Research Article
COVID-19 Data Analytics Using Extended Convolutional Technique

Anand Kumar Gupta,1 Asadi Srinivasulu,1 Olutayo Oyeyemi Oyerinde,2 Giovanni Pau,3 and C. V. Ravikumar4

1Data Science Research Laboratory, BlueCrest University College, Monrovia, Liberia
2School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg 2050, South Africa
3Kore University of Enna, Faculty of Engineering and Architecture, Enna 94100, Italy
4Sense, Vellore Institute of Technology, Vellore, Tamil Nadu, India

Correspondence should be addressed to Anand Kumar Gupta; head.sp@bluecrest.edu.lr

Received 18 August 2022; Revised 25 October 2022; Accepted 27 October 2022; Published 7 November 2022

Academic Editor: Payam Behzadi

Copyright © 2022 Anand Kumar Gupta et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The healthcare system, lifestyle, industrial growth, economy, and livelihood of human beings worldwide were affected due to the triggered global pandemic by the COVID-19 virus that originated and was first reported in Wuhan city, Republic Country of China. COVID cases are difficult to predict and detect in their early stages, and their spread and mortality are uncontrollable. The reverse transcription polymerase chain reaction (RT-PCR) is still the first and foremost diagnostic methodology accepted worldwide; hence, it creates a scope of new diagnostic tools and techniques of detection approach which can produce effective and faster results compared with its predecessor. Innovational through current studies that complement the existence of the novel coronavirus (COVID-19) to findings in the thorax (chest) X-ray imaging, the projected research’s method makes use of present deep learning (DL) models with the integration of various frameworks such as GoogleNet, U-Net, and ResNet50 to novel method those X-ray images and categorize patients as the corona positive (COVID + ve) or the corona negative (COVID -ve). The anticipated technique entails the pretreatment phase through dissection of the lung, getting rid of the environment which does now no longer provide applicable facts and can provide influenced consequences; then after this, the preliminary degree comes up with the category version educated below the switch mastering system; and in conclusion, consequences are evaluated and interpreted through warmth maps visualization. The proposed research method completed a detection accuracy of COVID-19 at around 99%.

1. Introduction

The ailment referred to as “the extreme acute breathing syndrome coronavirus 2 (SARS-CoV-2)” was determined by the end of December 2019. Accordant to reports, this disease was originated in China and has become the reason for the disorder referred to as “Corona Virus Disease 2019,” in short, “COVID-19.” The specialized health institution of the USA, the “World Health Organization” (WHO), stated that this disorder is a “deadly disease” in mid-March 2020 [1–5]. As per reviews delivered to date with the aid of using worldwide health organizations, authorities/entities, and governments, the pandemic has affected tens of thousands and thousands of human beings globally. The maximum severe contamination due to COVID-19 is associated with the lungs, which includes pneumonia. The signs and indications of the disorder may range and consist of excessive body temperature (high fever), dyspnea, coryza, and cough. These instances can be normally recognized through the usage of lung X-ray evaluation of the irregularities [6]. This research is the latest in its own field, targets the global pandemic caused by the coronavirus, and helps scholars to implement combat tools and techniques following the proposed method in this research work.
Throughout this hasty period, several scholars have tried to expand numerous transmission gear and diagnosing systems, such as the real-time reverse transcription polymerase chain reaction (RT-PCR), which is still claimed as a vital testing technique to discover extremely severe breathing disease SARS-COV2 [1–3] and in addition to COVID-19. Though RT-PCR is considered to be the best method of screening so far, it still has limitations. The working system of the RT-PCR is complex and cumbersome [2–5]. The input dataset, which has 39,256 X-ray images, was collected from various publicly available dataset repositories such as Kaggle, UCI, Google, and GitHub. This research is purely conducted on the sample datasets collected from these various repositories, which require no consent from human being. Thus, trials were attempted for detecting COVID-19 disease through lung X-ray images which include CT (computed tomography) or lung x-beam photographs. It is said the investigative significance and accurateness of CT lung photographs over RT-PCR in COVID-19 have been offered with the aid of using authors together with Li et al. [13], Xu et al. [14], Gozes et al. [15], and Shi et al. [16]. Authors [14] mentioned as follows as the COVID-19 epidemic shows its traits which can change starting with different varieties of virus-related pneumonitis, such as biological influenza-A pneumonia [5,17,18]. The goal of this research study has become to broaden a preliminary testing outline for COVID-19 with the aid of using automatic respiratory CT scans (CT photographs) of COVID-19, pneumonia, and ordinary instances. They considered 628 CT scans and X-ray test pattern images earlier than expansion, and their version acquired an accuracy of 90%. The writers’ approach consists of image preprocessing, dissection of more than one region (patches) accepting V-Net (volumetric network) [19] based on separation version V-Net-IR-RPN [20], that has skilled for pulmonic tuberculosis resolution.

