# Monte Carlo Dropout for Uncertainty Analysis and ECG Trace Image Classification

**Abstract.** Cardiovascular diseases (CVDs), such as arrhythmias (abnormal heartbeats) are the prime cause of mortality across the world. ECG graphs are utilized by cardiologists to indicate any unexpected cardiac activity. Deep Neural Networks (DNN) serve as a highly successful method for classifying ECG images for the purpose of computer-aided diagnosis. However, DNNs can not quantify uncertainty in predictions, as they are incapable of discriminating between anomalous data and training data. Hence, a lack of trust in automated diagnosis and the potential to cause severe decision-making issues is created, particularly in medical practises. In this paper we propose an uncertainty-aware ECG classification model where Convolutional Neural Networks (CNN), combined with Monte Carlo Dropout (MCD) is employed to evaluate the uncertainty of the model, providing a more trustworthy process for real-world scenarios. We use ECG images dataset of cardiac and covid-19 patients containing five categories of data, which includes COVID-19 patients' ECG records as well as data from other cardiovascular disorders. Our proposed model achieves 93.90% accuracy using this dataset.

**Keywords:** ECG. COVID-19. Cardiovascular diseases. Uncertainty Analysis. Convolutional Neural Network. Monte Carlo Dropout.

## 1 Introduction

Deep Learning (DL) models have been immensely popularized over the last few years due to their contribution to various fields of machine learning based systems. However, DNNs can not measure the uncertainty which ends up as a drawback in real world applications of DNNs, especially those concerning medical image processing. ECG Classification is a classical example of medical image and signal processing where DNNs are utilized. Worldwide, the Electrocardiogram (ECG) is applied in the examination of CVDs as it is relatively cheap and rather non-invasive in nature. The procedure of using an ECG machine involves tracing the electric potential generated between each heartbeat onto a graph. Our heart beats in a consistent manner, periodically exciting the myocardium (the muscular layer of the heart) as it pumps blood throughout the body. When the myocardium contracts, slight current is generated by the heart and transmitted to the body's surface, causing potential changes in each part of the body. During an ECG examination, there are multiple electrodes attached to different parts of the body, which record these potential changes using an electrocardiograph or a vector electrocardiograph. Thus, an ECG is obtained. ECGs can greatly assist in helping us to effectively diagnose and classify any existing heart disease. Diseases relating to the heart or blood vessels are collectively called CVDs. In 2019, the American Heart Association announced that CVDs are the major cause of death. It accumulated over 17.6 million deaths in 2016 and the number is estimated to reach 23.6 million by 2030 [1]. The way ECGs can methodically point out irregularities in our heartbeat is vital to the detection of CVDs. The sooner such heart disorders are discovered, the better it is, since most are actually treatable at an early stage. The risk of detriment and untimely death due to such ailments only increases with time. Research indicates that ECG is of great significance in foreshadowing both short and long-term consequences. For instance, a patient suffering from myocardial infarction has a greater chance of recovery the sooner the abnormal cardiac activity is detected by an ECG [2].

Researchers have been trying to automate the diagnosis of ECG signals by emphasizing on the results related to atrial fibrillation (AF) since 1957. An automated ECG diagnostic system holds the capability to examine the general populace and provide a valuable second opinion for health care practitioners. Since the population of the world is vastly affected by COVID-19, uncertain decisions made by the prediction model could be dangerous. It is reported that COVID-19 infection can possibly lead to severe myocarditis in previously healthy patients [3]. Statistics reveal that 27.8% of COVID19 patients experienced an increase in their troponin level, which is beyond the 99th percentile of the upper reference limit. This revelation points toward a case of acute myocardial injury in early cases reported from China [4]. The percentage is nearly ten times higher than the rate of influenza contagion (2.9%) [5]. Despite the acute myocardial damage, the majority of the COVID-19 patients recover without any obvious cardiac issues. Regardless, it is a matter of concern, since there may be obscure impairments (sub-clinical or hidden cardiac injury) causing harm in the long run. Currently, the world is emerging from the depths of the pandemic so the focus of health care shifts to finding out whether cardiac monitoring in COVID-19 survivors is necessary or not. All arguments considered, uncertainty analysis is imperative in guaranteeing better, definitive results, and accordingly, making them more applicable for real-life scenarios[6].

