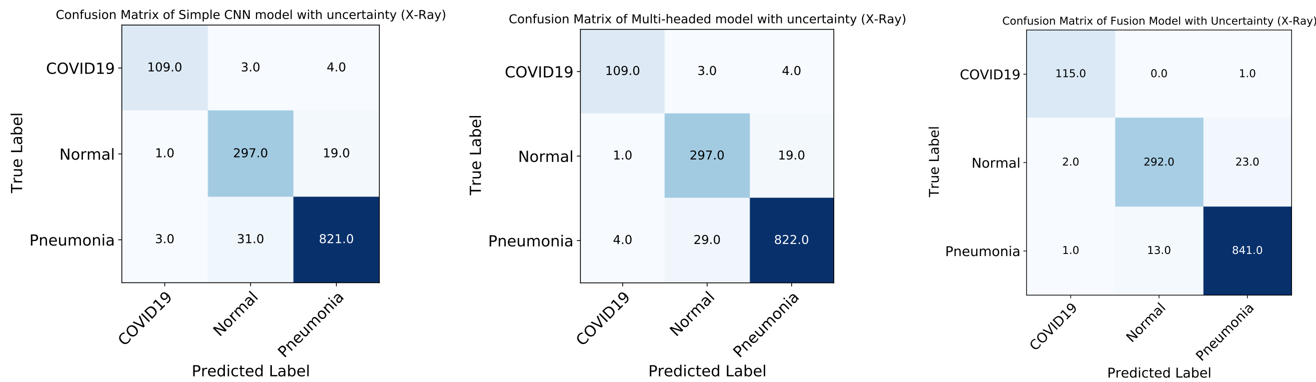
Fig. 9: Confusion matrices obtained using different CT scan dataset methods with quantifying uncertainty.

CNN, multi-headed deep learning and our proposed fusion model. Both STD coefficient and Mean coefficient are used to add noises (Gaussian noise) to our models. The values of STD coefficient are 0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4,



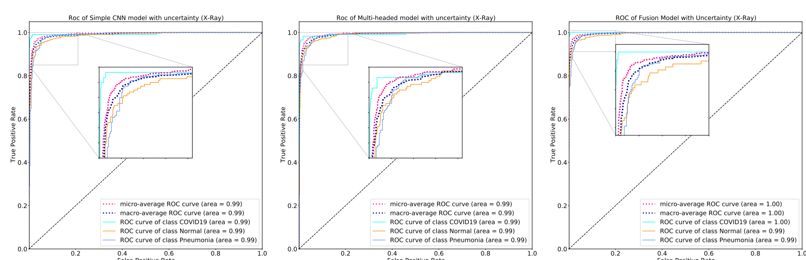
(a) Simple CNN (b) Multi-headed (c) Fusion Model

Fig. 10: ROC obtained using different CT scan datasets methods with quantifying uncertainty.



(a) Simple CNN (b) Multi-headed (c) Fusion Model

Fig. 11: Confusion matrices obtained using different X-Ray datasets methods with quantifying uncertainty.



(a) Simple CNN (b) Multi-headed (c) Fusion Model

Fig. 12: ROC of different methods applied to X-Ray dataset with quantifying uncertainty.

0.5, and 0.6 while the value of Mean coefficient is 0. The obtained results using CT scan and X-Ray datasets are presented in Tables VI and VII, respectively. It may be noted from Table VI that, both simple CNN and multi-headed deep learning models did not perform well with noise as compared to our fusion model when noise is added to CT scan dataset. The results in Table VI show that all metrics of the simple CNN and multi-headed deep learning models are dramatically decreased with the increase of noise. In contrast, our fusion model is more robust against noise when noise is added to CT scan dataset.

Table VII shows the performance of deep learning methods with noise in X-Ray dataset. Both simple CNN and multi-headed deep learning methods did not perform well with the addition of noise while our model is found to be more robust against noise. It should be noted that, our fusion model performed well when using CT dataset with noise than using X-Ray dataset with noise. Nevertheless, our proposed fusion model is more robust against noise compared to the other two deep learning methods.

This stage of the experiment is particularly necessary to demonstrate the stability of the applied models against noise. We clearly showed in this section that our proposed fusion model is very robust against noise for both CT scan and X-Ray image datasets.

D. Unknown Detection

In this sub-section, we evaluate the performance of deep learning models when they are fed by an unknown image. This part of the study emphasizes that models either do not know or can not clearly show their uncertainty during predictions. To achieve this goal, we fed one sample image from the well-known MNIST dataset (see Fig. 13). The mean and STD values of the simple CNN, multi-headed and our

TABLE VI: Comprehensive comparison of robustness against noise using different methods for COVID-19 detection using CT scan dataset (%). Please note that $STD-Cooef = \frac{\sigma_{\epsilon}}{\sigma_0}$, where σ_{ϵ} is standard deviation of the noise and σ_0 is standard deviation of the noise.

