COVLIAS 3.0 XEDL : Multicentre, Cloud-Based, Explainable Ensemble Artificial Intelligence Deep Learning System for COVID- 19 in Computed Tomography Scans

Arun K. Dubey
Bharati Vidyapeeth's College of Engineering
Sushant Agarwal
GBTI
Gian Luca Chabert
Azienda Ospedaliero Universitaria (A.O.U.)
Prabhav Sanga
Bharati Vidyapeeth's College of Engineering
John Laird
Adventist Health St. Helena
Inder M Singh
Manudeep K Kalra
Massachusetts General Hospital
Klaudija Viskovic
University Hospital for Infectious Diseases
Narpinder Singh
Deemed to be University
Mostafa M. Fouda
Idaho State University
Rajesh Singh
Uttaranchal Institute of Technology, Uttaranchal University
Deepak Garg
SR University
Gobinath Ravindran
SR University
Luca Saba
Azienda Ospedaliero Universitaria (A.O.U.)
Jasjit S. Suri (jasjit.suri@atheropoint.com)
GBTI

Research Article

Keywords: COVID-19, Control, ResNet-SegNet, feature fusion, ensemble deep learning, validation

Posted Date: December 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3688115/v1

Additional Declarations: No competing interests reported.

COVLIAS 3.0$_{XEDL}$: Multicentre, Cloud-Based, Explainable Ensemble Artificial Intelligence Deep Learning System for COVID-19 in Computed Tomography Scans

Arun K. Dubey, PhD[1], **Sushant Agarwal**, BTech[2], **Gian Luca Chabert**, MD[3], **Prabhav Sanga**, B.Tech[4], **John Laird**, MD[5], **Inder M Singh**, MD[6], **Manudeep K Kalra**, MD[7], **Klaudija Viskovic**, MD PhD[8], **Narpinder Singh**, PhD[9], **Mostafa M. Fouda**[10], **Rajesh Singh**, PhD[11], **Deepak Garg,** PhD[12], **Gobinath Ravindran,** PhD[13], **Luca Saba**, MD[14], **Jasjit S. Suri**, PhD, MBA, FIEEE, FAIMBE, FSVM, FAIUM, FAPVS[15]*

[1,4]Bharati Vidyapeeth's College of Engineering, New Delhi 110063, INDIA
[2]Advanced Knowledge Engineering Centre, GBTI, Roseville, CA, USA
[3]Dept. of Radiology, Azienda Ospedaliero Universitaria (A.O.U.), Cagliari, ITALY
[5]Heart and Vascular Institute, Adventist Health St. Helena, St. Helena, CA 94574, USA
[6]Stroke Monitoring and Diagnostic Division, AtheroPoint™, Roseville, CA 95661, USA
[7]Department of Radiology, Massachusetts General Hospital, Boston, MA, USA
[8]Dept. of Radiology and Ultrasound, University Hospital for Infectious Diseases, 10000 Zagreb, CROATIA
[9]Dept. of Food Science, Graphic Era, Deemed to be University, Dehradun, Uttarakhand, INDIA
[10]Dept. of ECE, Idaho State University, Pocatello, ID 83209, USA
[11]Division of Research and Innovation, Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, India -248007
[12]School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana-506371
[13]Department of Civil Engineering, SR University, Warangal, Telangana-506371
[14]Dept. of Radiology, Azienda Ospedaliero Universitaria (A.O.U.), di Cagliari - Polo di, Cagliari, ITALY.
[15]Stroke Monitoring and Diagnostic Division, AtheroPoint™, Roseville, CA 95661, USA

***Corresponding Author:**
Dr. Jasjit S. Suri, PhD., MBA, FIEEE[p], FAIMBE[q], FAIUM[r], FSVM[s], FAPVS[t]
[p]Fellow, Institute of Electrical and Electronics Engineers
[q]Fellow, American Institute of Medical and Biological Engineering
[r]Fellow, American Institute of Ultrasound in Medicine
[s]Fellow, Society of Vascular Medicine
[t]Fellow, Asia Pacific Vascular Society
Stroke Diagnosis and Monitoring Division, AtheroPoint™, Roseville, CA 95661, USA
Phone: (916)-749-5628; Email**:** jasjit.suri@atheropoint.com

COVLIAS 3.0$_{XEDL}$: Multicentre, Cloud-Based, Explainable Ensemble Artificial Intelligence Deep Learning System for COVID-19 in Computed Tomography Scans

Abstract

Background and Motivation: Lung computed tomography (CT) techniques have been utilized in the intensive care unit (ICU) for COVID-19 disease characterization due to its high-resolution imaging. Artificial Intelligence (AI) has significantly helped researchers in diagnosing COVID-19, and the proposed study hypothesized that the cloud-based *explainable ensemble deep learning* (XEDL) paradigm is superior to transfer learning (TL) models for disease classification.

Methodology: We propose a cloud-based ensemble deep learning (EDL) approach to classify COVID-19 versus Control patients. In the proposed study two cohorts are used: (i) 80 Croatian COVID-19 and (ii)70 Italian COVID-19 patients and 30 Control Italian patients. ResNet-SegNet-based lung segmentation of CT scans on five different data combinations (DC1-DC5) using two cohorts have been designed. Five deep convolutional neural network models namely, DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B1, and EfficientNet-B6 models are utilized for ensemble. The focal loss function is used with a gamma value of 2. Five-fold cross-validation has been performed during model training and testing on unseen data. Statistical analysis and heatmaps are generated to validate the model. This model was also available for global use on Amazon Web Services as COVLIAS 3.0$_{XEDL}$. The proposed COVLIAS 3.0$_{XEDL}$ is superior to TL models.

Results: The XEDL showed an accuracy of 99.99%, AUC 1 (p<0.0001) for DC1, 98.23%, AUC 0.97 (p<0.0001) for DC5, 96.45%, AUC 0.92 (p<0.0001) for DC2, 88.20%, AUC 0.85 (p<0.0001) for DC3, and 87.87%, AUC 0.81 (p<0.0001) for DC4. The proposed XEDL accuracy was 8.59% superior to the mean TL accuracy.

Conclusions: Our hypothesis holds true where XEDL is superior to TL in a cloud-based explainable framework using heatmaps.

*Keywords***:** COVID-19, Control, ResNet-SegNet, feature fusion, ensemble deep learning, validation.

1. Introduction

COVID-19 has affected globally since 2020. Deep learning (DL) techniques have been used to classify COVID-19 diseased patients against normal patients based on segmented computed tomography (CT) scans [1-3]. Although transfer learning (TL) models are effective, their inability to combine them has forced the use of ensemble methods [4-7]. Ensemble deep learning (EDL) is a powerful technique for improving prediction accuracy by combining the results of multiple machine learning [8, 9] and deep learning models [6, 10-13]. Earlier EDL methods were focused on classification, but statistical analysis, heatmaps to detect COVID lesions and model validations were neglected [14-18].

We hypothesize that the fusion of features of TL models can improve the system performance. Previous studies have used various DL techniques to predict COVID-19 [11, 19, 20], this literature review helped us select the right TL models for the ensemble paradigm [21, 22]. We propose an EDL approach to classify COVID-19 and Control patients using five pre-trained models, namely, DenseNet169, DenseNet121, DenseNet201, EfficientNetB1, and EfficientNetB6. Further to feature fusion, the proposed EDL uses multicentre cohorts and presents novelty in terms of explainability in cloua d-based framework. We used Croatian (CroMED) and Novara, Italian (NovMED) cohorts for the validation of our hypothesis and experimentations [2, 23, 24].

2. Methodology

The proposed model uses five pre-trained models, namely DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B1, and EfficientNet-B6, to extract features from the dataset [25-27]. The extracted features are combined using ensemble learning to improve the prediction accuracy. The focal loss function was used for training the model, and the training validation loss was 0.1%. Statistical analysis, receiver operating characteristics (ROC), and explainable EDL heatmaps are used to validate the model [28-30].

2.1 Image Acquisition and Data Demographics

In the proposed study, two distinct cohorts were utilized. Physicians and radiologists are the co-authors of this article have already validated this dataset. The first cohort, also known as the experimental data

set, consists of 80 CroMED COVID-19 positive individuals, out of which 57 are male and the remaining 31 are female. Sample images are shown in Figure 1. We have used this data [2] in our previous publications. A detailed descriptofboutthe datasepresentedesent in Appendix A.