Our method includes three (3) essential experiments to assess the overall performance of the predication and determine an effect on the distinctive levels of the procedure. The respective test follows the workflow. The distinction among trials is the dataset used from various repositories [21, 22]. In all occurrences, identical photographs of COVID-19 effective instances were used. Meanwhile, 3 distinctive datasets for poor instances were utilized. In that direction, experiments 1 and 2 included comparing effective vs. poor instances datasets Figure1, and experiment three contains pre-COVID generation images (from 2015–2017), further Table1 explains the comparative analysis of the proposed model (ECNN) with the existing system (CNN) in details through various parameters Figure 2.

Muhammad Irfan et al. developed a combined model by using CT and X-ray imaging concept named as HDNNs (hybrid deep neural networks) to predict severity levels of COVID-19 infectious by classifying the data into 3 classes such as pneumonia, normal, and COVID-19 positive [20]. Yassir Edrees Almalki et al. proposed a unique framework named as COVID Inception-ResNet (CoVIRNet) model, which uses chest X-ray images to assess the infection of COVID-19 naturally. The projected calculation has different initiation remaining blocks that take special care of data by utilizing various profundities highlight maps at various scales, with the different layers [23].

3. Methodology

3.1. Existing System. The referred existing system has approachable superficial learning strategies, for example, The Convolution Neural Organization and Intermittent Neural Organization. CNN computation drawbacks are as follows:

(i) Less efficient
(ii) High time complexity
(iii) High execution time
(iv) Highly error prone
(v) Supports only a small dataset

3.2. Proposed System. The proposed research work was carried out by using the available deep learning method convolutional NN (neural network) by using combined
methods of AI, ML, and DL and hence was named as an extended convolutional neural network (ECNN).

ECNN algorithm advantages are as follows:

(i) Highly accurate
(ii) Less time consuming
(iii) Less Performance time
(iv) Less error rate
(v) Big Data Scope

4. Experimental Results

The essential thought is to execute, guarantee, and ensure that the coronavirus illness severer impacted job gathered measurements worked in the manner that can force readiness and development from their most memorable viewpoint.

The proposed prototype outperformed with an accuracy of 99.34% via using 39,256 X-ray image datasets Figure 3. The following section describes the findings of the research, which includes the algorithm steps Table 2, input dataset Figure 4, advantages of the proposed prototype compared against existing techniques Table 1, and final results obtained by using the proposed prototype against various metrics such as time consumption Figures 5 and 6, accuracy Figure 7, loss Figure 8, precision, etc. through demonstrated figures from Figures 1–3 and 5–9.