Method	STD-Cooef	Precision	Recall	F-Measure	Accuracy
Deep 1 (Simple CNN)	0.0001	98.852	98.831	98.842	98.831
	0.001	98.868	98.848	98.858	98.848
	0.01	98.754	98.730	98.742	98.730
	0.1	95.785	95.614	95.700	95.614
	0.2	89.270	87.284	88.266	87.284
	0.3	86.370	82.593	84.439	82.593
	0.4	82.628	75.194	78.736	75.194
	0.5	78.303	64.053	70.465	64.053
	0.6	75.770	57.534	65.405	57.534
Deep 2 (Multi-headed)	0.0001	98.526	98.493	98.509	98.493
	0.001	98.524	98.493	98.508	98.493
	0.01	98.558	98.526	98.542	98.526
	0.1	93.447	92.871	93.158	92.871
	0.2	86.943	83.423	85.147	83.423
	0.3	80.168	69.065	74.203	69.065
	0.4	76.032	58.465	66.102	58.465
	0.5	73.837	54.690	62.837	54.690
	0.6	72.914	53.149	61.482	53.149
Proposed (Fusion model)	0.0001	99.085	99.085	99.085	99.085
	0.001	99.119	99.119	99.119	99.119
	0.01	99.194	99.187	99.190	99.187
	0.1	99.098	99.085	99.092	99.085
	0.2	98.828	98.814	98.821	98.814
	0.3	98.109	98.086	98.097	98.086
	0.4	96.956	96.884	96.920	96.884
	0.5	96.201	96.088	96.145	96.088
	0.6	95.804	95.665	95.734	95.665

TABLE VII: Comprehensive comparison of robustness against noise by different methods for COVID-19 detection using X-Ray dataset (%).

Method	STD-Cooef	Precision	Recall	F-Measure	Accuracy
Deep 1 (Simple CNN)	0.0001	95.408	95.341	95.375	95.341
	0.001	95.408	95.341	95.341	95.341
	0.01	95.338	95.263	95.301	95.263
	0.1	94.554	94.254	94.404	94.254
	0.2	91.540	89.285	90.398	89.285
	0.3	88.534	82.065	85.176	82.065
	0.4	85.770	73.136	78.951	73.136
	0.5	84.294	64.518	73.092	64.518
	0.6	82.545	57.375	67.696	57.375
Deep 2 (Multi-headed)	0.0001	95.474	95.419	95.446	95.419
	0.001	95.188	95.108	95.148	95.108
	0.01	95.404	95.341	95.372	95.341
	0.1	93.922	93.322	93.621	93.322
	0.2	88.861	82.453	85.537	82.453
	0.3	83.781	58.074	68.598	58.074
	0.4	82.207	40.062	53.871	40.062
	0.5	81.750	31.521	45.499	31.521
	0.6	81.366	27.639	41.262	27.639
Proposed (Fusion model)	0.0001	96.498	96.506	96.502	96.506
	0.001	96.568	96.583	96.576	96.583
	0.01	96.492	96.506	96.499	96.506
	0.1	96.363	96.350	96.357	96.350
	0.2	94.403	94.254	94.329	94.254
	0.3	91.769	91.071	91.418	91.071
	0.4	88.225	85.714	86.951	85.714
	0.5	84.181	78.804	81.404	78.804
	0.6	81.082	68.322	74.157	68.322

proposed fusion methods are reported in Table VIII. It can clearly be observed that, our fusion model showed its uncertainty towards unknown data more clearly than the other two deep learning methods.

We fed Fig. 13 to all three deep learning methods trained by CT scan and X-Ray datasets and then predicted the class of this unknown image sample.

Estimating uncertainty of machine and deep learning algorithms using different UQ methods is vital during crit-

Fig. 13: The MNIST sample image fed to the deep learning methods as an unknown sample.

ical predictions such as medical case studies. The applied machine and deep learning with different UQ methods should be able to capture a portion of uncertainties such as epistemic and aleatoric. In this study, we applied a fusion model to classify two types of COVID-19 images: CT scan and X-Ray datasets. In Table VIII, we computed *Mean* and *STD* for three deep learning methods to evaluate the predictions of these models when fed with unknown input data. It should be noted that *Mean* is the prediction and *STD* is the uncertainty of models. As can be seen in Table VIII, our model usually has 0 (or nearly zero) for one out of three classes in both datasets.

V. Discussion

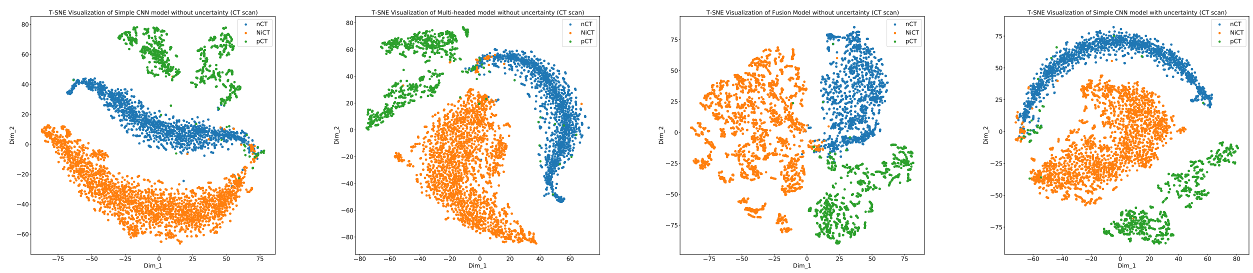
The early, timely and accurate detection of COVID-19 is crucial role in dealing with coronavirus as quickly and effectively as possible. It may be noted that various fields of science have been proposing new methods to deal with this disease. Among them, ML and DL methods are found to be more efficient and have been widely used for segmentation and detection of COVID-19 tasks. In this work, we mainly focused on the detection task using both CT scan and X-Ray images. In this study, we proposed a new, simple yet very efficient fusion model called $UncertaintyFuseNet$ and compared its performance with several classical and DL algorithms. It can be noted from the obtained results that our developed fusion model is accurate in detecting the COVID-19 cases. Besides these, we have also shown the superiority of our model in dealing with noise over the other two models. Also, the obtained results in the earlier section reveal the highest ability of our $UncertaintyFuseNet$ model in detecting features of the unseen images.
Finally, the performance of the proposed fusion model is compared with other existing methods in the literature. In Table 14, we compared the performance of the proposed fusion model with other state-of-the-art techniques developed using CT and X-Ray image datasets.