For estimating uncertainty, Bayesian neural networks (BNN) [7, 8] are a probabilistic variant of neural networks that are fundamentally suitable. The variational inference [8] is generally applied to calculate the posterior model by using variational (such as Gaussian distribution) distribution. However, this approach is not very convenient and does not improve the accuracy. Hence A method especially suitable for BNNs which may not always outperform other state-of-the-art models, is the MCD method. Gal and Ghahramani [9] introduced this idea of using dropouts to determine the model's uncertainty. They noticed that, while dropouts are generally applied to prevent overfitting, it may very well be implemented to depict a model's rough estimate of the weight's posterior. The concept relies on making more than one forward passes with the trained model, when dropout is applied at the testing phase, so that predictions and model uncertainty may be methodically calculated, for a given input. This method is capable of working with large amounts of data without affecting the architecture of the model.

Considering the risks of classifying ECG images, we propose an Uncertainty-aware CNN model using MCD to safely continue the autonomous diagnosis process. Our main contributions are:

- Using CNN to classify five distinct categories of ECG images including COVID-19.
- Using MCD to analyze the uncertainty of the model
- Finding uncertain samples, for safer diagnosis

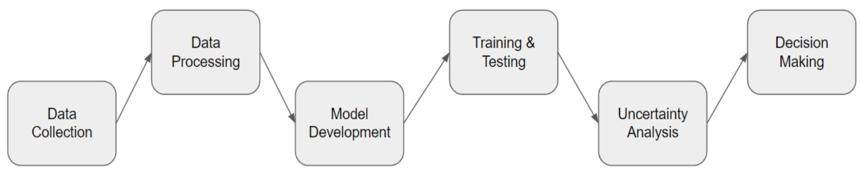



Fig. 1: Illustration of workflow

## 2 Literature Review

Many studies on ECG classification have been conducted over the years, given the importance of the problem. Faust et al[10] and Hong et al[11] provide a good review of ECG. We only discuss about the relevant studies here.

### 2.1 ECG Classification

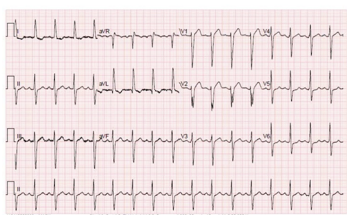
In most of the ECG classification studies, the input is either a 1-dimensional numerical ECG signal values [12] or a 2-dimensional image of the ECG signal [13]. Here, we train a 2-dimensional CNN on 2-dimensional images of ECG signals for disease classification. For our study, we use a ECG trace image dataset of covid-19 and cardiac patients[14].

Irmak et al.[15] present a CNN model to detect COVID-19 with the help of ECG trace images. They achieve 83.05% five-class classification accuracy using their approach. Another CNN based model proposed by [16] attempt three-class classification, reaching 90.8% accuracy. Sobahi et al.[17] propose an attention-based 3-dimensional CNN model with residual connections (RC). Their approach reaches 92.0% accuracy on four-class classification. Attallah et al.[18] develops five distinct DL models. Discrete wavelet transform is used to merge features extracted from upper layers, which are subsequently combined with lower-layer features. The predictions of three machine learning classifiers is merged using a classification system, which is developed following the implementation of a feature selection approach. The model achieves 91.7% accuracy in three-class classification.

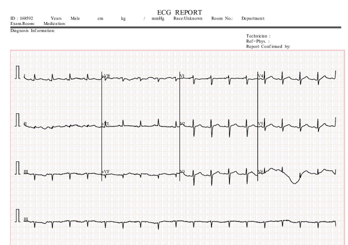
## 2.2 Uncertainty Estimation in Medical Image Analysis

Studies regarding uncertainty in DNN prototypes are conducted in order to increase the reliability of the prediction methods. Evaluation of uncertainty in medical applications has been frequently applied for diagnoses of diseases, such as covid-19 [19], tuberculosis [20], and cancer [21] etc. Nevertheless, there is no uncertainty analysis on covid-19 ECG images classification as per our knowledge, till date.