4.1. ECNN Algorithm. Two trials of one or the other CC or MLO seen should be adjusted utilizing the picture enlistment method. At that argument, a dissimilar picture was received by removing the former trial from the existing trial and subsequently scaling to the full-range force. The territorial pictures from the refined district proposition are trimmed from the three pictures and scaled to $224 \times 224 \times 3$ for each picture, which are utilized for ECNN floodlight extraction. As stated in Figure 10, the three channels are rehashed from one-channel grayscale pictures (e.g., the current sweep of $224 \times 224 \times 1$) since the pertained ECNN and ECNN models expect 3-channel pictures. Multimeasurements of three-state in floodlights (from the earlier sweep, current output, and contrast pictures) are made to prepare a CNN model. For instance, The ECNN is floodlights utilizing ResNet-60V3 of $2048 \times 3$ measurements for each view (CC or MLO) of a subject’s side (left or right bosom). Remember that the earlier sweep consistently relates to the ordinary (sound) status in any event for a destructive subject. According to Table 2, assuming we code sound and carcinogenic as 0 and 1 individually, at that point, the ground realities (yields) compared with the three states (earlier, current, distinction) of the destructive view are [0 1 1] Figure 11. This coding instrument can be handily stretched out to at least two earlier sweeps.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Name of the parameter</th>
<th>CNN</th>
<th>ECNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Accuracy</td>
<td>88.28%</td>
<td>99.34%</td>
</tr>
<tr>
<td>2.</td>
<td>Error rate</td>
<td>0.46</td>
<td>0.13</td>
</tr>
<tr>
<td>3.</td>
<td>Val_Loss</td>
<td>4.62</td>
<td>3.94</td>
</tr>
<tr>
<td>4.</td>
<td>Val_Accuracy</td>
<td>0.79</td>
<td>0.67</td>
</tr>
<tr>
<td>5.</td>
<td>Size of the dataset</td>
<td>1.75 GB</td>
<td>1.75 GB</td>
</tr>
<tr>
<td>6.</td>
<td>No. of epochs</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>7.</td>
<td>Time-complexity</td>
<td>O(n^2)</td>
<td>O(n^2)</td>
</tr>
<tr>
<td>8.</td>
<td>Execution time</td>
<td>1370 ms</td>
<td>1129 ms</td>
</tr>
</tbody>
</table>

Figure 1: COVID-19 ECNN graph comparing Val_Accuracy vs. Val_Loss.

Table 1: Comparison metrics with the existing system and ECNN technique.
4.2. Algorithm. The following are the ECNN algorithm steps:

The COVID-19 disease infection data index, i.e., the absolute 39,256 pictures, of our experiment involved the related following steps:

1. Introduces mandatory collection
2. Introduces training dataset
3. Executes in the floodlight ordering of change data
4. Alignment with 70-time segments and 2 yield
5. Introduces Keras (Keras is a deep learning library)
6. Resets ECNN
7. Enhances ECNN part and regulates loss calculation function.
8. Improvement of yield part.
9. Adds the ECNN
10. Adjusts ECNN in the assessment dataset
11. Loads the COVID-19 disease infection test image data for the year 2020
12. Predict COVID-19 disease severity
13. Envision delayed consequences with expected or real Coronavirus illness contamination

Input Dataset: the input dataset has 16 columns with the target class as severity level of the COVID-19 disease and Figure 4 consists of 39,256 X-ray images with an average of 2111 pixels X 1509 pixels (width X height).

The ECNN calculation benefits: the benefits are as follows:

(i) Tremendous accuracy
(ii) A lesser measure of time
(iii) Less implementation time
(iv) Less fault frequency
(v) Enormous data amount

The proposed block diagram Figure 10 demonstrates the number of input layers, hidden layers, dropout layers, pooling layers, dense layers, activation layers, and convolutional layers to predict and forecast the targeted output.

5. Discussion of Results

The outcomes of COVID-19 disease detection and prediction by incorporating X-ray images dataset with the ECNN technique are illustrated through figures.

Figure 11 shows the execution flow of the ECNN code on COVID-19 database analyzing time taken, accuracy, loss, Val_Loss, and Val_Accuracy with respect to epochs.

The proposed model in Figure 9 achieves an accuracy of 99.34% on the database collected and used from Kaggle and UCI repositories.

The image in Figure 5 displays the CPU and related resources occupancy of the computer during ECNN code execution on the COVID-19 database.

5.1. Performance Evaluation Methods. The general trial result is estimated and introduced utilizing the most broadly utilized factual methodologies, for example, exactness, accuracy, review, F1-score, responsiveness, and particularity. For study use, because of the restricted examples, the generally speaking measurable outcomes are addressed with a 95% certainty stretch followed by recently revealed writing that likewise utilized a little dataset [13, 23]. In our dataset, Omicron may be delegated genuine positive (Tp) or genuine negative (Tn),
Figure 3: COVID-19 ECNN graph comparing accuracy vs. epoch.