To identify the important features in our proposed fusion model for each class, the Grad-CAM visualization approach is used to both CT scan and X-Ray image datasets. Figs. 16 and 17 show the most important features used by our fusion model for identifying each class separately.

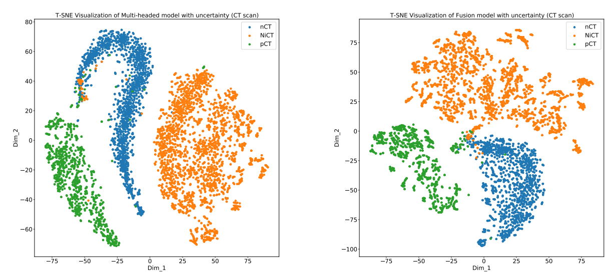
Finally, the output posterior distributions of our proposed fusion model for both datasets is presented in Fig. 18. It can be noted from the figure that, the correctly classified samples do not overlap with the other two classes (incorrect classes).

TABLE VIII: Unknown class detection by different deep learning methods when fed with Fig. 13.

Method		CT scan			X-Ray		
		nCT	NiCT	pCT	COVID-19	Normal	Pneumonia
Deep 1	Mean	0.02	0.98	0.0	0.57	0.15	0.28
(Simple CNN)	STD	0.10	0.10	0.01	0.39	0.26	0.35
Deep 2	Mean	0.05	0.30	0.65	0.68	0.22	0.10
(Multi-headed)	STD	0.19	0.43	0.45	0.32	0.27	0.16
Proposed	Mean	**0.56**	**0.0**	**0.44**	**0.41**	**0.59**	**0.0**
Fusion	STD	**0.50**	**0.07**	**0.50**	**0.49**	**0.49**	**0.0**



(a) Simple CNN without UQ (b) Multi-headed without UQ (c) Fusion Model without UQ (d) Simple CNN with UQ



(e) Multi-headed with UQ (f) Fusion Model with UQ

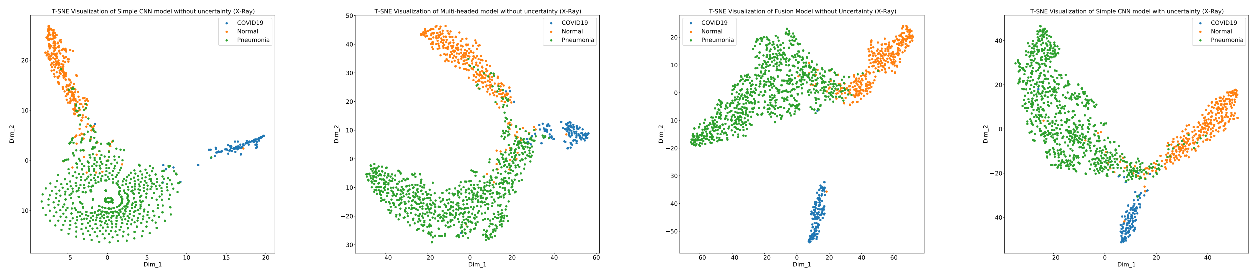
Fig. 14: T-SNE visualisation of different methods applied to CT scan dataset without and with quantifying uncertainty.

In our study, we have evaluated our *UncertaintyFuseNet* model using both CT scan and X-Ray datasets. This research attempts to fill the gap in [64] which are reported in the literature.

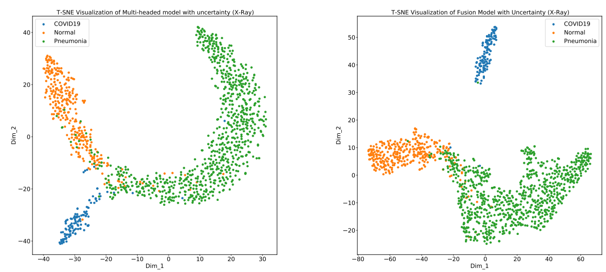
Indeed, this study is a work entirely related to past study developed to fill their gap(s) [64]. Wang et al. [64] proposed a deep feature fusion model for COVID-19 detection using CT scan dataset. Their proposed model performed significantly better in detecting COVID-19 using CT scan dataset. However, they stated the proposed model may not perform well for other types of data such as X-Ray. In another study, Tang et al. [65] proposed an ensemble deep learning model for COVID-19 detection using X-Ray images only. Most of previous studies on COVID-19 detection only focused on performance rather than the uncertainty analysis of their models. We firmly believe in obtaining accurate predictions, having impressive performance. Shamsi Jokandan et al. [5] considered uncertainty in their study; however, they used very small datasets in their study. In this work, we proposed a novel fusion model using two big CT scan and X-Ray datasets. Also, we applied an uncertainty quantification method (*i.e.*, EMC) to improve the quality of predictions.

