Hybrid approach for COVID-19 detection from chest radiography

Esraa Fady Dawood Department of Computer Science Faculty of Computers and Information Menoufia University, Egypt esraa_fady@yahoo.com Nader Mahmoud Department of Computer Science Faculty of Computers and Information Menoufia University, Egypt nader.mahmoud@ci.menofia.edu.eg Ashraf Elsisi Department of Computer Science Faculty of Computers and Information Menoufia University, Egypt ashraf.elsisi@ci.menofia.edu.eg

Abstract-Automatic and rapid screening of COVID-19 from the chest X-ray and Computerized Tomography (CT) images has become an urgent need in this pandemic situation of SARS-CoV-2 worldwide. However, accurate and reliable screening of patients is a massive challenge due to the discrepancy between COVID-19 and other viral pneumonia in both X-ray and CT images. Several models were introduced, but always there was a glitch that might be due to the use of a single classifier, and this reduces their accuracy. In this paper, we study the use of multi-classifiers and show their effect on different models working on X-ray and CT images. We perform a comparison study to show the high impact of ensemble stacking approach on top performer CNN models that recorded the highest detection accuracy in image detection and classification: COVID-Net, VGG16, ResNet, Bayesian, DenseNet, and DarkNet. We presented multi-classifiers instead of a single classifier stacked in an ensemble stacking approach for the diagnosis of the COVID19 from the Chest CT and Xray images. We provide a quantitative evaluation of the proposed ensemble stacking approach on two types of datasets: X-ray images and CT images datasets, with percentages reaching 99%.

Keywords— COVID-19, stacked algorithm, ensemble technique, deep learning, chest X-ray images, Computerized Tomography (CT) images.

I. INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic has put the lives and health of a large number of people at risk. It has wreaked havoc on the public life of the entire world's population. As the number of cases rises, all nations are experiencing a lack of resources to identify COVID-19 [1]. This necessitates the use of a COVID-19 detection technique that is readily available, inexpensive, and automated.

SARS (Severe Acute Respiratory Syndrome) can be triggered by coronavirus infection, which can have significant health consequences. The diagnosis of COVID-19 cases must be confirmed by molecular assays, such as the reverse transcription polymerase chain reaction (RT-PCR) pathological test using throat swab samples, according to the World Health Organization (WHO). However, radiography imaging sensitivity outperforms the usual PCR technique [3]. The detection technique based on chest radiography images offers a number of benefits over traditional blood [4] and PCR testing. In comparison to other procedures, these are quick and inexpensive. In addition, many instances may be tested at the same time.

Furthermore, radiography-based detection methods are a viable alternative for COVID-19 testing kit shortages due to

the widespread availability of radiology imaging equipment in hospitals. Machine and deep learning have become foundational fields in artificial intelligence in recent years [5]. For COVID-19 detection and classification on an automated basis, deep learning approaches is being extensively researched for COVID-19 detection and classification on automated manner [6]. As a result, deep learning has become an important component of automated clinical decision-making [7] [8].

In this paper, we proposed an ensample stacking approach based on multi classifiers rather than the typical single classifier theme in current state-of-the-art models. We quantitatively demonstrate the substantial effect of the proposed ensemble stacking approach on various models accuracy with chest X-ray and CT images.

This paper's structure is organized as follows: Section 2 presents the related work. Section 3 describes the proposed stacked CNN model for the classification of COVID-19 X-ray and CT images. Section 4 describes the details of the experimental results, performance comparison. Finally, the conclusion is drawn in Section 5.

II. RELATED WORKS

The Convolution Neural Network is the core idea of deep learning techniques in computer vision, and it has resulted in excellent results in a variety of pattern recognition tasks, including image classification [9], object localisation, segmentation, and detection. It has also demonstrated its superiority in medical image analysis for image classification and segmentation issues [10], particularly in lung-related illnesses including lung nodule detection, pneumonia detection, and pulmonary TB detection [11]. Many researchers have recently presented various approaches for detecting COVID-19 using radiological imaging. Various CNN models were examined in terms of 2D and 3D CNN models for the categorization of chest radiography images into COVID-19, viral pneumonia, or no-infection [12].

Sethy et al. [13] extracted the deep features from the chest radiography images using the fully connected layer of the pre-trained models. Then, features matching has been done using SVM classifier. The authors utilized 13 different pre-trained net and the highest classification accuracy of 98.66 % was obtained using ResNet50. Kumar et al. [14] utilized the DL ResNet-152 for prediction of COVID 19 patients on CRIs. Synthetic Minority Oversampling Technique (SMOT) is used with DL to balance the imbalanced data Points. The extracted learned features are fed to Random Forest (RF) and XGBoost (XGB) classifiers

for 3-class classification. The proposed model provided 97.7 % accuracy for XGB and 97.3 for RF classifiers.