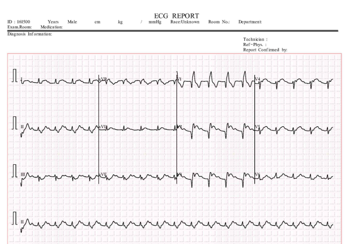
Ghoshal et al.[22] used drop-weights based BNN to quantify uncertainty in DL techniques. They show that the predictive accuracy of the model was notably associated with its uncertainty. Milanés-Hermosilla et al.[23] implement Shallow CNN and with an ensemble model to classify motor imagery. Besides, they use MCD to find the uncertainty of their predictions, which make their model more reliable.



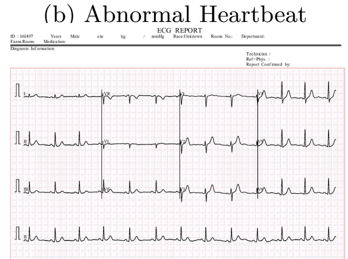
(a) Covid-19



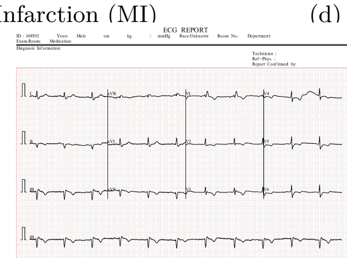

(b) Abnormal Heartbeat




(c) Myocardial Infarction (MI)




(d) Normal




(e) Previous History of MI

Fig. 2: Sample of each classes from the dataset

## 3 Proposed Methodology

We describe our proposed methodology in this section. Fig. 1 represents workflow of the following method. We discuss about the dataset collection and analysis, data processing, MCD based deep learning model creation.

### 3.1 Dataset Collection and Analysis

We use ECG Images dataset of Cardiac and COVID-19 Patients [14], which has 1937 distinct patient records. Data is gathered using ECG Device 'EDAN SERIES-3' placed in different health institutes across Pakistan. The dataset is a 12-lead based standard ECG images for differents patients with five labels such as, **COVID-19**, **Abnormal Heartbeat (HB)**, **Myocardial Infarction (MI)**, **Previous History of MI (PMI)**, and **Normal Person**. Fig. 2 shows some samples from the dataset.

### 3.2 Data Preprocessing

We have preprocessed the ECG images from the dataset by converting them from RGB (three channel) to grayscale (one channel) images and resizing them into $70 \times 70$ resolution. We use the cropped version of this dataset, where the key features of the ECG are cropped [24]. This requires less computation cost and better classification performance. Fig. 3 exhibits the cropping process.

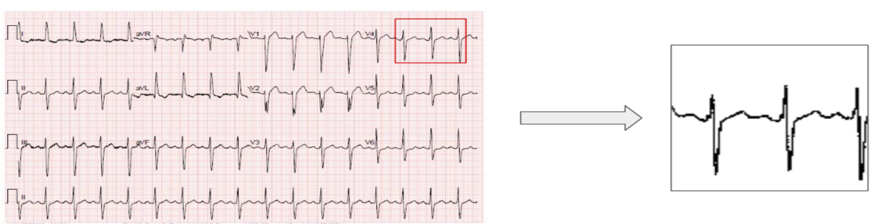


Fig. 3: Processed Data

### 3.3 CNN Architecture

Our CNN model starts with an input image with the resolution of $70 \times 70$ pixels. Then we use six 2-dimensional convolutional layers and with $3 \times 3$ kernel size. Also, to avoid overfitting in the model, we added a batch normalization layer after each convolutional layer. Besides that, we use a max pooling layer after each batch normalization layer with a pool size of $2 \times 2$ to reduce the computational cost. And for all convolutional layers, we use default strides which is (1,1). As the activation function, we use Relu as the non-saturation of its gradient accelerates the convergence of stochastic gradient descent (SGD) compared to the