There are a limited number of studies that have considered noise in the COVID-19 datasets which is another important research gap that we are facing in real scenarios. Also, uncertainty of models during predictions must be demonstrated in practice. One way is to give unknown samples and make the model predict their classes. We considered both tests to evaluate the effectiveness of our *UncertaintyFuseNet* model. The proposed fusion model in this study undoubtedly provides a practical solution for the real-world scenarios. It should be noted that although the proposed fusion model fusion is found to be suitable for COVID-19 detection, but it is a general model, and we think it is also effective in classifying other diseases. To sum up, the most important features of our proposed fusion model in this study are listed below:

1) Obtained the highest COVID-19 detection performance compared to both datasets (CT and X-Ray).
2) The developed model is able to incorporate model uncertainty during detection.
3) Proposed model is robust against noise.
4) Able to detect unknown data with high accuracy.



(a) Simple CNN without UQ (b) Multi-headed without UQ (c) Fusion Model without UQ (d) Simple CNN with UQ



(e) Multi-headed with UQ (f) Fusion Model with UQ

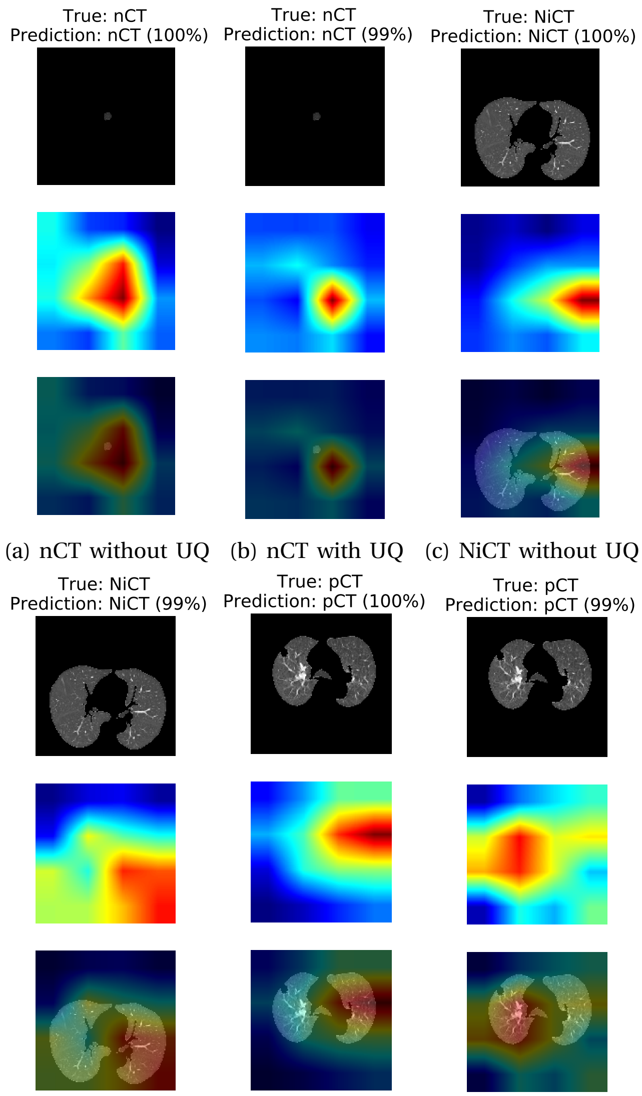
Fig. 15: T-SNE visualisation of different methods applied to X-Ray dataset without and with quantifying uncertainty.

TABLE IX: Comprehensive comparison of results obtained by our proposed model with state-of-the-art techniques on automated detection of COVID-19 cases using both CT scan and X-Ray image datasets.