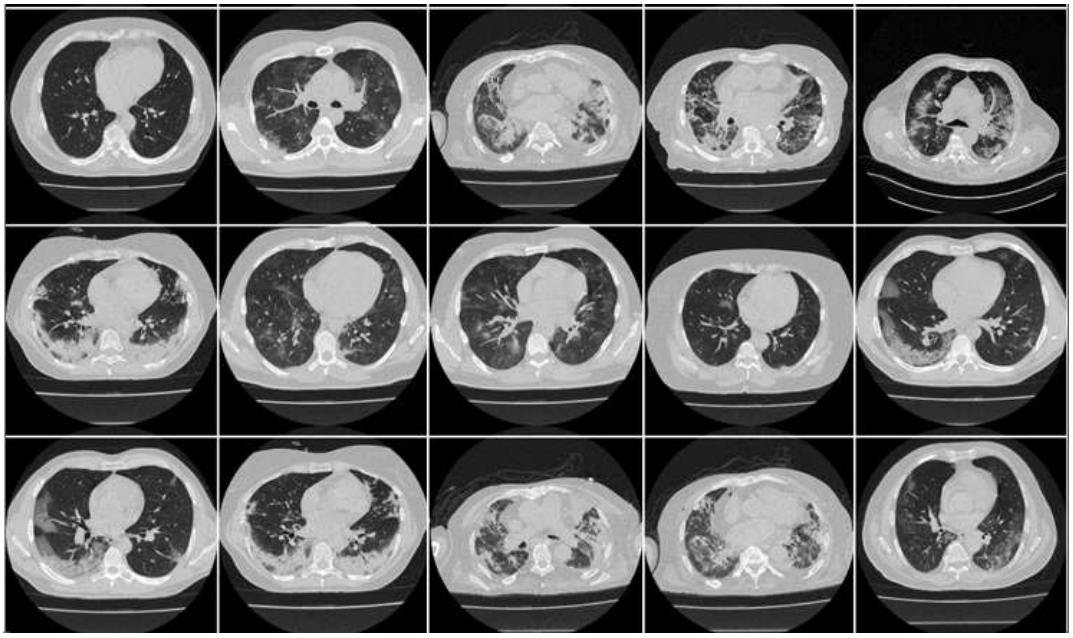

Figure 1. Raw "COVID-19 CT lung slices" taken from the CroMED dataset.

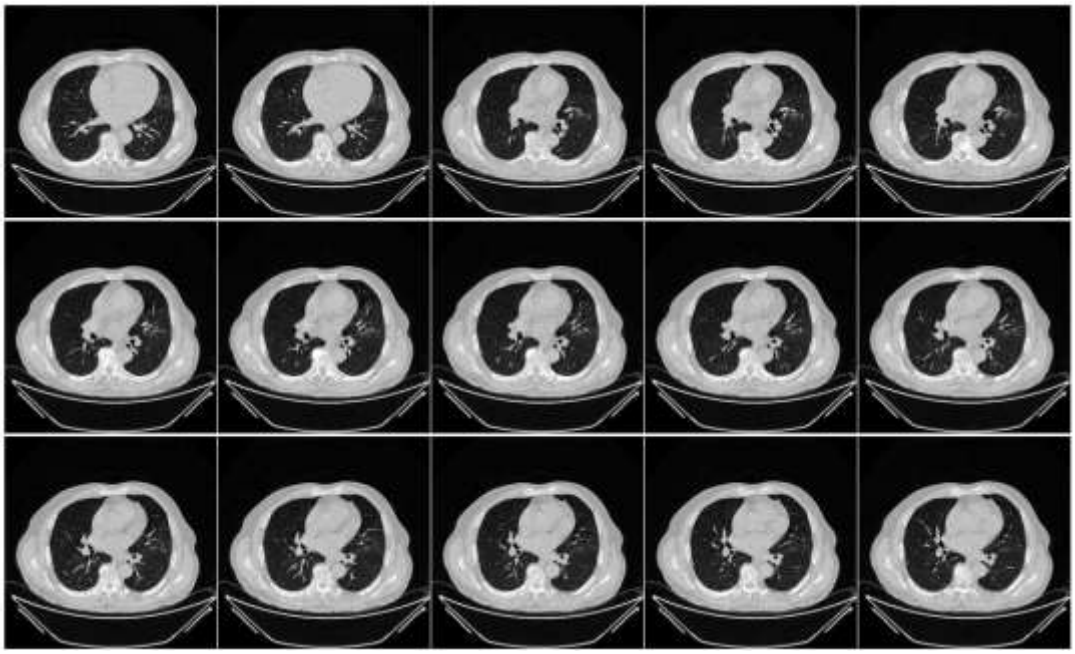

Figure 2. Raw "COVID-19 CT lung slices" taken from the NovMED dataset.

Figure 2 depicts 72 NovMED COVID-19 positive individuals, including 47 males and the remaining 29 females. The CroMED COVID-19 dataset consists of 5396 raw images with 512x512 dimensions,

while the NovMED COVID-19 dataset contains 5797 raw images with 768x768 dimensions and the Control (Italy) dataset has 1855 raw images with 768x768 dimensions . Figure 3 depicts the Control patient data slices. This data appears imimbalancedbut synthetic data is not advisable in tedical research. We have used focal loss to handle the data imbalance situation.

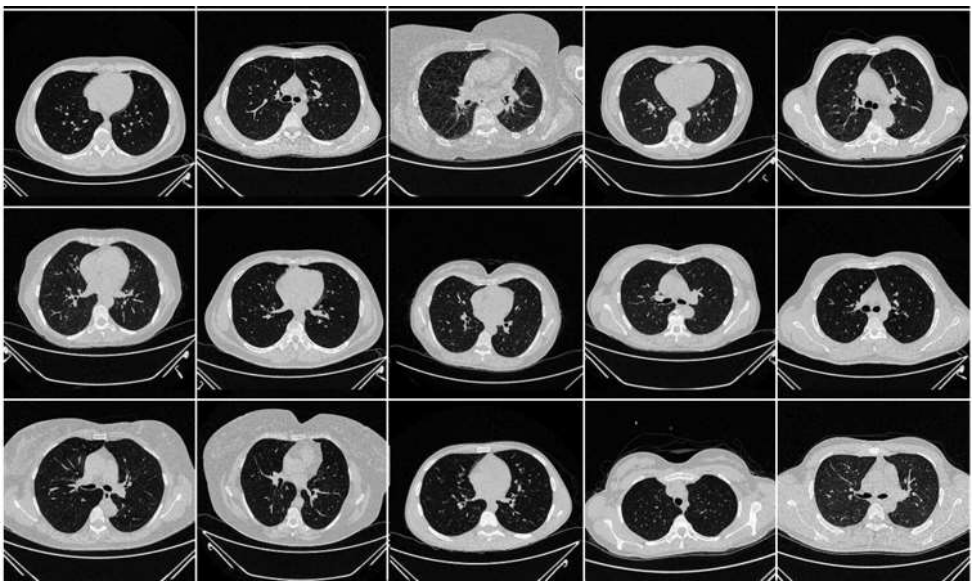

Figure 3. Raw "Control CT slices" taken from the NovMED dataset.

2.2 Global Architecture

The global architecture is illustrated in Figure 4. This consists of the cascade of four key blocks, namely CT volume acquisition along with quality control (block 1), HDL-based lung segmentation (block 2), TL-based feature extractor with explainable heatmaps (block 3), feature fusion-based EDL and classification (block 4), followed by performance evaluation. The HDL-based lung segmentation is superior compared to solo deep learning models [31]. For classification, we used five TL models, It is selected based on empiricalexperimentst over many TL models [32] , DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B1, and EfficientNet-B6, to visualize lesion sections in a grayscale CT image, which was later explainable using heatmaps. Hyperparameters are optimizer: Adam, learning rate (lr=0.001), Regularizer: L2 (0.01), Dropout: 0.2, Batch Size: 64, classification layer activation function: Sigmoid, intermediate layer activation function: ReLu, and Epoch:30. Extracted features from TL models were then concatenated for COVID-19 patients' prediction. Fully connectedlayersr and dropout layers were added for classification. The sigmoid activation function performed the binary

classification between COVID-19 versus Control. K-cross-validationohasve been used for training (80%) and validation (10%) then trained modes are used for testing (10%).

Figure 4. COVLIAS 3.0$_{XEDL}$: CT volume acquisition, ResNet-SegNet-based segmentation, TL-based feature extraction and explainability, and feature fusion-based EDL classification.