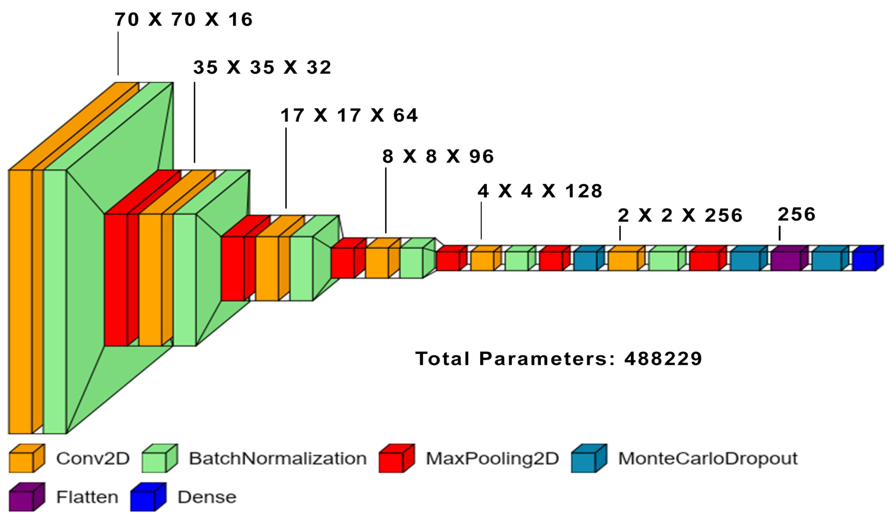



Fig. 4: Visual Representation of Proposed Deep CNN Architecture

Sigmoid/Tanh function. After that, we flatten the values to single dimension and use 20% MCD from that layer. Softmax activation function is added at the end of the output since here all the nodes are classified. The input to the function is transformed into a value between 0 and 1. We use MCD in the last two convolutional block as well, to reduce overfitting and to measure uncertainty later. The model consists of 488229 parameters. Fig. 4 is the visual representation of our proposed architecture.

### 3.4 Monte Carlo Dropout

The general purpose of dropout is to decrease the model complexity and prevent overfitting [25]. During the training phase, each neuron's output, in the dropout layer, is multiplied by a binarymask which is obtained from a Bernoulli distribution. Some of the neurons are set to zero this way. Afterwards, the neural network is used at test phase. The dropout technique can be used as an approximation of probabilistic Bayesian models in deep Gaussian processes, which was presented by Gal and Ghahramani [9]. MCD is a technique of performing numerous stochastic forward passes in a neural network using activated dropout throughout the testing phase to generate an ensemble of predictions that may reflect uncertainty estimations. If we are given a trained neural network model with dropout $f_{nn}$. To derive the uncertainty for one sample $x$ we collect the predictions of $T$ inferences with different dropout masks. Here $f_{nn}^{d_i}$ represents the model with dropout mask $d_i$. So we obtain a sample of the possible model outputs for sample $x$ as

$$f_{nn}^{d_0}(x), ....., f_{nn}^{d_T}(x) \tag{1}$$

We obtain an ensemble prediction by computing the mean and the variance of this sample. The prediction is the mean of the model's posterior distribution for this sample and the estimated uncertainty of the model regarding $x$.

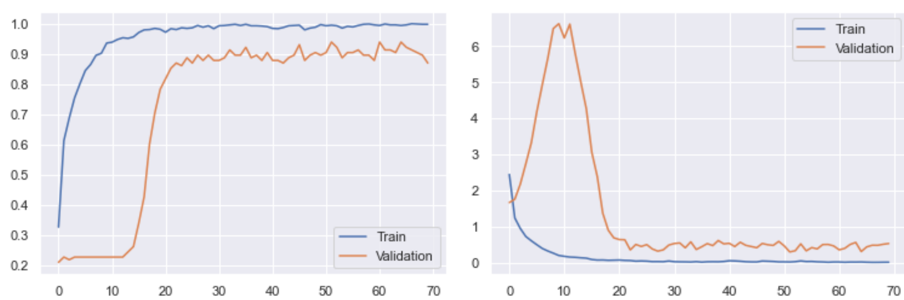
$$Predictive\ Posterior\ Mean,\ p = \frac{1}{T}\sum_{i=0}^{T} f_{nn}^{d_i}(x) \tag{2}$$

$$Uncertainty,\ c = \frac{1}{T}\sum_{i=0}^{T}[f_{nn}^{d_i}(x) - p]^2 \tag{3}$$

We do not alter the dropout neural network model itself but simply collect the results of stochastic forward passes from the model. This is done in order to assess the prognostic mean and uncertainty of the model. Consequently, this data can be implemented with the existing NN models trained with dropout.