Dataset	Study	Year	# of Samples	Performance					UQ	Code
				Precision	Recall	F-measure	Accuracy	AUC		
CT scan	Li et al. [41]	2020	1540 (3 classes)	N/A	82.60	N/A	N/A	0.918	×	×
	Jaiswal et al. [42]	2020	2492 (2 classes)	96.29	96.29	96.29	96.25	0.970	×	×
	Wang et al. [43]	2020	640 (2 classes)	96.61	97.71	97.14	97.15	N/A	×	×
	Sharma [44]	2020	2200 (3 classes)	N/A	92.10	N/A	91.00	N/A	×	×
	Panwar et al. [45]	2020	1600 (2 classes)	95.00	95.00	95.00	95.00	N/A	×	×
	Do and Vu [46]	2020	746 (2 classes)	85.00	85.00	85.00	85.00	0.922	×	×
	Singh [47]	2020	N/A (2 classes)	N/A	91.00	89.97	93.50	N/A	×	×
	Pham [48]	2020	746 (2 classes)	N/A	91.14	93.00	92.62	0.980	×	×
	Martinez [49]	2020	746 (2 classes)	94.40	86.60	90.30	90.40	0.965	×	×
	Loey et al. [50]	2020	11012 (2 classes)	N/A	80.85	N/A	81.41	N/A	×	×
	Ning et al. [51]	2020	19685 (3 classes)	N/A	N/A	N/A	N/A	0.978	×	×
	Han et al. [52]	2020	460 (3 classes)	95.90	90.50	92.30	94.30	0.988	×	×
	Shamsi Jokandan et al. [5]	2021	746 (2 classes)	N/A	86.50	N/A	87.90	0.942	√	×
	Ours	**2021**	**19685 (3 classes)**	**99.08**	**99.08**	**99.08**	**99.08**	**1.00**	√	√
X-Ray	Khan et al. [32]	2020	1251 (4 classes)	90.00	89.92	89.80	89.60	N/A	×	√
	Ozturk et al. [53]	2020	1125 (3 classes)	89.96	85.35	87.37	87.02	N/A	×	√
	Mesut and [54]	2020	458 (3 classes)	98.89	98.33	98.57	99.27	N/A	×	√
	Mahmud et al. [55]	2020	1220 (4 classes)	82.87	83.82	83.37	90.30	0.825	×	√
	Heidari et al. [56]	2020	2544 (3 classes)	N/A	N/A	N/A	94.50	N/A	×	×
	Rahimzadeh and Attar [57]	2020	11302 (3 classes)	72.83	87.31	N/A	91.40	N/A	×	√
	Pereira et al. [58]	2020	1144 (7 classes)	N/A	N/A	64.91	N/A	N/A	×	×
	De Moura et al. [59]	2020	1616 (3 classes)	79.00	79.33	79.33	79.86	N/A	×	×
	Yoo et al. [60]	2020	1170 (2 classes)	97.00	99.00	97.98	98.00	0.980	×	×
	Chandra et al. [61]	2020	2346 (2 classes)	N/A	N/A	N/A	91.32	0.914	×	×
	Zhang et al. [62]	2020	2706 (2 classes)	77.13	N/A	N/A	78.57	0.844	×	√
	Shamsi Jokandan et al. [5]	2021	100 (2 classes)	N/A	99.90	N/A	98.60	0.997	√	×
	Ahmad [63]	2021	4200 (4 classes)	93.01	92.97	92.97	96.49	N/A	×	×
	Ours	**2021**	**6432 (3 classes)**	**96.35**	**96.37**	**96.36**	**96.35**	**0.993**	√	√

VI. Conclusion

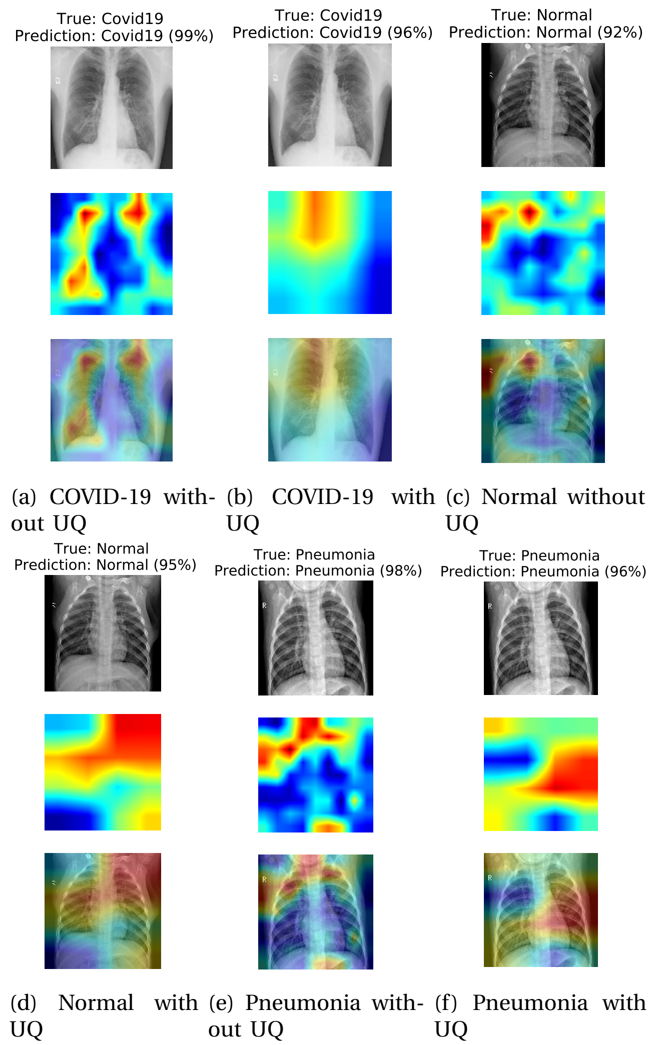
In this study, we have proposed a new fusion deep learning model to accurately detect COVID-19 using CT scan and X-Ray datasets. In order to detect the COVID-19 accurately, we considered the uncertainty issues while detecting the disease. Our proposed $UncertaintyFuseNet$ model addressed the possible uncertainties and demonstrated stronger robustness to noise and unknown data. In addition, a class-wise analysis procedure has been implemented to obtain robust performance of the applied



(a) nCT without UQ (b) nCT with UQ (c) NiCT without UQ

(d) NiCT with UQ (e) pCT without UQ (f) pCT with UQ

Fig. 16: Grad-CAM visualization for our proposed fusion model without and with UQ for COVID-19 (17a and 17b), Normal (17c and 17d), and Pneumonia (17e and 16f) using CT scan dataset.