Table 2: ECNN proposed model steps.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Import required libraries</td>
</tr>
<tr>
<td>2</td>
<td>Preprocessing of the dataset</td>
</tr>
<tr>
<td>3</td>
<td>Combined CNN with extended neurons</td>
</tr>
<tr>
<td>4</td>
<td>Perform 10-folded cross-validation with 2 classes</td>
</tr>
<tr>
<td>5</td>
<td>Import Keras deep learning library with all supported libraries</td>
</tr>
<tr>
<td>6</td>
<td>Reset all parameters of ECNN</td>
</tr>
<tr>
<td>7</td>
<td>Enhance the ECNN part and regulation of the loss calculation function</td>
</tr>
<tr>
<td>8</td>
<td>Enhancement of yield part of 10-folded with 2 classes</td>
</tr>
<tr>
<td>9</td>
<td>Accumulate the ECNN parameters</td>
</tr>
<tr>
<td>10</td>
<td>Adjusting the ECNN in the preparation of model</td>
</tr>
<tr>
<td>11</td>
<td>Load the COVID-19 disease infection image dataset</td>
</tr>
<tr>
<td>12</td>
<td>Predicting the infection severity by classifying the dataset into 2 classes</td>
</tr>
<tr>
<td>13</td>
<td>Outcome of the trained model and stop the model</td>
</tr>
</tbody>
</table>

Figure 4: Sample input dataset of the projected prototype for COVID-19 disease detection.
Figure 5: Processor and related resources occupancy of the computing device.

Figure 6: ECNN_COVID-19 graph comparing epochs vs. time.

Figure 7: COVID-19 ECNN graph comparing accuracy vs. loss.
assuming people are analyzed precisely, and it very well may be characterized as bogus positive (Fp) or misleading pessimistic (Fn) if misdiagnosed. The assigned measurable measurements are made sense of in subtleties beneath.

Accuracy: the exactness is the general number of effectively recognized occasions across all cases. Utilizing the accompanying recipes, precision is not entirely settled.

\[
\text{Accuracy} = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}.
\]

(1)

Precision: precision is evaluated as the proportion of precisely anticipated positive results out of completely anticipated positive results.

\[
\text{Precision} = \frac{Tp}{Tp + Fp}.
\]

(2)

Recall: recall alludes to the proportion of significant results that the calculation precisely distinguishes.

\[
\text{Recall} = \frac{Tp}{Tn + Fp}.
\]

(3)

Sensitivity: sensitivity alludes to the main exact positive metric comparative with the complete number of events and can be estimated as follows:

\[
\text{Sensitivity} = \frac{Tp}{Tp + Fn}.
\]

(4)

Specificity: it distinguishes the quantity of precisely recognized and determined genuine negatives and can be tracked down utilizing the accompanying recipe:

\[
\text{Specificity} = \frac{Tn}{Tn + Fp}.
\]

(5)

F1-score: the F1 score is the symphonious mean of accuracy and review. The greatest conceivable F score is 1, which shows an amazing review and accuracy (99.34%).

Figure 8: COVID-19 ECNN graph comparing epochs vs. loss.

Figure 9: Final results of COVID-19 by using the ECNN approach.
\[ F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \]  \hspace{1cm} (6)  

\[ \text{AUC} = \frac{\sum ri(xp) - xp((xp + 1)/2)}{xp + xn}. \]  \hspace{1cm} (7)  

5.2. Evaluation Methods. The following are measurements of evaluation methods or metrics.

\[ \text{Quality} = \frac{BP + VM}{BP + VP + BM + VM}, \] \hspace{1cm} (8)  

\[ \text{Preciseness} = \frac{BP}{BP + VP}, \]  

\[ \text{Callback} = \frac{BP}{BP + VM}, \]  

\[ F - \text{measure} = \frac{2 \times \text{Preciseness} \times \text{Callback}}{\text{Preciseness} + \text{Callback}}. \]
5.3. Input Dataset. The input dataset has 16 columns with target class, i.e., the severity level of the COVID-19 disease, and consists of 39,256 X-ray images. The X-ray image(s) resolution of the input dataset is as follows: Width 2111 pixels x Height 1509 pixels.

The graph Figure 6 demonstrates the time taken to complete the iteration of epochs.