Table 1: Results obtained from CNN-MCD method

| Accuracy | Precision | Sensitivity | Loss | F1 Score | AUC Score |
|---|---|---|---|---|---|
| 0.9177 | 0.9167 | 0.9167 | 0.4257 | 0.9167 | 0.9733 |




(a) Accuracy Graph (b) Loss Graph

Fig. 5: Training and Validation Graph for CNN-MCD Method

## 4 Result Analysis and Discussion

### 4.1 Experimentation

The model has been trained and tested grayscale images. Converting colored images from the dataset into grayscale images reduced the size or number of unnecessary features of the images. The images are randomly divided into 8:1:1 training, validation, and test datasets, which also maintains the ratio of disease and non-disease classes in every sub-datasets. After an in-depth experiment, we use 32 training samples per iteration, and set the epochs value to 70.

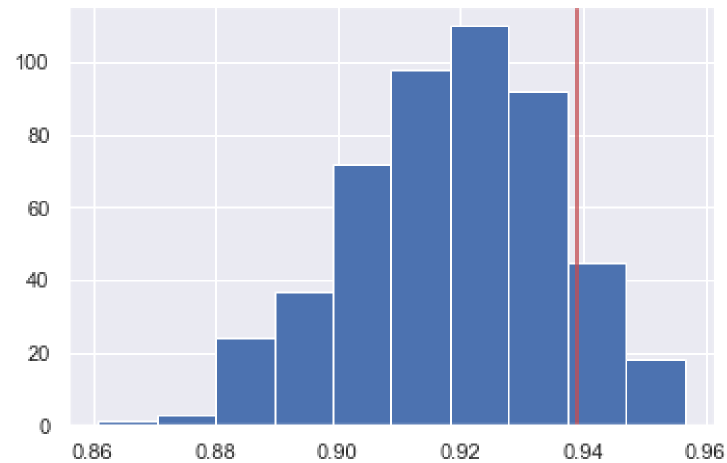



Fig. 6: Distributions of the monte carlo predictions and prediction of the ensemble (Red).

### 4.2 Result Analysis

Following any image classification task, evaluation of the quality of the model depends on performance evaluation metrics. For quantitative evaluation, renowned metrics of performance evaluation, such as Accuracy, Specificity, Sensitivity, and Precision, are employed to determine the performance of the CNN-MCD approach. There are other effective measures for performance evaluation, such as The Area Under the Curve of the Receiver Operating Characteristic (ROC), also known as AUC of ROC which can be applied too.

Table 1 depicts the results of the CNN-MCD approach. We obtained an accuracy of 0.9177; precision is 0.9167. Sensitivity is measured to be 0.9167 while Loss is found to be 0.4257. The overall F1 Score is 0.9167 and the AUC Score is 0.9733. Because of using a real-world dataset, CNN-MCD approach performs poorly with less number of epochs, after 20 epochs the results get more stable as shown in Fig. 5. We stop the training process after 70 epochs.

Fig. 6. shows the distributions of the MCD predictions and prediction of the ensemble, which is acquired by computing the average and the variance of possible model outputs. We find the uncertain samples using mean and variance of the predictions. We show two uncertain samples in Table 2. the probability distributions of their predictions. This helps us understand the dataset and indicates any existing issues within the model. In the graph, classes 0 to 4 are titled covid-19, HB, MI, Normal, PMI respectively. The first sample successfully predicts the label as 'HB', but from the distribution we can see that, the model is quite uncertain about this prediction as it also confuses the input as 'PMI'. The second sample fails to predict the label. The model is very uncertain about labelling its predictions as 'covid-19' or 'PMI'. Hence, MCD helps the model to point out uncertainty in a way that prevents uncertain diagnosis.