(a) COVID-19 without UQ (b) COVID-19 with UQ (c) Normal without UQ

(d) Normal with UQ (e) Pneumonia without UQ (f) Pneumonia with UQ

Fig. 17: Grad-CAM visualization for our proposed fusion model without and with UQ for COVID-19 (17a and 17b), Normal (17c and 17d), and Pneumonia (17e and 17f) using X-Ray dataset.

methods. The experimental results show that our fusion model can achieve the best performance using both CT and X-ray datasets. Utilizing the hierarchical feature features in several stages of the proposed fusion model yielded the highest detection performance compared to the other applied deep learning methods. Also, we added different types of noise to our model and showed that our fusion model performed better than other two deep learning methods. Also, the proposed fusion model handled the uncertainty much better when the unknown data is fed to the model. We have demonstrated the effectiveness of the proposed $UncertaintyFuseNet$ model using various experiments. The limitation of our proposed fusion model will be addressed in our future studies. In future, we intend to: (i) expand the COVID-19 datasets and test our fusion model using multi-modal data, (ii) include attention mechanism while fusing features, and (iii) integrate with other fusion techniques such as decision level fusion.

VII. Acknowledgment

This research was partially supported by the Australian Research Council's Discovery Projects funding scheme (project DP190102181).

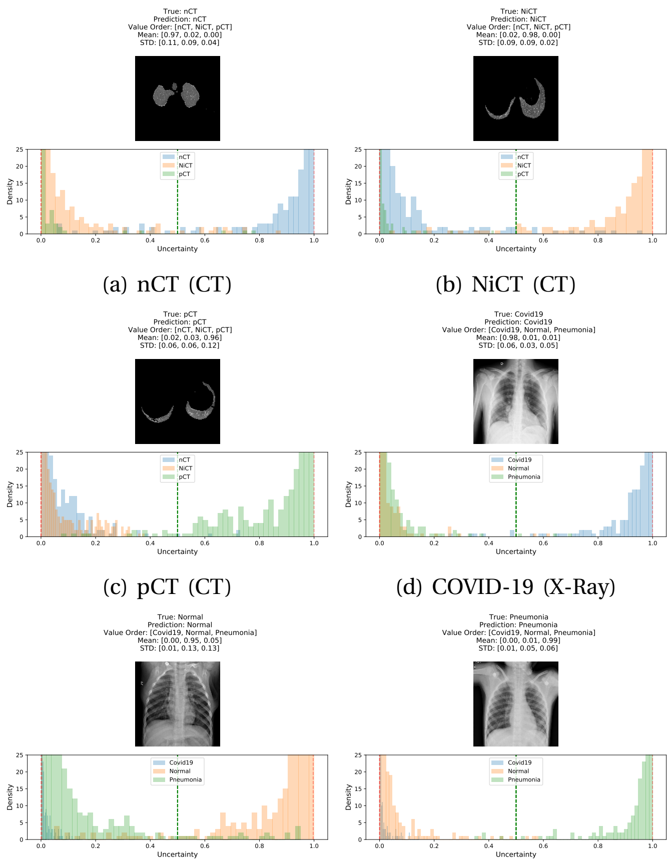
References

[1] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849*, 2020.

[2] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," 2020.

[3] F. Pourpanah, M. Abdar, Y. Luo, X. Zhou, R. Wang, C. P. Lim, and X.-Z. Wang, "A review of generalized zero-shot learning methods," *arXiv preprint arXiv:2011.08641*, 2020.

[4] Y. Luo, X. Wang, and F. Pourpanah, "Dual vaegan: A generative model for generalized zero-shot learning," *Applied Soft Computing*, p. 107352, 2021.



(a) nCT (CT)

(b) NiCT (CT)

(c) pCT (CT)

(d) COVID-19 (X-Ray)

(e) Normal (X-Ray)

(f) Pneumonia (X-Ray)

Fig. 18: The output posterior distributions of our proposed fusion model for nCT (18a), NiCT (18b), and pCT (18c) using CT scan dataset, and COVID-19 (18d), Normal (18e), and Pneumonia (18f) using X-Ray dataset, respectively.

[5] A. Shamsi, H. Asgharnezhad, S. S. Jokandan, A. Khosravi, P. M. Kebria, D. Nahavandi, S. Nahavandi, and D. Srinivasan, "An uncertainty-aware transfer learning-based framework for covid-19 diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[6] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," *arXiv preprint arXiv:2003.14395*, 2020.

[7] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label," *medRxiv*, 2020.

[8] L. O. Hall, R. Paul, D. B. Goldgof, and G. M. Goldgof, "Finding covid-19 from chest x-rays using deep learning on a small dataset," *arXiv preprint arXiv:2004.02060*, 2020.

[9] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network," *arXiv preprint arXiv:2003.13815*, 2020.

[10] X. Chen, L. Yao, and Y. Zhang, "Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images," *arXiv preprint arXiv:2004.05645*, 2020.

[11] F. Ali, S. El-Sappagh, S. R. Islam, D. Kwak, A. Ali, M. Imran, and K.-S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208–222, 2020.