Figure 8 explicates the loss ratio accordant with each epoch during execution.

Figure 3 demonstrates the accuracy (99.34%) achieved against each epoch during execution.

Figure 7 demonstrates how the loss is reduced and accuracy gained from the training model of ECNN with respect to each iteration.

Value loss and value accuracy gained from the ECNN model during training are explained in Figure 1.

At a glance representation of the comparison among epochs, loss, accuracy of the ECNN model is described in Figure 2.

5.4. Comparison Table. Table 1 explains the comparison of the proposed technique with respect to various parameters in the existing system.

Table 1 explains the comparison factors of both techniques based on multiple parameters.

This research proposes research gaps in the existing system, which are less efficient, high time-complexity prone, high execution-time system, produces high error rate, and has no support for big datasets. To bridge these gaps, the proposed model outperforms in terms of all and is highly efficient, has less time complexity, less execution-time system, produces fewer errors, and supports big datasets.

6. Conclusion

This method suggests how the present prototype may be beneficial for more than one task, specifically if it is far taken into consideration that the modified U-Net prototypes do now no longer have higher overall performance. Also, it is proven how X-ray images’ noise may produce predisposition with inside the prototypes. Most metrics display photographs without dissection as higher for categorizing COVID infections. The additional evaluation suggests that even though benchmarks are higher, those prototypes are primarily created on totally seen diagnosis throughout lung’s X-ray as clean proof of COVID-19, so actual correct prototypes ought to be centered on lung elements for classification. In this situation, dissection is desired for dependable outcomes by lowering this bias. The transfer is getting to know that changed into crucial of the outcomes offered. As proven categorized models, the use of this approach wants 40 and 50 epochs to converge, even as segmentation prototypes without modification were approximately 282. The sequence of prototypes was obtainable to decide COVID-19 disease in chest X-ray images with an overall precision of 99% by categorizing COVID-19-positive and COVID-19-negative images. In the meantime, solitary for the COVID-19 label, the method achieved an average of 98.58% accuracy through a metric based-deep analysis at the database for a threshold of 0.4. Changing the edge suggest a growth for the accuracy of proposed architecture by as much as 99%.

The segmentation work suggests an excessive opportunity of imparting more statistics to elements in all experiments, concluding the unconventional outcomes through dissecting lungs and including statistics mixed with surrounding noise. The noise is related to wires used in medical equipment’s, patient’s gender, and/or age, making photographs without lungs have extra information for classification in those situations. Either destiny efficacy or the use of prototypes without lungs should have been the very best possibilities for mislabeling photographs due to errors. Further research studies are required of section diagnosis recognized by the expert radiotherapist to make sure that any noise is a causing object for biased results. It is likewise critical to spot that the outcomes offered do now no longer always suggest the identical overall performance in each database. For example, the used database was collected of Asian victims; different international sufferers might also additionally display minor facts seize modifications or diagnosis, assuming a higher type is wanted the use of international databases. In addition, setting apart the databases through gender will offer extra statistics at the prototype’s scope because the tiny tissues of the chest might also additionally cover elements of the lungs, and it is far unidentified in case or not that it is taken into consideration a partiality with inside the forecast of the prototypical. The future work of this proposed approach can further be extended and implemented in real-time healthcare facilities to detect and predict COVID-19 infection in less time.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request (head.sp@bluecrest.edu.lr).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Anand Kumar Gupta contributed to the conceptualization, references collection, data curation, formal analysis, methodology, software, data curation, investigation, resources, software selection, and writing the original draft. Dr. Asadi Srinivasulu contributed to the methodology, supervision, writing review and editing, project administration, visualization, and investigation. Dr. Olutayo Oyeyemi Oyerinde participated in the formal analysis. Prof. Giovanni Pau contributed to proofreading of the research article in data collection and environment, and Dr. Ravikumar CV participated in the plagiarism check and correction process.
Acknowledgments

This research work was funded by the National Research Foundation (NRF) of South Africa in terms of both NRF Competitive Program (funding) for Rated Researchers with grant number 118547 and funding based on NRF/NCoR joint science, and technology research collaboration with grant number 118678.

References