Tree-class COVID classification method of CRIs was proposed in [15]. The method based on DL and 9-different CNN architectures. The best accuracy was 95 % for two models. Wang et al. [26] introduce model called COVID-Net to detect COVID cases from chest radiography images. . This model was designed for (normal vs pneumonia vs COVID) classification with overall accuracy of 94.4 %.

Brunese et al. [27] introduced a COVID-19 detection approach based on transfer learning of VGG-16 model. This model used to differentiate between healthy and disease Xray images with accuracy of 92 %. Then, it used for COVID-19 detection from disease X-ray chest dataset with accuracy of 94 %. The used DenseNet provided the same accuracy of 80 % for binary classification. Narin et al. [31] used 5-fold cross validation pre-trained models for COVID-19 detection from chest radiography images. The highest accuracy of 98 % is obtained using ResNet50 model. Panwar et al. [28] presented DL based method nCOVnet for fast COVID19 detection from x-ray chest images. This method used the VGG net and transfer learning using five-different layers that achieved 97.2 % detection accuracy.

Similarly, our previous works shows we tested six pretrained CNN models, including COVID-Net, VGG16, ResNet, Bayesian, DenseNet, and DarkNet, in order to demonstrate the applicability of deep learning in COVID-19 detection using X-ray and CT images [16]. The experimental results showed that both DarkNet and COVID-Net yielded an accuracy of 99% with minor variation in the values of PRECISION, RECALL, and F1 on the CT datase. For X-ray dataset, DarkNet model yield highest detection accuracy of 98% with PRECISION 0.98, RECALL 0.9, and F1 0.94.

However, other models yielded low accuracy. Thus, the work in this paper is an extension to our work [16]. We mainly focus on ensemble stacking classifiers and their effect on different models accuracy.

Stacked Generalization, or simply "Stacking," is a machine learning ensemble algorithm. On the same dataset, it entails integrating predictions from several machine learning models [20]. The algorithm takes the outputs of basic models as input and tries to figure out how to combine the input predictions in the best way possible to get a superior output prediction. The fact that each classifier contributes the same amount to the ensemble prediction, regardless of how well it performs, is a drawback of this technique [21]. A weighted average ensemble is a variant of this method that weights each ensemble member's contribution by the trust or expected performance. This allows the best classifiers to contribute more, while the worst-performing models contribute less [22].

Stacked generalization works by deducing the generalizer's biases in relation to a given learning set. This deduction works by generalising in a second space whose inputs are the original generalizers' predictions when they were taught a portion of the learning set and then asked to guess the remainder, and whose output is the right prediction.

As seen in fig. 1. Level-0 Models (Base Models): Models that fit the training data and provide predictions. Level-1 Model (Meta Model): Model that learns how to integrate the predictions of the basic models in the most effective way.

The meta-model is trained using out-of-sample data predictions generated by base models. That is, non-training data is fed into the base models, predictions are produced, and these predictions, along with the expected outputs, form the input and output pairs of the training dataset used to fit the meta-model [23].



Fig. 1. General ensemble stacking approach

III. METHODOLOGY

In this work, we present a deep learning-based approach for detecting the three COVID-19 classes from chest radiography images. We used deep learning approaches to collect the feature extraction, then we replacing the default single classifier with the multi-classifier (Ensemble Stacking approach). CNN learns a range of relevant feature representations automatically and merges feature extraction and classification stages into a single pipeline that can be trained end-to-end without the need for manual design or human involvement.

As illustrated in Fig. 2, the proposed COVID-19 detection stacking technique incorporates two models. The base model contains five classifiers named: SGD, SVM, Naive Bayes, Random Forest, and KNN. We made them five so that there would always be an identification of the majority, while if we used them in even numbers, the results could be divided equally between two categories. The second model known as the Meta model is the Logistic Regression.



Convolutional layers in convolutional neural network apply learnt filters to input images in a systematic way to generate feature maps that summarise the presence of certain characteristics in the input. After the convolutional layer, a new layer called a pooling layer is introduced. Specifically, when a convolutional layer's feature maps have been subjected to nonlinearity. The pooling layer works on each feature map independently to generate a new set of pooled feature maps with the same amount of features. Calculate the average value for each patch and use it to construct a feature map using average pooling.

The output of the CNN feature extraction phase pass through the base model which contains five classifiers working in parallel to get the most accurate result. The output of the base model is the input of the meta model, which works through logistic regression. Our model output is one of the three class cases which are COVID-19, pneumonia, or normal.

IV. EXPERIMENTAL SETUP

In this section we present: dataset information, evaluation metrics, evaluation of proposed method, and performance comparison. All experiments in this section were carried out on a PC with 8 GB of RAM and a gtx 1080ti GPU with 11 GB of RAM.