2.3 ResNet-SegNet Architecture for Lung Segmentation

Figure 5 shows the ResNet-SegNet architecture of CT lung segmentation consisting of convolution and maxpooling (yellow). Here, a new connection known as the skip connection (crossing lines) allowed gradients to pass through a predetermined number of layers, thus solving the vanishing gradient issue. Moreover, during the backpropagation step, the local gradient value stayed at one with the aid of one more addition to the network, namely an identity function (green) [33]. Here we have used Backbone-encoder: ResNet50, parameter: ~15 million, model size: 170 Mega Byte (MB).

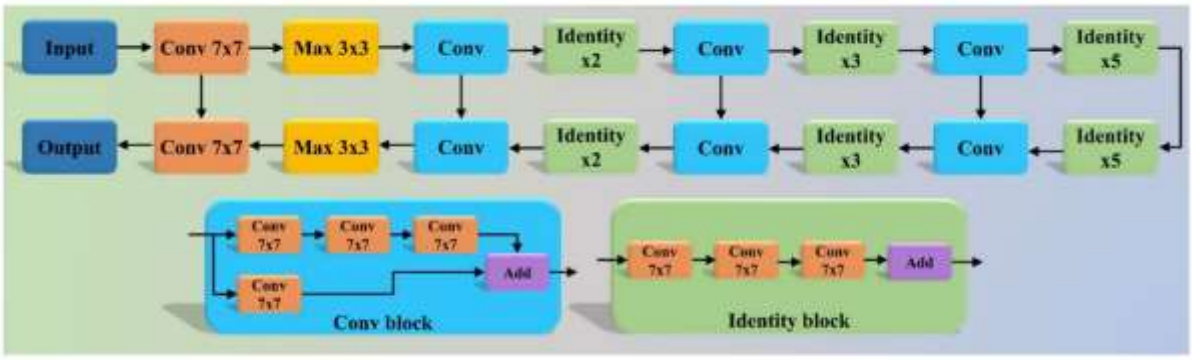

Figure 5. ResNet-SegNet architecture of CT lung segmentation [33].

2.4 Focal Loss Function

By introducing a hyperparameter gamma 'γ' (gamma = 2.0 has been decided by iterative method range from gamma 1.0 to 2.5 and its performance over model), referred s the focusing parameter, this loss function generalizes binary cross-entropy by penalizing tricky-to-classify examples far more severely than simple-to-classify examples [34, 35]. Focal loss (*FL*) is mathematically given by thuation 1.

$$FL\ (y, p) = -\alpha y(1-p)^{\gamma} \log(p) - (1-y)\, p^{\gamma} \log(1-p) \qquad (1)$$

Where, y Є [1, 0] is a binary class label, p Є [0,1] is the probability of COVID-19 classes, γ is the focusing parameter, and α is a hyperparameter (whose default value 1.0)

2.5 Performance Metric

To estimate the various performance evaluation metrics, we computed the true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Equation 2 to Equation 5 fundamental equations for computing F1-score, recall (Ʀ), recision (Ƥ) and mean accuracy (ɳ), respectively. Equation 6 shows the mean accuracy of the EDL ($\bar{\eta}_{EDL}$). quations 7 and 8, prespresentet of requested objects (ƶ) and a set of discovered items (y̓), one can the alculate Dice (ß) and Jaccard ($\mathcal{J}$). Additionally calculated for each model are the probability curve ROC and degree of separability AUC (the-area-under-the-curve).

$$Ƒ = 2\ x\ \left(\frac{Ƥ\ x\ Ʀ}{Ƥ+Ʀ}\right) \quad (2), \qquad Ʀ = \frac{TP}{TP+FN} \quad (3), \qquad Ʀ = \frac{TP}{TP+FN} \quad (4), \qquad ɳ = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

$$\bar{\eta}_{EDL} = \frac{\sum_{i=1}^{M} \eta_{EDL\ i}}{M} \quad (6), \qquad e(y̓, ƶ) = 2\frac{|y̓||ƶ|}{|y̓|+|ƶ|} \quad (7), \qquad \mathcal{J}(y̓, ƶ)$$

2.6 Experimental Protocol

Five Data Combinations

Using the multicentre cohorts from Italy and Croatia, we created five different sorts of data combination scenarios to ensure generalization using unseen data analysis.

- DC1: Training and Testing both using the combination of CroMED (COVID-19) + NovMED (Control).
- DC2: Training and Testing both using the data from Novara, Italy: NovMED (COVID-19) + NovMED (Control).
- DC3: Training using a combination of two cohorts: CroMED (COVID-19) + NovMED (Control) and Testing only from Novara, Italy: NovMED (COVID-19) + NovMED (Control).
- DC4: Training using Novara, Italy: NovMED (COVID-19) + NovMED (Control), and Testing using Croatia: CroMED (COVID-19) + NovMED (Control).

- DC5 Training and Testing on CroMED (COVID-19) + NovMED (COVID-19) + NovMED (Control), so-called, Mix data.

Experiment #1: CT Lung Segmentation Using Hybrid Deep Learning

In this experiment, lungs are segmented in CT scans using ResNet-SegNet-based HDL [1, 13]. K5-based cross-validation was used for training/testing combination and performance was evaluated with given ground truth lung borders, as traced by the radiologists.

Experiment #2: Feature Fusion-based Ensemble Deep Learning Models for Classification

We have extracted features using TL-based DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B1, and EfficientNet-B6 and concatenated their features. These fused features were then utilized for classification. This was then benchmarked against the voting-based EDL.

Experiment #3: Unseen Data Analysis for Classification

This experiment uses unseen data as part of the test data set ensuring the generalization concept [36, 37]. K5-based cross-validation was used for tthe raining/testing combination.

Experiment #4: System Reliability Test

The fourth and last experiment was performed to check the reliability of COVLIAS 3.0_{XEDL}. Two statistical analyses were conducted and heatmaps were visualized using Grad-CAM.

Hardware and Optimization: All models were executed using DC1 to DC5 on the NVIDIA GPU cluster at Idaho State University (ISU), Idaho, USA. We used the Tensorflow 2.0 libraries to create the program and MedCalc statistical software for performance evaluation. We have used optimizer: Adam, learning rate: 0.001, activation functions: ReLU, Sigmoid ,and selection of freezing layers vary among TL models to achieve best accuracy in unseen data analysis. Loss/Accuracy curves were generated.

3. Results and Performance Evaluation

3.1 Results

Results on Lung Segmentation using Hybrid Deep Learning: Figure 6 depicts the CroMED (COVID-19), NovMED (COVID-19), and NovMED (Control) HDL segmented dataset using ResNet-SegNet.

