Response to Reviewer 2 Comments

Current report (Round 2):

First comment: The authors have made an effort to address an important research question, and their work has the potential to contribute valuable insights to the field. However, the revised version of the manuscript has not addressed the key concerns raised in the previous report. As pointed out by the referee, the proposed approach lacks novelty and does not provide a clear explanation for its supposed superiority over previous papers. The lack of a comparison table that clearly summarizes the key aspects of the current work versus other reports is a major weakness of the manuscript. The authors MUST clearly state the novelty. What is unique, novel, or better about your work as compared to existing art? The table, which includes the summary and comparison of this current work versus other reports, should be added. (Table Columns: such as references, approaches, modes, data size, imaging techniques, etc.). The authors must claim: "there is no report on…. Now, no clear sentences in the abstract and introduction showing the novelty and the gap of knowledge that still requires you to just simply expand the scope of disease diagnosis by using both X-ray and CT images. Showing only a minor improvement (Table 3) is not enough to show the motivation and significance to be a new publication. A submission may be rejected if it is not significant to the readership, even if it is scientifically sound.

First response: According to first comment, I have revised and added more the detail following:

1.) **The abstract** has been revised to include the following details: In **lines 9-11**, it is noted that "CNN models VGG19, ResNet152V2, and DenseNet121 were combined with long short-term memory (LSTM) and gated recurrent unit (GRU) RNN models, which are convenient to develop because these networks are all available as features on many platforms." This serves to convey that the proposed method can be easily developed and implemented. In **lines 11-14**, it is stated that "The proposed method is evaluated using a large dataset totaling 16,210 X-ray and CT images (5,252 COVID-19 images, 6,154 pneumonia images, and 4,804 normal images) were taken from several databases, which had various image sizes, brightness levels, and viewing angles", this ensures that the performance of the proposed method does not overfit on any particular of the dataset. In **lines 17-19**, it is stated that "From the experimental results, the proposed method is highly effective in distinguishing lung diseases. Furthermore, both CT and X-ray images can be used as input for classification, allowing for the rapid and easy detection of COVID-19" to present the versatility of the proposed method in the classification of lung diseases using multiple types of data images.

2.) **The introduction** has been revised to include the following details: In **line 63**, I have added the limitation of Ref. **[11],** stating: "However, they considered only using X-ray images". In **lines 66-67**, I have added the limitation of Ref. **[12],** stating: "but the method can only classify images as COVID or non-COVID". In **lines 74-76**, I have added the limitation of Ref. **[12]**, stating: "Although the method has a high classification efficiency, it required two different CNN networks to operate at its two stages: segmentation and classification, demonstrating that they use many process to provide output". In **lines 95-96**, I have added the limitation of Ref. **[17]**, stating: "but they evaluate it only on CT images". In **lines 97-103**, I have provided a summary of the limitations of previous studies and proposed a method to address and improve upon these limitations, it is noted that "Even though previous studies have presented diverse approaches to detect COVID-19 and classify lung disease, most of the research has considered only X-ray or CT images, not both. Also, several researchers evaluated their method using a small number of datasets, making it difficult to ensure that the performance would be replicated when these methods were tested on a larger dataset. Thus, this research aims to present the combined CNN-RNN network to distinguish three classes of lung disease (i.e., COVID-19, pneumonia, and normal) that can be used in both X-ray and CT images as input". In **lines 104-112**, I have presented information on the rationale for combining CNN-RNN, and assert that no prior studies have compared the classification performance of different CNNs when combination with different versions of RNN for distinguish lung illnesses, it is

noted that “In this study, the CNN was combined with the RNN to improve the classification result because the CNN has high-efficiency characteristics for feature extraction but is un-connected across nodes within the same layer. while the RNN contains a property that analyzes dependencies and continuity from earlier information, which assists in pattern recognition from extracted features. The widely used CNN models VGG19, ResNet152V2, and DenseNet121 were combined with a version of RNN, including LSTM and GRU, to determine which combination of CNN and RNN architecture yielded the best results. There is no report that compares the performance of different CNNs when combined with different RNN versions”. Finally, In **lines 112-115**, I have noted the dataset utilized in this study to demonstrate that the proposed method is adaptable to input and exhibits high reliability upon evaluating with large datasets, stating: “In this work, we collected a large dataset of 16,210 X-ray and CT images from various sources with varying image sizes, brightness levels, and viewing angles to use for experimentation, which guaranteed the method’s flexibility to handle input data and high reliability.”