[12] Z. Shanshan, "Original paper multi-source information fusion technology and its engineering application."

[13] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest ct," *Information Fusion*, vol. 42, pp. 102–110, 2018.

[14] M. Martínez-Ballesteros, J. M. García-Heredia, I. A. Nepomuceno-Chamorro, and J. C. Riquelme-Santos, "Machine learning techniques to discover genes with potential prognosis role in alzheimer's disease using different biological sources," *Information Fusion*, vol. 36, pp. 114–129, 2017.

[15] M. Chen, P. Zhou, D. Wu, L. Hu, M. M. Hassan, and A. Alamri, "Ai-skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework," *Information Fusion*, vol. 54, pp. 1–9, 2020.

[16] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network," *Information Fusion*, vol. 53, pp. 174–182, 2020.

[17] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information fusion*, vol. 19, pp. 4–19, 2014.

[18] U. R. Acharya, H. Fujita, V. K. Sudarshan, M. R. K. Mookiah, J. E. Koh, J. H. Tan, Y. Hagiwara, C. K. Chua, S. P. Junnarkar, A. Vijayananthan *et al.*, "An integrated index for identification of fatty liver disease using radon transform and discrete cosine transform features in ultrasound images," *Information Fusion*, vol. 31, pp. 43–53, 2016.

[19] Q. He, X. Li, D. N. Kim, X. Jia, X. Gu, X. Zhen, and L. Zhou, "Feasibility study of a multi-criteria decision-making based hierarchical model for multi-modality feature and multi-classifier fusion: Applications in medical prognosis prediction," *Information Fusion*, vol. 55, pp. 207–219, 2020.

[20] A. Leslie, A. Jones, and P. Goddard, "The influence of clinical information on the reporting of ct by radiologists." *The British journal of radiology*, vol. 73, no. 874, pp. 1052–1055, 2000.

[21] W. W. Boonn and C. P. Langlotz, "Radiologist use of and perceived need for patient data access," *Journal of digital imaging*, vol. 22, no. 4, pp. 357–362, 2009.

[22] J. B. Jonas, Y. X. Wang, L. Dong, and S. Panda-Jonas, "High myopia and glaucoma-like optic neuropathy," *Asia-pacific Journal of Ophthalmology (Philadelphia, Pa.)*, vol. 9, no. 3, p. 234, 2020.

[23] E. S. Kumar and P. S. Jayadev, "Deep learning for clinical decision support systems: a review from the panorama of smart healthcare," *Deep Learning Techniques for Biomedical and Health Informatics*, pp. 79–99, 2020.

[24] M. W. Dusenberry, D. Tran, E. Choi, J. Kemp, J. Nixon, G. Jerfel, K. Heller, and A. M. Dai, "Analyzing the role of model uncertainty for electronic health records," in *Proceedings of the ACM Conference on Health, Inference, and Learning*, 2020, pp. 204–213.

[25] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *arXiv preprint arXiv:2011.06225*, 2020.

[26] M. Abdar, M. Samami, S. D. Mahmoodabad, T. Doan, B. Mazoure, R. Hashemifesharaki, L. Liu, A. Khosravi, U. R. Acharya, V. Makarenkov *et al.*, "Uncertainty quantification in skin cancer classification using three-way decision-based bayesian deep learning," *Computers in Biology and Medicine*, p. 104418, 2021.

[27] M. Z. Uddin, M. M. Hassan, A. Alsanad, and C. Savaglio, "A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare," *Information Fusion*, vol. 55, pp. 105–115, 2020.

[28] G. Quellec, M. Lamard, G. Cazuguel, C. Roux, and B. Cochener, "Case retrieval in medical databases by fusing heterogeneous information," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 108–118, 2010.

[29] L. Wu, L. Wang, N. Li, T. Sun, T. Qian, Y. Jiang, F. Wang, and Y. Xu, "Modeling the covid-19 outbreak in china through multi-source information fusion," *The Innovation*, vol. 1, no. 2, p. 100033, 2020.

[30] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Information Fusion*, vol. 53, pp. 155–164, 2020.

[31] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks," *Computers in Biology and Medicine*, p. 103795, 2020.

[32] A. I. Khan, J. L. Shah, and M. M. Bhat, "Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, p. 105581, 2020.

[33] V. K. R. Chimmula and L. Zhang, "Time series forecasting of covid-19 transmission in canada using lstm networks," *Chaos, Solitons & Fractals*, p. 109864, 2020.

[34] 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," *arXiv preprint arXiv:2004.02696*, 2020.

[35] Y. Oh, S. Park, and J. C. Ye, "Deep learning covid-19 features on cxr using limited training data sets," *IEEE Transactions on Medical Imaging*, 2020.

[36] N. S. Punn and S. Agarwal, "Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks," *arXiv preprint arXiv:2004.11676*, 2020.

[37] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34 – 45, 2019.

[38] G. Luo, S. Dong, W. Wang, K. Wang, S. Cao, C. Tam, H. Zhang, J. Howey, P. Ohorodnyk, and S. Li, "Commensal correlation network between segmentation and direct area estimation for bi-ventricle quantification," *Medical image analysis*, vol. 59, p. 101591, 2020.

[39] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection," *arXiv preprint arXiv:2003.10769*, 2020.