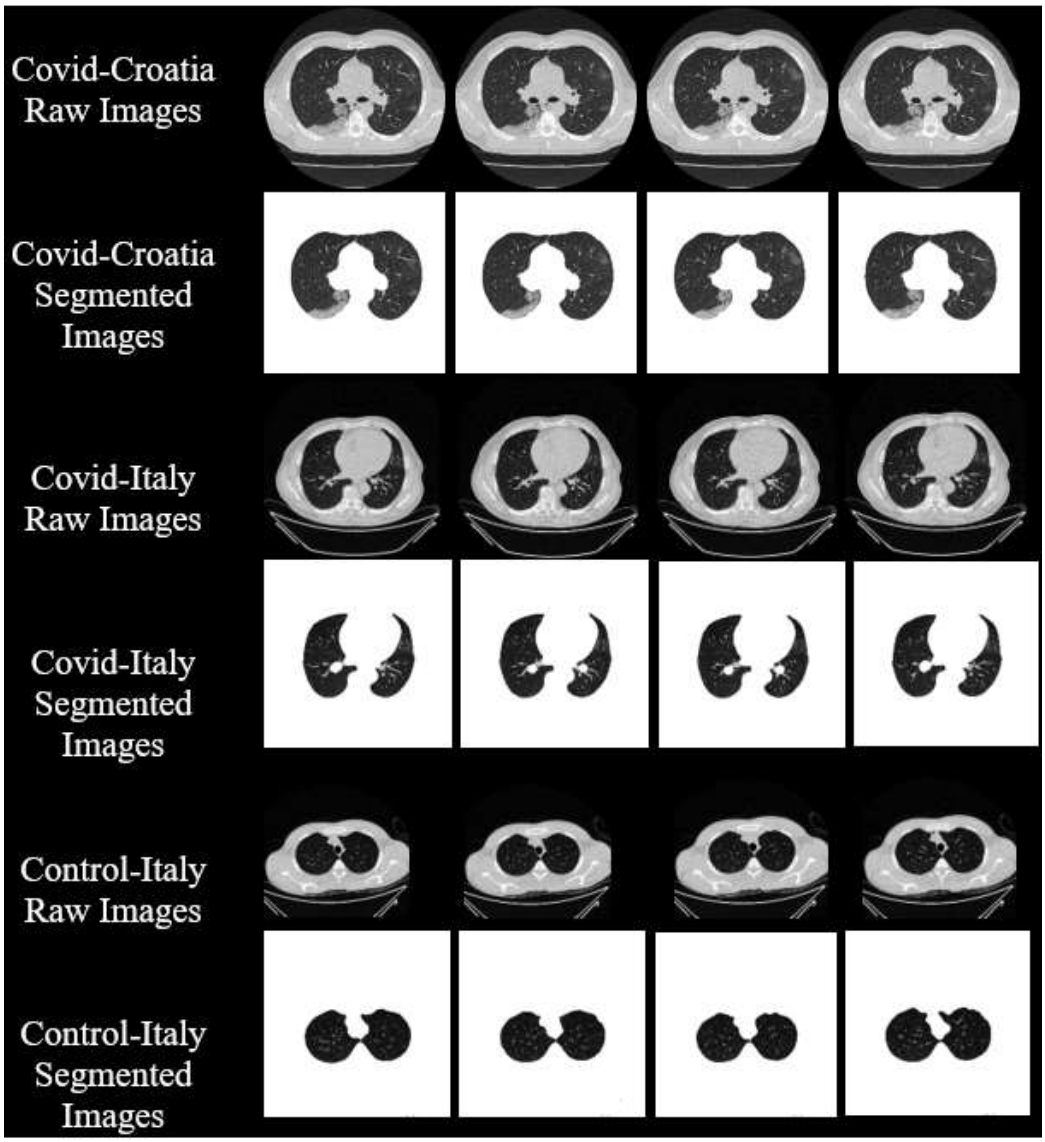

Figure 6. ResNet-SegNet-based segmentation. Row #1: Raw (Croatia), Row #2: Segmented (Croatia), Row #3: Raw (Italy), Row #4: Segmentation (Italy), Row #5: Raw (Control) Row #6: Segmented (Control).

Figure 7 shows the cumulative frequency of Jaccard (Right) and Dice (Left). Figure 8 (a) shows the regression curve of the GT area *vs*. the AI area, while Figure 8 (b) shows the Bland-Altman (BA) plots. The CC for regression is 0.99 while the BA plshowhoa ws low standard deviation.

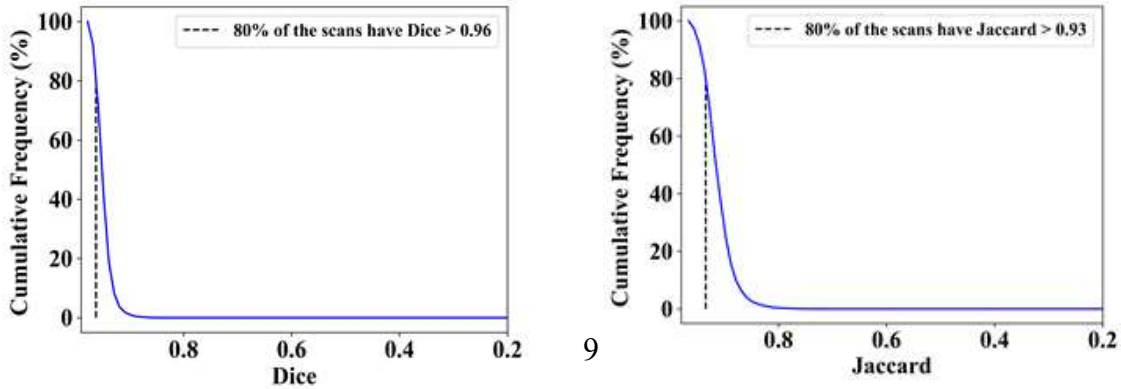

Figure 7. Cumulative frequency distribution of Dice and Jaccard indices.

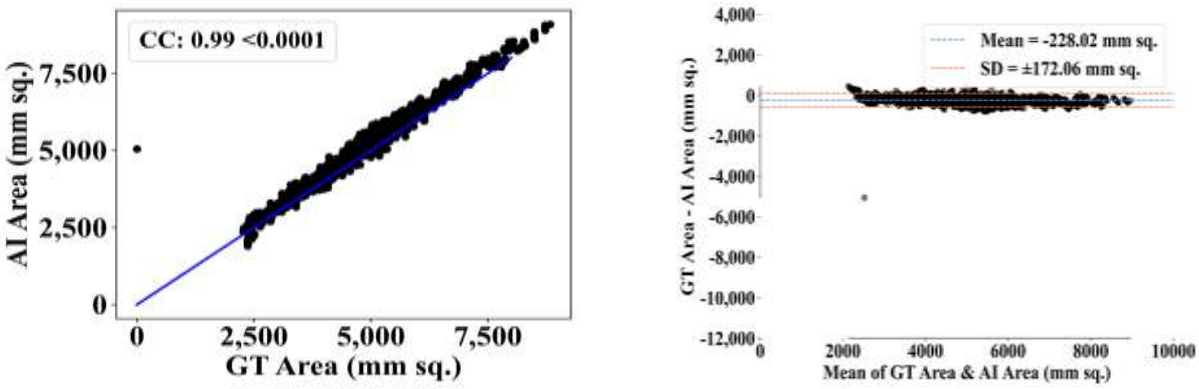

Figure 8. Left (a): AI Area vs. GT Area using CroMED dataset. Right (b): Bland-Altman plots.

Results on Feature Fusion-based Ensemble Deep Learning Models for Classification: The outcomes of experiments 1 are displayed in Table 1. DC1 hasane accuracy of 99.99% which is 3.67% better than DC2, 1.79% better than DC5, 13.59% better than DC3, and 13.79% better than DC4.

Results on Unseen Data Analysis for Classification: Unseen data analysis is shown in Table 1 where DC3 and DC4 had an accuracy of 88.20% and 87.87% respectively. It demonstrates that extra steps should be taken to reliably forecast COVID-19 patients.

We have also applied the voting ensemble method [10, 16, 38] using five TL models the and mean accuracy is 98.64%. The proposed EDL is 1.3% bettthe er than voting method.

Results on System Reliability Test: We have performed the fourth experiment to validate the system. Statistical T-test and Chi-square test have demonstrated very low *p-values* (p<0.0001) using MedCalc software. The heatmap has showna clear lesion section.

Table 1. EDL comparative results on five data combinations (DC1-DC5) using six attributes.

EDL Statistics					
Data Combination	**DC1**	**DC2**	**DC3**	**DC4**	**DC5**

Mean ACC (%)	99.99	96.45	88.20	87.87	98.23
COVID-19 Precision (%)	100	100	93	99	98
Control Precision (%)	100	86.7	75	68	100
COVID-19 Recall (%)	100	95.4	84	84	100
Control Recall (%)	100	99.4	97	98	85.9
COVID-19 F1-score (%)	100	97.6	91.4	91	99.4
Control F1-score (%)	100	92.4	84	80	92.6
AUC [0-1]	1	0.97	0.85	0.81	0.92
Sensitivity (%)	100	99.2	99.37	99.37	84.9
Specificity (%)	100	94.9	68.5	67.7	100
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

3.2 Performance Evaluation

Statistical Test

We have performed a T-test to compare the lesion region and a categorical Chi-square test among different groups of patients. Additionally, the *p*-value is computed, and it serves as a gauge of the strength of the evidence opposing the null hypothesis. The null hypothesis should be rejected if the observed frequencies are significantly different from the predicted frequencies, as shown by a low *p*-value <0.0001 as shown in Table 2.

Table 2. T-test and Chi-square test *p*-values.

EDL using DC	T-Test	Chi-square Test
DC1	$p<0.0001$	$p<0.0001$
DC2	$p<0.0001$	$p<0.0001$
DC3	$p<0.0001$	$p<0.0001$
DC4	$p<0.0001$	$p<0.0001$
DC5	$p<0.0001$	$p<0.0001$

Explainability

This paper includes explainability using the Grad-CAM technique. Grad-CAM helps in (i) debugging the working of the AI model, (ii) validating the predicted results, and (iii) visualizing and determining what the AI model considers as important features while making its prediction. It generates a heatmap lDopplerpler [39] of the important places that make the image belong to a particular class according to the trained EDL model. Figure 9 below depicts the heatmap produced by the trained model. We have already done extensive work on heatmap in our previous publications [2, 40].