3.) **The discussion** has been revised to describe the proposed method's superiority over previous methods, it is noted that “Another advantage of our method is flexibility, as it was able to classify between COVID-19, pneumonia, and normal cases from both X-ray and CT images. A comparison between proposed method and another methods in terms of accuracy, precision, recall, and F1-score is shown in Table 3. From Table 3 is demonstrated that the proposed method is more accurate than [14,15,33–38], which evaluated the effectiveness of their approaches only using X-ray or CT images. Perumal et al. [16] used X-ray and CT images for lung disease classification, but their methods were tested on a small dataset and provided an accuracy lower than the proposed method. However, Aslan et al. [13] reported a better method, since it not only had complex processes to deliver output but also used only chest X-ray images to classify COVID-19, pneumonia, and normal cases. Hamed et al. [17] achieved high performance, but they only discussed binary classification (COVID or non-COVID) based on CT images to classify, showing that their method is non-flexible for input format and has less categorical ability than the proposed method, which can classify three lung disease categories” in **lines 293-306**.

Second comment: Furthermore, the authors have failed to provide a detailed letter addressing the specific points raised by the referee. It is essential for authors to carefully consider and respond to the feedback provided by reviewers to improve the quality of their work. Ignoring or failing to address the comments and concerns of the reviewers is a serious issue. The authors MUST describe what specific changes have been made in response to the points in the revised manuscript and on which pages; send us your revised paper with tracking changes too.

Second response: I apologize for this and thank you very much for your suggestion. I have already revised the previous report that is shown in pages 3-5 of this report.

Previous report (Round 1):

Point 1: Please rewrite some complex sentences.
Response 1: I already rewrote it and sent the manuscript to Cambridge Proofreading for English revisions.

Point 2: For the abstract, add one sentence demonstrating a knowledge gap. Now that you've outlined what we already know, tell us what we don't know. How does your work address a specific question?
Response 2: In **lines 5-9**, I have addressed the solution of combining CNN and RNN, as well as the advantages of utilizing the CNN-RNN approach. Specifically, I have stated that "we combined convolutional neural network (CNN) and recurrent neural network (RNN) models by replacing the fully connected layers of CNN with a version of RNN. In this framework, the attributes of CNNs were utilized to extract features and those of RNNs to calculate dependencies and classification base on extracted features."

Point 3: While it mentions that the method has "high level of dependability and comparable performance," it does not provide any information about what this means in terms of the accuracy, sensitivity, or specificity of the method.
Response 3: I already revised that sentence to "From the experimental results, the proposed method is highly effective in distinguishing lung diseases. Furthermore, both CT and X-ray images can be used as input for classification, allowing for the rapid and easy detection of COVID-19." in **line 17-19** and provide the information about accuracy, precision, recall, and F1-score in section ***3.3. Evaluation***.

Point 4: To make the introduction and discussion more substantial, the authors may wish to introduce more examples and discussion. (such as 10.1016/j.asoc.2021.107918; 10.1016/j.chemolab.2022.104695, etc)
Response 4: In the section ***1. Introduction***, I have referred to the papers 10.1016/j.asoc.2021.107918 and 10.1016/j.chemolab.2022.104695 as ref. **[13,17]** and I have discussed their methods in section ***4.Discussion***.

Point 5: Please give some details about resolution each original dataset. And how to control.
Response 5: In section ***2.1 Data Sets*** I have included the image size details for each original dataset. In section ***3.1 Data Pre-processing*** I have explained the pre-processing methods utilized for the control image data.

Point 6: Line 74-77: We increase the scope of disease diagnosis by making use of both X-ray images and CT scans. This allows us to take advantage of the computational speed and prediction accuracy of CNNs and RNNs, respectively, in order to forecast X-ray and CT scans of COVID-19 images.
Response 6: Apologies, I had assumed that your specific pertained to point 7.
Point 7: Please add a result about "computational speed" ?
Response 7: Training and test time durations are presented in **Table 2** and further analyzed in **lines 258-264**. I have stated that "In terms of training time, the ResNet152V2 with GRU model utilizing the original images took the least amount of time to train the network—99 minutes and 57 seconds—followed by the DenseNet121 with LSTM model utilizing the normalized images and the VGG19 with LSTM model utilizing the normalized images—103 minutes and 55 seconds and 115 minutes and 1 second, respectively. Meanwhile, in terms of predicted time, DenseNet121 achieved the fastest time at 0.08 seconds per image, followed by VGG19 at 0.16 seconds per image and ResNet152V2 at 0.21 seconds per image."