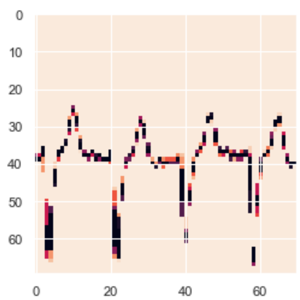
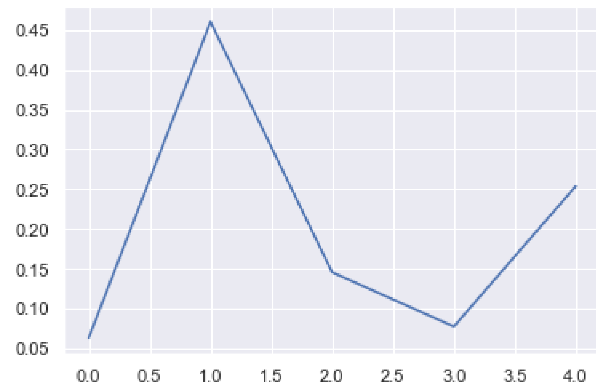
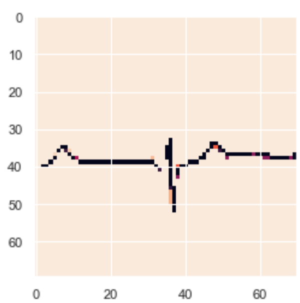
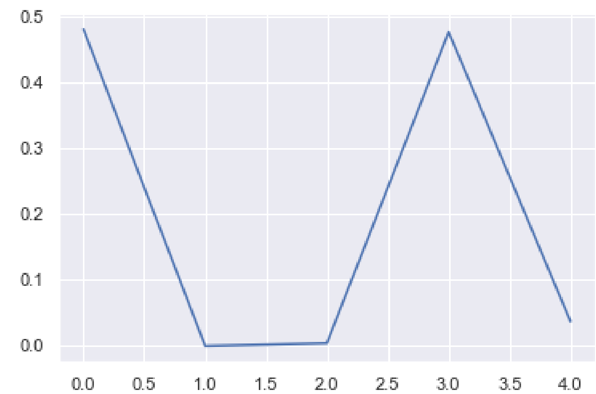
<table>
<tr><th>Uncertain Sample</th><th>Probabilty Distribution of Prediction</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Table 2: Uncertain samples and their predictions

In terms of model performance, our proposed model outperforms all the existing models in multi-class classification shown in table 3. State-of-the-art models such as VGG-16, VGG-19, ResNet-50, DenseNet, InceptionV3 achieves accuracy from 79%-83%. Other methods mentioned in table 3. conducts their classification on three or four classes while our five-class classification accuracy is higher than those.

## 5 Conclusion and Future Work

In our study, we present the uncertainty of a CNN model in ECG Classification using MCD. MCD quantifies the uncertainty and improves the predictive accuracy of the CNN for safer and better analysis of quantitative metrics. Besides, it keeps the model structure the same as before and adds minimal cost to inference time. We use MCD method to do such which is very significant to decrease the risk factors in cardiovascular disease diagnosis. Our proposed

| Method | Number of Classes | Accuracy |
|---|---|---|
| ResNet-50[15] | 5 | 78.08 |
| DenseNet[15] | 5 | 76.83 |
| InceptionV3[15] | 5 | 79.35 |
| VGG-16[15] | 5 | 83.74 |
| VGG-19[15] | 5 | 83.32 |
| CNN Model of [15] | 5 | 83.05 |
| Attention-based CNN with Residual Connections[17] | 4 | 92.00 |
| COV-ECGNET[16] | 3 | 90.8 |
| ECG-BiCoNet[18] | 3 | 91.7 |
| CNN-MCD (Proposed) | 5 | 91.77 |
| CNN-MCD Ensemble (Proposed) | 5 | 93.90 |

Table 3: Comparison among different models.

CNN-MCD method not only surpass state-of-the-art models in a huge margin but also allows users to find uncertain samples to reduce risk factors. Our proposed method can be used only by taking an image of the ECG. This will help to diagnose patients quickly if the specialist is not currently available. The MCD method used in our proposed model helps to identify out-of-scope samples which can be prevented as these are very risky to deal with. Our future goal is to develop a Semi-Supervised model to classify ECG images, as such medical data are very hard to acquire.

## References


1. Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation* **139,** e56–e528. https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000659 (2019).
2. Lown, B., Klein, M. D. & Hershberg, P. I. Coronary and precoronary care. *The American Journal of Medicine* **46.** Symposium on Alherosderosis, 705–724. ISSN: 0002-9343. https://www.sciencedirect.com/science/article/pii/0002934369900229 (1969).
3. Lecun, Y., Kavukcuoglu, K. & Farabet, C. *Convolutional Networks and Applications in Vision* in (May 2010), 253–256.
4. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* **128,** 336–359. https://doi.org/10.1007%2Fs11263-019-01228-7 (Oct. 2019).
5. Chattopadhyay, A., Sarkar, A., Howlader, P. & Balasubramanian, V. N. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV),* 839–847 (2018).