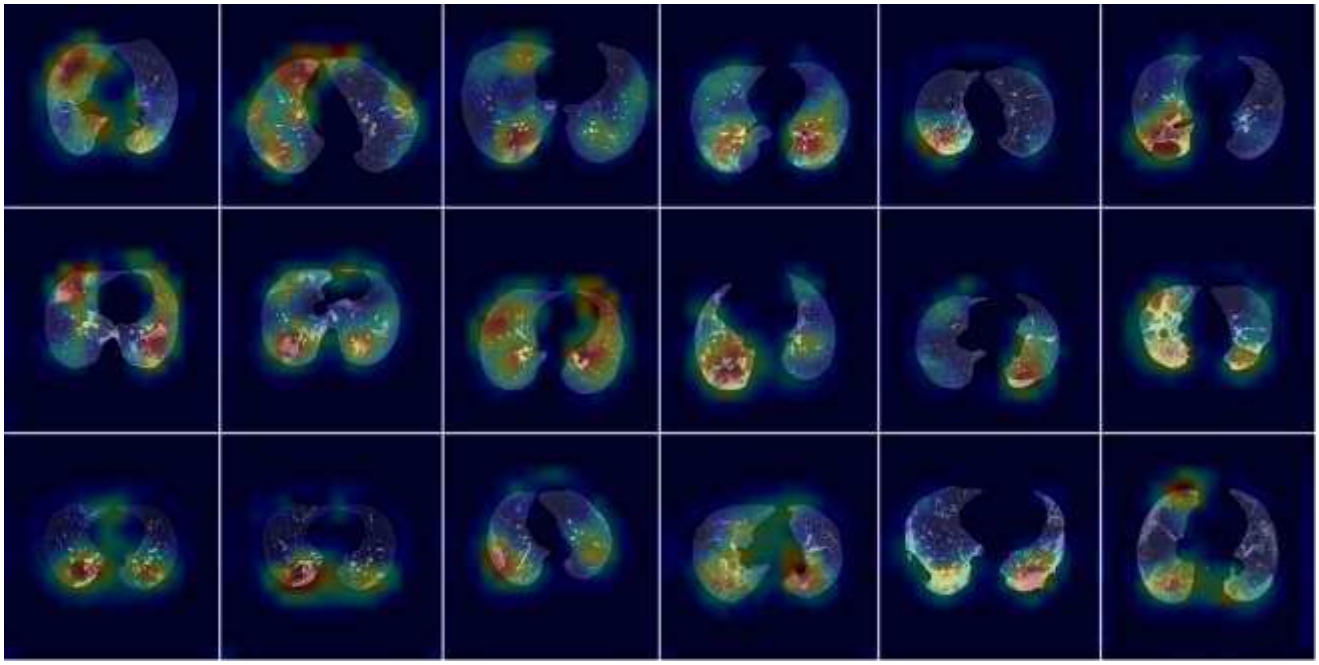

Figure 9. Grad-CAM-based proposed EDL heatmaps for COVLIAS 3.0_{XEDL}.

Receiver Operating Characteristics

Figure 10 shows the loss vs. accuracy graph using DC1, DC2, DC3, DC,4 and DC5. The ROC of the EDL model using five data combinationhasve been shown in Figure 11. It shows the AUC over DC1, DC2, DC3, DC,4 and DC5 are 1, 0.97, 0.85, 0.92, and 0.81. It presents that DC1 is the best model in seen analysis and DC3 is the best in unseen analysis.

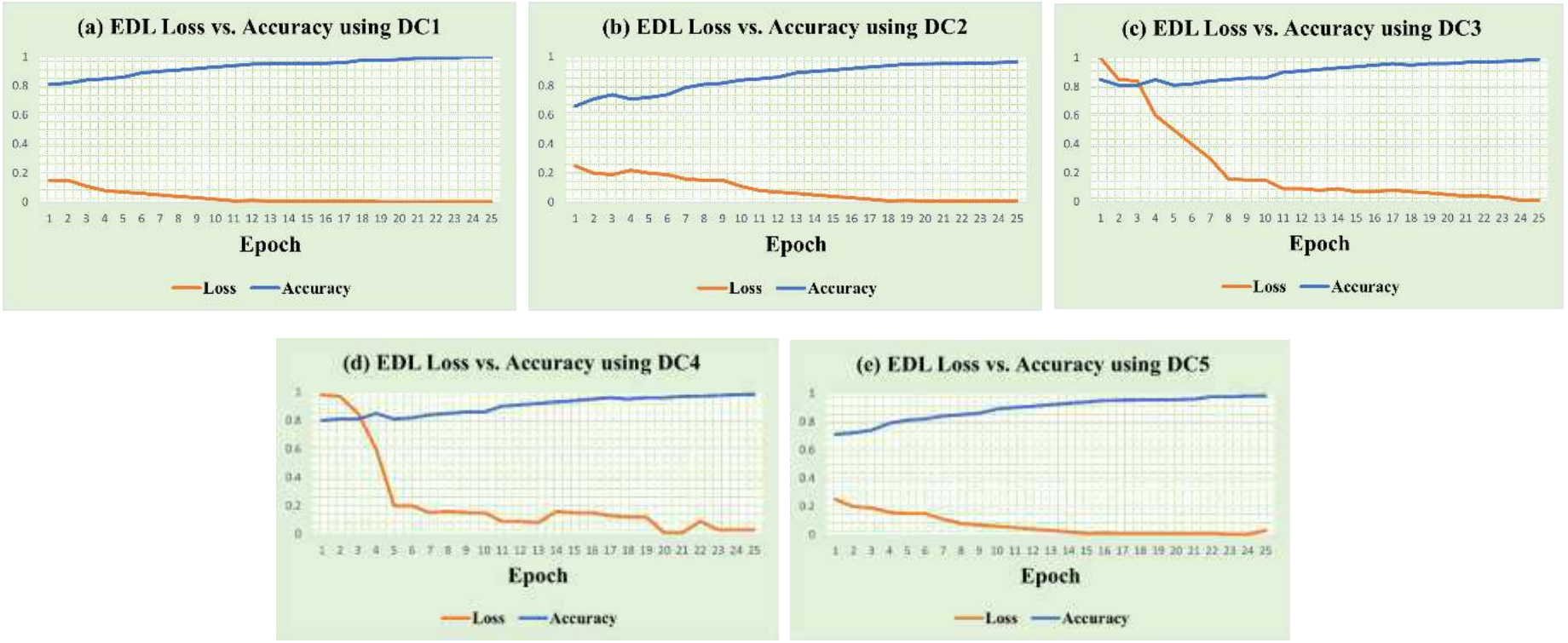

Figure 10. Loss vs. Accuracy of EDL model using DC1, DC2, DC3, DC4, and DC5.