[40] E. Rusak, L. Schott, R. S. Zimmermann, J. Bitterwolf, O. Bringmann, M. Bethge, and W. Brendel, "A simple way to make neural networks robust against diverse image corruptions," in *European Conference on Computer Vision.* Springer, 2020, pp. 53–69.

[41] K. Li, Y. Fang, W. Li, C. Pan, P. Qin, Y. Zhong, X. Liu, M. Huang, Y. Liao, and S. Li, "Ct image visual quantitative evaluation and clinical classification of coronavirus disease (covid-19)," *European radiology*, pp. 1–10, 2020.

[42] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the covid-19 infected patients using densenet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, pp. 1–8, 2020.

[43] S.-H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y.-D. Zhang, "Covid-19 classification by fgcnet with deep feature fusion from graph convolutional network and convolutional neural network," *Information Fusion*, vol. 67, pp. 208–229, 2020.

[44] S. Sharma, "Drawing insights from covid-19-infected patients using ct scan images and machine learning techniques: a study on 200 patients," *Environmental Science and Pollution Research*, vol. 27, no. 29, pp. 37 155–37 163, 2020.

[45] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, and V. Singh, "A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images," *Chaos, Solitons & Fractals*, vol. 140, p. 110190, 2020.

[46] C. Do and L. Vu, "An approach for recognizing covid-19 cases using convolutional neural networks applied to ct scan images," in *Applications of Digital Image Processing XLIII*, vol. 11510. International Society for Optics and Photonics, 2020, p. 1151034.

[47] D. Singh, V. Kumar, and M. Kaur, "Classification of covid-19 patients from chest ct images using multi-objective differential evolution–based convolutional neural networks," *European Journal of Clinical Microbiology & Infectious Diseases*, pp. 1–11, 2020.

[48] T. D. Pham, "A comprehensive study on classification of covid-19 on computed tomography with pretrained convolutional neural networks," *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.

[49] A. R. Martinez, "Classification of covid-19 in ct scans using multi-source transfer learning," *arXiv preprint arXiv:2009.10474*, 2020.

[50] M. Loey, G. Manogaran, and N. E. M. Khalifa, "A deep transfer learning model with classical data augmentation and cgan to detect covid-19 from chest ct radiography digital images," *Neural Computing and Applications*, pp. 1–13, 2020.

[51] W. Ning, S. Lei, J. Yang, Y. Cao, P. Jiang, Q. Yang, J. Zhang, X. Wang, F. Chen, Z. Geng *et al.*, "ictcf: an integrative resource of chest computed tomography images and clinical features of patients with covid-19 pneumonia," 2020.

[52] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, and W. Zhang, "Accurate screening of covid-19 using attention-based deep 3d multiple instance learning," *IEEE transactions on medical imaging*, vol. 39, no. 8, pp. 2584–2594, 2020.

[53] 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," *Computers in Biology and Medicine*, p. 103792, 2020.

[54] 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," *Computers in Biology and Medicine*, p. 103805, 2020.

[55] T. Mahmud, M. A. Rahman, and S. A. Fattah, "Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization," *Computers in biology and medicine*, vol. 122, p. 103869, 2020.

[56] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving the performance of cnn to predict the likelihood of covid-19 using chest x-ray images with preprocessing algorithms," *International journal of medical informatics*, vol. 144, p. 104284, 2020.

[57] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2," *Informatics in Medicine Unlocked*, p. 100360, 2020.

[58] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr, and Y. M. Costa, "Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios," *Computer Methods and Programs in Biomedicine*, p. 105532, 2020.

[59] J. De Moura, L. R. García, P. F. L. Vidal, M. Cruz, L. A. López, E. C. Lopez, J. Novo, and M. Ortega, "Deep convolutional approaches for the analysis of covid-19 using chest x-ray images from portable devices," *IEEE Access*, vol. 8, pp. 195 594–195 607, 2020.

[60] S. H. Yoo, H. Geng, T. L. Chiu, S. K. Yu, D. C. Cho, J. Heo, M. S. Choi, I. H. Choi, C. Cung Van, N. V. Nhung *et al.*, "Deep learning-based decision-tree classifier for covid-19 diagnosis from chest x-ray imaging," *Frontiers in medicine*, vol. 7, p. 427, 2020.

[61] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, and S. S. Netam, "Coronavirus disease (covid-19) detection in chest x-ray images using majority voting based classifier ensemble," *Expert systems with applications*, vol. 165, p. 113909, 2021.

[62] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen *et al.*, "Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection," *IEEE transactions on medical imaging*, 2020.

[63] F. Ahmad, A. Farooq, and M. U. Ghani, "Deep ensemble model for classification of novel coronavirus in chest x-ray images," *Computational Intelligence and Neuroscience*, vol. 2021.

[64] S.-H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y.-D. Zhang, "Covid-19 classification by fgcnet with deep feature fusion from graph convolutional network and convolutional neural network," *Information Fusion*, vol. 67, pp. 208–229, 2021.

[65] S. Tang, C. Wang, J. Nie, N. Kumar, Y. Zhang, Z. Xiong, and A. Barnawi, "Edl-covid: Ensemble deep learning for covid-19 cases detection from chest x-ray images," *IEEE Transactions on Industrial Informatics*, 2021.