6. Begoli, E., Bhattacharya, T. & Kusnezov, D. F. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence (Online)* **1** (Jan. 2019).
7. Neal, R. *Bayesian Learning via Stochastic Dynamics* in *Advances in Neural Information Processing Systems* (eds Hanson, S., Cowan, J. & Giles, C.) **5** (Morgan-Kaufmann, 1992). https://proceedings.neurips.cc/paper/1992/file/f29c21d4897f78948b91f03172341b7b-Paper.pdf.
8. Gal, Y. *Uncertainty in Deep Learning* PhD thesis (University of Cambridge, 2016).
9. Gal, Y. & Ghahramani, Z. *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning* in *Proceedings of The 33rd International Conference on Machine Learning* (eds Balcan, M. F. & Weinberger, K. Q.) **48** (PMLR, New York, New York, USA, 20–22 Jun 2016), 1050–1059. https://proceedings.mlr.press/v48/gal16.html.
10. Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S. & Acharya, U. R. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine* **161,** 1–13. ISSN: 0169-2607. https://www.sciencedirect.com/science/article/pii/S0169260718301226 (2018).
11. Hong, S., Zhou, Y., Shang, J., Xiao, C. & Sun, J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine* **122,** 103801. ISSN: 0010-4825. https://www.sciencedirect.com/science/article/pii/S0010482520301694 (2020).
12. Attia, Z. *et al.* An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet* **394** (Aug. 2019).
13. Jun, T. J. *et al. ECG arrhythmia classification using a 2-D convolutional neural network* 2018. https://arxiv.org/abs/1804.06812.
14. Khan, A. H., Hussain, M. & Malik, M. K. ECG Images dataset of Cardiac and COVID-19 Patients. *Data in Brief* **34,** 106762. ISSN: 2352-3409. https://www.sciencedirect.com/science/article/pii/S2352340921000469 (2021).
15. ırmak, E. COVID-19 disease diagnosis from paper-based ECG trace image data using a novel convolutional neural network model. *Physical and Engineering Sciences in Medicine* **45,** 167–179 (Mar. 2022).
16. Rahman, T. *et al.* COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network. *Health Information Science and Systems* **10** (2022).
17. Sobahi, N., Sengur, A., Tan, R.-S. & Acharya, U. R. Attention-based 3D CNN with residual connections for efficient ECG-based COVID-19 detection. *Computers in Biology and Medicine* **143,** 105335. ISSN: 0010-4825. https://www.sciencedirect.com/science/article/pii/S0010482522001275 (2022).
18. Attallah, O. ECG-BiCoNet: An ECG-based pipeline for COVID-19 diagnosis using Bi-Layers of deep features integration. *Computers in Biology*

*and Medicine* **142,** 105210. ISSN: 0010-4825. https://www.sciencedirect.com/science/article/pii/S0010482522000026 (2022).

19. Ghoshal, B. & Tucker, A. *On Cost-Sensitive Calibrated Uncertainty in Deep Learning: An application on COVID-19 detection* in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)* (2021), 503–509.
20. Ul Abideen, Z. *et al.* Uncertainty Assisted Robust Tuberculosis Identification With Bayesian Convolutional Neural Networks. *IEEE Access* **8,** 22812–22825 (2020).
21. Jungo, A. *et al.* in, 682–690 (Sept. 2018). ISBN: 978-3-030-00927-4.
22. Ghoshal, B. & Tucker, A. *Estimating Uncertainty and Interpretability in Deep Learning for Coronavirus (COVID-19) Detection* 2020. https://arxiv.org/abs/2003.10769.
23. Milanés-Hermosilla, D. *et al.* Monte Carlo Dropout for Uncertainty Estimation and Motor Imagery Classification. *Sensors* **21.** ISSN: 1424-8220. https://www.mdpi.com/1424-8220/21/21/7241 (2021).
24. NKENGUE, M. J. *ECG_Image_Cropped* https://www.kaggle.com/datasets/marcjuniornkengue/ecg-image-cropped. Version 4. 2022.
25. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* **15,** 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html (2014).