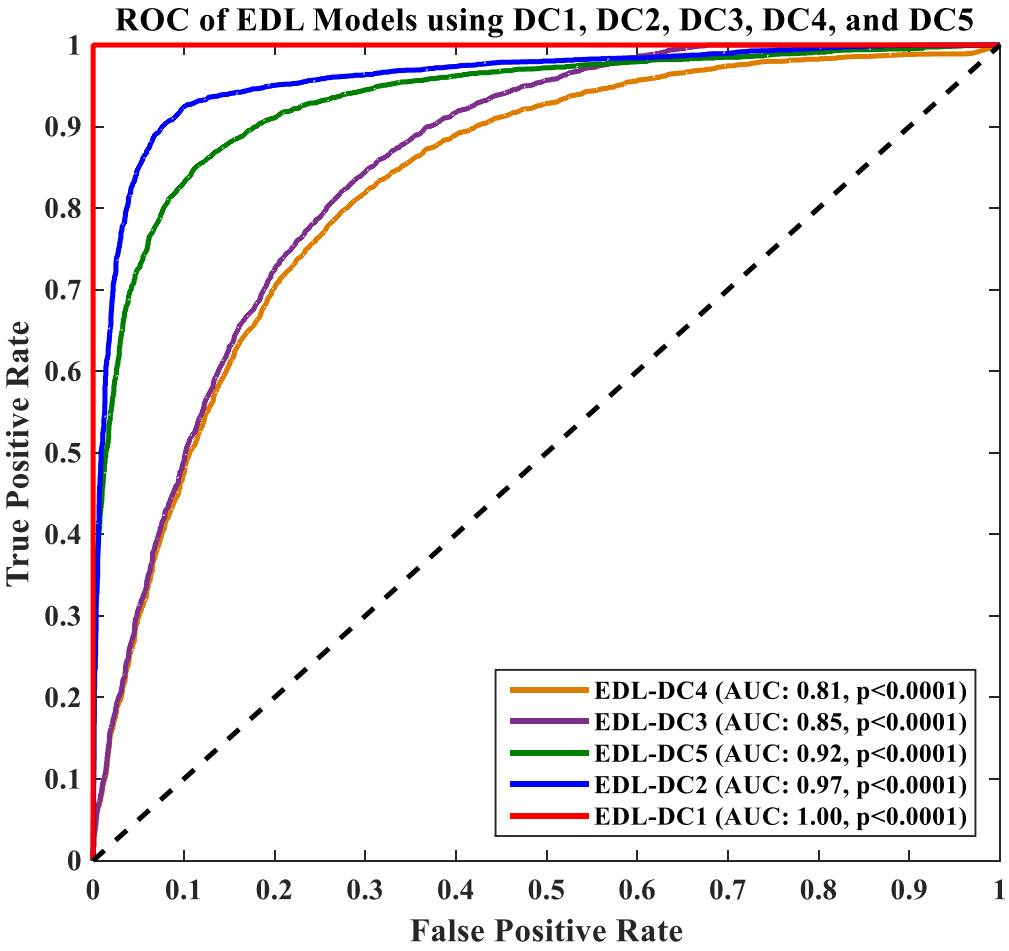

Figure 11. ROC of EDL model using DC1, DC2, DC3, DC4, and DC5.

4. Cloud-based COVLIAS 3.0$_{XEDL}$

This study also presents a cloud app for COVLIAS 3.0$_{XEDL}$. The app takes in CT lung images (Figure 12) as input and processes them using the trained EDL model. The results of the prediction, including the predicted class and probability scores, are displayed on the results page (Figure 13). Each uploaded image is given a unique ID, which can be used to view the prediction results at any time. This app is deployed using Amazon Web Services and utilizes a 6-core CPU and 16 GB of RAM. COVLIAS 3.0$_{XEDL}$ uses multiprocessing to process batch images in a parallel manner, resulting in a faster turnaround time. COVLIAS 3.0XEDL system is currently only available for internal use and accessible through restricted IP addresses. We faced some challenges at the time of deployment like data security, performance, and contingency planning, weovercamee these issuey.

Figure 12. Home page showing the COVLIAS 3.0$_{XEDL}$ Cloud App.

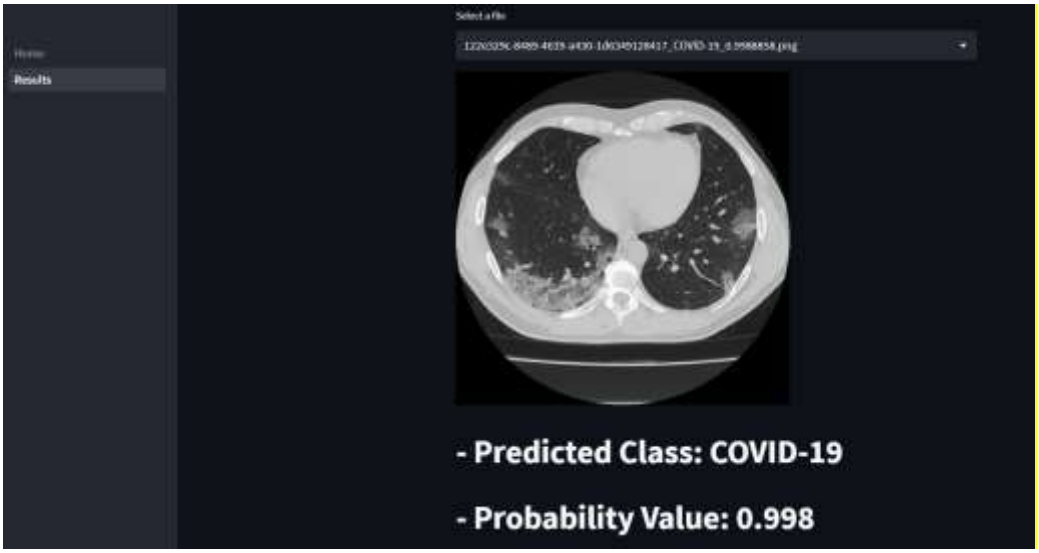

Figure 13. COVLIAS 3.0$_{XEDL}$ Cloud App showing a COVID-19 CT scan.

5. Discussion

The proposed ensemble learning model showed promising results in predicting COVID-19 and Control patients. The focal loss function was effective in training the model, and the low training validation loss indicated that the model was not overfitting. The statistical analysis and heatmap provided additional validation of the model.

Principal Findings

The proposed ensemble model outperformed the benchmark model in terms of accuracy, precision, recall, *p*-value, and F1 score. The statistical analysis and heatmap showed that the model was reliable and accurate. The system had the following novelties: (i) usage of HDL model, namely ResNet-SegNet for segmentation, (ii) Five TL models, their extracted features and feature fusion, (iii) the generalized

COVLIAS system using unseen data, and (iv) explainability, system stability/reliability, and cloud-based application. The ensemble model makes use of the variety among base models through feature fusion EDL. The use of EDL is essential for adding more features and improving the performance of the ensemble model.

Benchmarking

The proposed model was compared to a benchmark model using a five-fold cross-validation approach. The ensemble model outperformed the benchmark model in terms of mean accuracy, precision (Pre), recall (Re), F1-score (F1), *p*-value, clinical validation (Cli. Val.), and scientific validation (Sci. Val.). Our EDL model achieved an accuracy of 99.99%, and other performance matrix values were better than existing models, as presented in Table 3. Scientific validation has also been done in this paper. We have also compared the TL model to the proposed EDL in Table 4. The mean accuracy for five TL was 92.08% which is 8.59% lower tthe han proposed model.

Table 3. Proposed model comparison of accuracy (Accu), precision (Pre), recall (Re), F1-score (F1), AUC, *p*-value, clinical validation (Cli. Val.), and scientific validation (Sci. Val.) with existing EDL.

Year	**2021**	**2022**	**2021**	**2021**	**2021**	**2022**	**2022**	**2023**
Reference	Pathan *et al.* [41]	Shaik *et al.*[42]	Kundu *et al.* [43]	Zhou *et al.*[26]	Cruz *et al.* [38]	Lu *et al.*[44]	Huang *et al.*[45]	Proposed
DL Models for ensemble	Five Pre-trained model	Eight pre-trained model	Five Pre-trained model	Three Pre-trained model	Six Pre-trained model	DL model with a Loss function	Five Pre-trained model	Five Pre-trained model
Dataset	COVID-CT	COVID-CT	SARS-CoV-2	COVID-CT	COVID-CT	COVID-CT	COVID-CT	DC1
Data Size	746	746	2482	2933	746	746	746	7652
Accu (%)	97	93.33	98.93	99.05	90.7	94.3	98.84	99.99
Pre (%)	97	93.6	98.93	-	93.27	0.94	98.87	100
Re (%)	97	92.97	98.93	-	89.69	0.93	98.93	100
F1 (%)	97	93.21	98.93	98.59	94.05	0.94	98.92	100
AUC [0-1]	0.97	0.92	0.98		0.95	0.98	0.99	1
p-value	-	-	-	0	-	<0.0001	0	<0.0001
Cli. Val.	✗	✗	✗	✗	✗	✗	✗	✗
Sci. Val.	✗	✗	✗	✗	✗	✗	✗	✓

Table 4. TL comparison to the proposed EDL model.

Year	**2021**	**2021**	**2022**	**2022**	**2023**	**2023**	
Reference	Alshazly *et al*. [46]	Alshazly *et al*. [46]	Lunagaria *et al.* [47]	Masud *et al*.[48]	Xu *et al.* [49]	Proposed	
Models	DenseNet 201	DenseNet 169	Efficient NetB1	EfficientNe tB6	DenseNet 121	Proposed EDL	
Dataset	COVID-CT	COVID-CT	COVID X-Ray	COVID X-Ray	COVIDx-CT 2A	DC1*	
Data Size	746	746	4575	6432	3745	5797	1855
Accu (%)	92.9	91.2	97.6	79.3	99.44	99.99	
Pre (%)	91.3	88.1	96.4	69	99.89	100	
Re (%)	-	-	96.4	91	✗	100	
F1 (%)	92.5	90.8	96.4	78	✗	100	
AUC [0-1]	0.93	0.91	0.97	0.87	✗	1	
p-value	✗	✗	✗	✗	✗	<0.0001	

*CroMED (COVID), NovMED (Control).