Point 8: Line 105-106 "the remaining images were excluded, due to their small resolution." What is the minimum resolution required for this research? Please add references.
Response 8: I apologize for the mistake. I have revised the sentence in **line 137-138** to read: "The remaining images were randomly excluded to balance the number of data in each class of lung diseases"

This was done in order to balance the data for the three classes used in our study, and not for the purpose of removing small resolution images.

Point 9: Line 193: Why this research use 300 epochs for training? Please add references.
Response 9: I have added reference **[32]** in order to provide support that it offers the best accuracy for training CNN to detect COVID-19 based on the experiments conducted by the authors of that work.

Point 10: Line 237: "During the testing phase, Figure 6 depicts the confusion matrix of our proposed". The "Figure 6" is not refer to the figure directly. Please edit it.
Response 10: I already refer "Figure 6" to the figure directly.

Point 11: Table 3. For Saxena et al. [31] show F1-score as 93.78 which is higher than the proposed method, so it should be bold.
Response 11: I already fixed it and also include your suggest papers in point 4 (10.1016/j.asoc.2021.107918 and 10.1016/j.chemolab.2022.) to **Table 3.**

Point 12: Figure 4. The combined CNN–RNN network structure. The figure shown the dropout dashed line, but there is no mention about "dropout" in the context. Please add more details and references.
Response 12: I have included more details about the use of the dropout technique in the sentence: “In the fully connected layer, the dropout technique was used to avoid overfitting in the networks” and have added references **[30,31]**.

Point 13: Figure 4. The combined CNN–RNN network structure. Is the last layer of among VGG19, ResNet152V2, and DenseNet121 used same reshape parameters for combining with the RNN model? Please add more details and references.
Response 13: I have added more details about the reshape parameters for combining CNN-RNN in **lines 200-203**, stating: “We reshaped the output of VGG19 (none, 7, 7, and 512), ResNet152V2 (none, 7, 7, and 2,048), and DenseNet121 (none, 7, 7, and 1,024) to (49,512), (49, 2048), and (49,1024) respectively.” and have included reference **[11]** to support this technique.

Point 14: For the Table 3, Is it fair to compare the other research with the number of dataset lower than the proposed method?
Response 14: The use of a larger number of images, featuring various sizes, brightness levels, noise levels, and angles of view, to evaluate the performance of the method demonstrates its greater flexibility in handling input data and increased reliability. As shown in **Table 3**, the methods described in references **[35, 36]** are similar to the proposed methods in terms of the number of datasets utilized. Also describe about the number of data to evaluate method in **lines 99-101**, stating: "several researchers evaluated their method using a small number of datasets, making it difficult to ensure that the performance would be replicated when these methods were tested on a larger dataset" and **lines 112-115**, stating: "In this work, we collected a large dataset of 16,210 X-ray and CT images from various sources with varying image sizes, brightness levels, and viewing angles to use for experimentation, which guaranteed the method's flexibility to handle input data and high reliability."

Point 15: The authors should describe more clearly the ethics. For ethical considerations, experiments including human data and relevant approval protocols should properly be documented.
Response 15: In this research, we are mindful of the ethical concerns surrounding the use of personal data and have taken a cautious approach. However, we have used an existing public dataset to ensure that there is no infringement on individuals' privacy.

Point 16: Please add limitation or drawbacks in this work?

Response 16: I already added limitations in this work at **lines 307-310**, stating: “The limitations of our proposed method are the GPU utilization and high RAM consumption (minimum of 16 GB) for the model training. This is because our method involved a deep CNN network and combined it with an RNN model, which may not be suitable for low-resource devices.”.

Point 17: For ‘Conclusions’, the challenges in studying this work, that the authors have addressed, should be emphasized.

Response 17: In the revised conclusion, I have emphasized the challenges of this study, including the use of CT and X-ray images, which are different types of diagnostic imaging, and the varied image sizes, brightness levels, noise levels, and angles of view from various data sources used to evaluate the proposed method.