Strengths, Weaknesses, and Extensions

HDL models are powerful solutions for lung segmentation. The suggested model's usage of TL models that have already been pre-trained on substantial datasets is one of its advantages. The model also takes advantage of ensemble learning, which increases the precision of predictions. The model's need for a sizable dataset for training and validation is one of its limitations that can be fulfilled with a large dataset. Characterization of COVID-19 lung tissues can incorporate similar patterns [50].A combinationn of feature fusion and votimethodshod can also be adopted using massdatasetsaset [51, 52].

6. Conclusion

The proposed model uses a cloud-based ensemble learning approach to predict COVID-19 and Control patients, which is slightly better than existing models. We have used a feature fusion-based ensemble learning approach to improve the performance of previous EDL models. Best values are achieved over DC1 with a mean accuracy, precision, recall, F1-score, and p-values of 99.99%, 100%, 100%, F1-score 100%, and a *p*-value<0.0001. COVLIAS 3.0_{XEDL} outperformed the benchmark EDL and TL models.

Further, the statistical analysis, heatmap, and ROC provided additional validation. The proposed model has the potential to be used as a reliable and accurate tool for COVID-19 diagnosis.

Author contributions: *Conceptualization*, J.S.S., S.A.,A.K.D., L.S., M.F.; *Data curation,* G.L.C., J.L., I.M.S., M.K.K., and K.V.; *Software,* A.K.D.,P.S., and S.A.; *Writing*—review & editing, J.S.S., A.K.D., M.K.K., K.V,G.R., D.G., R.S., and L.S.; All authors have read and agreed to the published version of the manuscript.

Funding: No funding was received for this project

Data availability statement: Due to its propriety nature, supporting data cannot be made available openly.

Conflict of interest All authors declare no competing interests.

References

[1] J.S. Suri, S. Agarwal, P. Elavarthi, R. Pathak, V. Ketireddy, M. Columbu, L. Saba, S.K. Gupta, G. Faa, I.M. Singh, Inter-variability study of COVLIAS 1.0: hybrid deep learning models for COVID-19 lung segmentation in computed tomography, Diagnostics, 11 (2021) 2025.

[2] J.S. Suri, S. Agarwal, G.L. Chabert, A. Carriero, A. Paschè, P.S. Danna, L. Saba, A. Mehmedović, G. Faa, I.M. Singh, COVLIAS 2.0-cXAI: cloud-based explainable deep learning system for COVID-19 lesion localization in computed tomography scans, Diagnostics, 12 (2022) 1482.

[3] J.S. Suri, A. Puvvula, M. Biswas, M. Majhail, L. Saba, G. Faa, I.M. Singh, R. Oberleitner, M. Turk, P.S. Chadha, COVID-19 pathways for brain and heart injury in comorbidity patients: A role of medical imaging and artificial intelligence-based COVID severity classification: A review, Computers in biology medicine, 124 (2020) 103960.

[4] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, P. Xie, Sample-efficient deep learning for COVID-19 diagnosis based on CT scans, medrxiv, (2020) 2020.2004. 2013.20063941.

[5] S. Serte, H. Demirel, medicine, Deep learning for diagnosis of COVID-19 using 3D CT scans, Computers in Biology Medicine, 132 (2021) 104306.

[6] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Chong, Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, IEEE transactions on computational biology bioinformatics18 (2021) 2775-2780.

[7] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19), European radiology, 31 (2021) 6096-6104.

[8] J.S. Suri, M. Bhagawati, S. Paul, A.D. Protogerou, P.P. Sfikakis, G.D. Kitas, N.N. Khanna, Z. Ruzsa, A.M. Sharma, S. Saxena, A powerful paradigm for cardiovascular risk stratification using multiclass, multi-label, and ensemble-based machine learning paradigms: a narrative review, Diagnostics, 12 (2022) 722.

[9] R. Cau, F. Pisu, M. Porcu, F. Cademartiri, R. Montisci, P. Bassareo, G. Muscogiuri, A. Amadu, S. Sironi, A. Esposito, Machine learning approach in diagnosing Takotsubo cardiomyopathy: The role of the combined evaluation of atrial and ventricular strain, and parametric mapping, International Journal of Cardiology, 373 (2023) 124-133.

[10] M. Mouhafid, M. Salah, C. Yue, K. Xia, Deep ensemble learning-based models for diagnosis of covid-19 from chest ct images, Healthcare, MDPI, 2022, pp. 166.

[11] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, F. Cabitza, Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study, Journal of medical systems, 44 (2020) 1-12.
[12] G.S. Tandel, A. Tiwari, O.G. Kakde, N. Gupta, L. Saba, J.S. Suri, Role of Ensemble Deep Learning for Brain Tumor Classification in Multiple Magnetic Resonance Imaging Sequence Data, Diagnostics, 13 (2023) 481.
[13] G.S. Tandel, A. Balestrieri, T. Jujaray, N.N. Khanna, L. Saba, J.S. Suri, Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm, Computers in Biology Medicine, 122 (2020) 103804.
[14] S. Tang, C. Wang, J. Nie, N. Kumar, Y. Zhang, Z. Xiong, A. Barnawi, EDL-COVID: ensemble deep learning for COVID-19 case detection from chest x-ray images, IEEE Transactions on Industrial Informatics, 17 (2021) 6539-6549.
[15] R. Arnaout, L. Curran, Y. Zhao, J.C. Levine, E. Chinn, A.J. Moon-Grady, An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease, Nature medicine, 27 (2021) 882-891.
[16] Liu, Manhua Zhang, Daoqiang Shen, Dinggang, Ensemble sparse classification of Alzheimer's disease, NeuroImage, 60 (2012) 1106-1116.
[17] H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma, W. Qian, An interpretable ensemble deep learning model for diabetic retinopathy disease classification, 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2019, pp. 2045-2048.
[18] J. Ruiz, M. Mahmud, M. Modasshir, M. Shamim Kaiser, f.t. Alzheimer's Disease Neuroimaging Initiative, 3D DenseNet ensemble in 4-way classification of Alzheimer's disease, Brain Informatics: 13th International Conference, BI 2020, Padua, Italy, September 19, 2020, Proceedings 13, Springer, 2020, pp. 85-96.
[19] Hai-tao Zhang, Jin-song Zhang, Hai-hua Nan, Yan-dong Zhao, Ying Fu, En-qing Xie, Yong-hong Liu, Wei Li, Wang-ping Zhang, Hong-jun, Automated detection and quantification of COVID-19 pneumonia: CT imaging analysis by a deep learning-based software, European journal of nuclear medicine molecular imaging 47 (2020) 2525-2532.
[20] W. Zhao, W. Jiang, X. Qiu, Deep learning for COVID-19 detection based on CT images, Scientific Reports, 11 (2021) 1-12.
[21] C. Shorten, T.M. Khoshgoftaar, B. Furht, Deep Learning applications for COVID-19, Journal of big Data, 8 (2021) 1-54.
[22] S. Aslani, J. Jacob, Utilisation of deep learning for COVID-19 diagnosis, Clinical Radiology, 78 (2023) 150-157.
[23] J.S. Suri, S. Agarwal, B. Jena, S. Saxena, A. El-Baz, V. Agarwal, M.K. Kalra, L. Saba, K. Viskovic, M. Fatemi, Five strategies for bias estimation in artificial intelligence-based hybrid deep learning for acute respiratory distress syndrome COVID-19 lung infected patients using AP (ai) Bias 2.0: a systematic review, IEEE Transactions on Instrumentation Measurement (2022).
[24] J.S. Suri, S. Agarwal, S.K. Gupta, A. Puvvula, M. Biswas, L. Saba, A. Bit, G.S. Tandel, M. Agarwal, A. Patrick, A narrative review on characterization of acute respiratory distress syndrome in COVID-19-infected lungs using artificial intelligence, Computers in Biology Medicine 130 (2021) 104210.
[25] X. Li, W. Tan, P. Liu, Q. Zhou, J. Yang, Classification of COVID-19 chest CT images based on ensemble deep learning, Journal of Healthcare Engineering, 2021 (2021).
[26] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, Y. Dong, The ensemble deep learning model for novel COVID-19 on CT images, Applied soft computing, 98 (2021) 106885.
[27] A.K. Das, S. Ghosh, S. Thunder, R. Dutta, S. Agarwal, A. Chakrabarti, Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network, Pattern Analysis Applications 24 (2021) 1111-1124.

[28] L.A. Waller, G.M. Jacquez, Disease models implicit in statistical tests of disease clustering, Epidemiology, (1995) 584-590.
[29] A.S. Datta, S. Biswas, Comparison of haplotype-based statistical tests for disease association with rare and common variants, Briefings in bioinformatics, 17 (2016) 657-671.
[30] Y. Roggo, L. Duponchel, C. Ruckebusch, J.P. Huvenne, Statistical tests for comparison of quantitative and qualitative models developed with near infrared spectral data, Journal of molecular structure, 654 (2003) 253-262.
[31] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, P. Xie, COVID-CT-dataset: a CT scan dataset about COVID-19, arXiv preprint arXiv:.13865, (2020).
[32] A.K. Dubey, G.L. Chabert, A. Carriero, A. Pasche, P.S. Danna, S. Agarwal, L. Mohanty, Nillmani, N. Sharma, S. Yadav, Ensemble Deep Learning Derived from Transfer Learning for Classification of COVID-19 Patients on Hybrid Deep-Learning-Based Lung Segmentation: A Data Augmentation and Balancing Framework, Diagnostics, 13 (2023) 1954.
[33] J.S. Suri, S. Agarwal, A. Carriero, A. Paschè, P.S. Danna, M. Columbu, L. Saba, K. Viskovic, A. Mehmedović, S. Agarwal, COVLIAS 1.0 vs. MedSeg: artificial intelligence-based comparative study for automated COVID-19 computed tomography lung segmentation in Italian and Croatian Cohorts, Diagnostics, 11 (2021) 2367.
[34] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988.
[35] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, J. Yang, Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, Advances in Neural Information Processing Systems, 33 (2020) 21002-21012.
[36] P.K. Jain, N. Sharma, M.K. Kalra, A. Johri, L. Saba, J.S. Suri, Far wall plaque segmentation and area measurement in common and internal carotid artery ultrasound using U-series architectures: An unseen Artificial Intelligence paradigm for stroke risk assessment, Computers in Biology Medicine, 149 (2022) 106017.
[37] M. Agarwal, S. Agarwal, L. Saba, G.L. Chabert, S. Gupta, A. Carriero, A. Pasche, P. Danna, A. Mehmedovic, G. Faa, Eight pruning deep learning models for low storage and high-speed COVID-19 computed tomography lung segmentation and heatmap-based lesion localization: A multicenter study using COVLIAS 2.0, Computers in biology medicine, 146 (2022) 105571.
[38] J.F.H. Santa Cruz, An ensemble approach for multi-stage transfer learning models for COVID-19 detection from chest CT scans, Intelligence-Based Medicine, 5 (2021) 100027.
[39] S.-F. Huang, R.-F. Chang, W.K. Moon, Y.-H. Lee, D.-R. Chen, J.S. Suri, Analysis of tumor vascularity using three-dimensional power Doppler ultrasound images, IEEE transactions on medical imaging, 27 (2008) 320-330.
[40] J.S. Suri, S. Agarwal, G.L. Chabert, A. Carriero, A. Paschè, P.S. Danna, L. Saba, A. Mehmedović, G. Faa, I.M. Singh, COVLIAS 1.0 Lesion vs. MedSeg: An Artificial Intelligence Framework for Automated Lesion Segmentation in COVID-19 Lung Computed Tomography Scans, Diagnostics, 12 (2022) 1283.
[41] S. Pathan, P. Siddalingaswamy, P. Kumar, M.P. MM, T. Ali, U.R. Acharya, Novel ensemble of optimized CNN and dynamic selection techniques for accurate Covid-19 screening using chest CT images, Computers in Biology Medicine 137 (2021) 104835.
[42] N.S. Shaik, T.K. Cherukuri, Transfer learning based novel ensemble classifier for COVID-19 detection from chest CT-scans, Computers in Biology Medicine 141 (2022) 105127.
[43] R. Kundu, P.K. Singh, S. Mirjalili, R. Sarkar, COVID-19 detection from lung CT-Scans using a fuzzy integral-based CNN ensemble, Computers in Biology Medicine 138 (2021) 104895.
[44] S.S. Skandha, A. Nicolaides, S.K. Gupta, V.K. Koppula, L. Saba, A.M. Johri, M.S. Kalra, J.S. Suri, A hybrid deep learning paradigm for carotid plaque tissue characterization and its validation in multicenter cohorts using a supercomputer framework, Computers in biology medicine, 141 (2022) 105131.

[45] M.-L. Huang, Y.-C. Liao, Stacking Ensemble and ECA-EfficientNetV2 Convolutional Neural Networks on Classification of Multiple Chest Diseases Including COVID-19, Academic Radiology, (2022).
[46] H. Alshazly, C. Linse, E. Barth, T. Martinetz, Explainable COVID-19 detection using chest CT scans and deep learning, Sensors, 21 (2021) 455.
[47] M. Lunagaria, V. Katkar, K. Vaghela, Covid-19 and Pneumonia Infection Detection from Chest X-Ray Images using U-Net, EfficientNetB1, XGBoost and Recursive Feature Elimination, International Journal of Advanced Computer Science Applications, 13 (2022).
[48] M. Masud, A light-weight convolutional Neural Network Architecture for classification of COVID-19 chest X-Ray images, Multimedia systems, 28 (2022) 1165-1174.
[49] Y. Xu, H.-K. Lam, G. Jia, J. Jiang, J. Liao, X. Bao, Improving COVID-19 CT classification of CNNs by learning parameter-efficient representation, Computers in Biology Medicine, 152 (2023) 106417.
[50] F. Molinari, A. Mantovani, M. Deandrea, P. Limone, R. Garberoglio, J.S. Suri, Characterization of single thyroid nodules by contrast-enhanced 3-D ultrasound, Ultrasound in medicine biology, 36 (2010) 1616-1625.
[51] A. El-Baz, J.S. Suri, Big data in multimodal medical imaging, CRC Press2019.
[52] J.S. Suri, Two-dimensional fast magnetic resonance brain segmentation, IEEE Engineering in Medicine Biology magazine, 20 (2001) 84-95.

Appendix A

CT Dataset imaging techniques

"Italian cohort: All chest CT scans were performed in a supine position with a single full inspiratory breath-hold, utilizing a 128-slice multidetector-row "Philips Ingenuity Core" CT scanner from Philips Healthcare (Netherlands). There was no intravenous or oral injection of contrast media. A soft tissue kernel with 512 × 512 matrix (mediastinal window) and a lung kernel with 768 × 768 matrix (lung window) were utilized to rebuild one-mm thick pictures. The CT tests were carried out with a 120 kV, 226 mAs/slice (using Philips' automatic tube current modulation—Z-DOM), 1.08 spiral pitch factor, 0.5-s gantry rotation time, and 64 × 0.625 detector configuration. The CT data of 72 COVID-positive individuals were used in the proposed investigation. CT volumes of patients were selected based on two criteria (i) the image quality should be reasonable and should have no artifacts or blurriness due to body movement, and (ii) there is no metallic object present in the scan area. Each patient consisted of approximately 200 slices from which the radiologist [LS] selected 65–70 slices from the visible lung region, yielding a total of 5000 images.

Croatian Cohort: A CROATIAN patient of seven COVID-19 positive patients (500 images) was used to validate the AI system (COVLIAS). All chest multidetector CT images (MDCT) were performed in a supine position with a single full inspiratory breath-hold utilizing FCT Speedia HD (Fujifilm Corporation, Tokyo, Japan, 2017) 64-detector MDCT scanner to acquire images of the thorax in craniocaudal direction. Images were acquired with a standard algorithm and viewed with Hitachi, Ltd. Whole Body X-ray CT System Supria Software (System Software Version: V2.25, Copyright Hitachi, Ltd. 2017). There was no contrast media available for intravenous or oral administration. The used scanned parameters were: volume scan, large focus, tube voltage 120 kV, tube current 350 mA with automatic tube current modulation mode (IntelliEC mode), and rotation speed 0.75 s. Parameters used for reconstruction were: field of view (FOV) 350 mm, slice thickness 5 mm (0.625 × 64), table pitch 1.3281, picture filter 32 with multi recon option: picture filter 22 (lung standard) with Intelli IP Lv.2 iterative algorithm (WW1600/WL600), slice thickness 1.25 mm, recon index 1 mm and picture filter 31 (mediastinal) with Lv.3 Intelli IP iterative algorithm (WW450/WL45), slice thickness 1.25 mm, recon index 1 mm. CT volumes of patients were selected based on two criteria (i) the image quality should be reasonable and should have no artifacts or blurriness due to body movement and (ii) there is no metallic object present in the scan area."