A. Data Set

We have considered publicly available datasets that had images of COVID-19 infection that were either positive or suspicious. The usage of public datasets aids in comparing the suggested ensemble stacking technique to state-of-the-art models in terms of accuracy gain. COVIDx [24] is the major dataset used to train and evaluate X-ray image models. It contains 16,756 chest radiography images for 13,645 patients. TABLE I displays the dataset distribution and Fig. 3 exhibits X-ray image samples from the COVIDx dataset with different cases.



Fig. 3. Samples of COVIDx dataset [24], (a) COVID-19, (b) pneumonia, and (c) normal case

TABLE I. COVIDx dataset distribution	
--------------------------------------	--

Туре	Normal	Pneumonia	COVID-19	Total
Train	5325	2400	2328	10053
Test	1088	1088	1176	3352
Validation	1088	1088	1176	3352
				16756

For CT images we have used China association of Chest CT Image Investigation (CCCCII)[25] dataset. It consists of 104,009 CT images for 1,489 cases. Its images show chest CT volumes from three different forms of infections: COVID-19, pneumonia, and uninfected. Figure 4 shows CT image examples with various illnesses from the COVIDx-CT dataset. The dataset distribution is shown in TABLE II. Slices identified as harbouring lung abnormalities were highlighted in the COVID-19 and pneumonia scan volumes, as seen in the figures.

We divided these two datasets with 60% for training, 20% for validation, and 20% for testing, which resulted in the best accurate performance of all the models tested. The dividing portions (60%, 20%, and 20%) are constructed at random in each run. The average of those 60 times is then taken into account in the stated findings. The CT and X-ray datasets were chosen based on their size and universality. It is the clearest and most comparable to genuine photos without any alterations or additions since it is the biggest data collection accessible thus far.



Fig. 4. Samples of COVID CT dataset [25], (a) COVID-19, (b) pneumonia, and (c) normal case

TABLE II. COVID CT dataset distribution

Туре	Normal	Pneumonia	COVID-19	Total
Train	28202	12635	21568	62405
Test	8658	4893	7251	20802
Validation	Validation 8658 4893		7251	20802
				104009

B. Evaluation Metrics

To measure the network's performance, we use the following most popular metrics: Precision, Recall, F1- Score, and Accuracy:

• Precision: The proportion of COVID19 positive patients who were properly predicted to the total number of positive predictions (i.e. True positives and false positives). This measure allows an algorithm to calculate the rate of false positives. The lower the false positives, the higher the precision.

Precision = *True positives* (*TP*)/ (*True positives* (*TP*)+ *false positives* (*FP*))

• Recall: The algorithm's sensitivity is also known as recall. It's the proportion of accurately anticipated positive outcomes (also known as true positives) to

actual class observations (i.e. True positives and False negatives).

Recall = *True positives (TP)/ (True positives (TP)+ False Negative (FN))*

• F1 Score: calculates false positives and false negatives by averaging the previously listed metrics. In instances where class distribution is unequal, the F1 score is beneficial.

F1 = 2 * (*Precision* * *Recall*) / (*Precision* + *Recall*)

• Accuracy: This is the most often used and straightforward categorization metric. It's calculated as the number of right predictions divided by the number of samples. Although high accuracy is an excellent measure, in some cases where the class distribution is not symmetric, it may not be the best.

Accuracy = (True positives (TP) + True negatives (TN))/(Total samples)

C. Accuracy Evaluations

TABLE III and TABLE IV shows performance evaluations validated on CT and X-ray datasets. As a result, all models were trained and assessed using the CT dataset, as shown in TABLE III. Similarly, as shown in TABLE IV, all models were trained and assessed using the X-ray dataset.

 TABLE III.
 SIX CT MODELS ON COVIDX-CT DATASET WITH

 SINGLE CLASSIFIER
 VIA

Model	Information					
Name	Precision	Recall	f1	Accuracy	Time	
COVID-Net [26]	0.98	0.97	0.99	99%	3 hours	
VGG [27]	0.90	0.90	0.90	94%	4 hours	
ResNet [28]	0.92	0.94	0.93	94%	5 hours	
Bayesian [29]	0.90	0.71	0.79	78%	6 hours	
DarkNet [30]	0.99	0.90	0.95	99%	3 hours	
DenseNet [31]	0.82	0.79	0.80	80%	6 hours	

TABLE IV. SIX X-RAY MODELS ON COVIDX- DATASET WITH SINGLE CLASSIFIER

Model Name	Information					
Widder Manie	Precision	Recall	f1	Accuracy	Time	
COVID-Net	0.96	0.96	0.95	94%	4 hours	
VGG	0.70	0.72	0.70	85%	5 hours	
ResNet	0.93	0.94	0.94	95%	4 hours	
Bayesian	0.70	0.65	0.69	70%	6 hours	
DarkNet	0.98	0.90	0.94	98%	3 hours	
DenseNet	0.82	0.80	0.84	86%	5 hours	

On the initial steps with a single classifier, DarkNet outperformed COVID-Net on the CT dataset, with an accuracy rate of 99% and some change in the values of

PRECISION, RECALL, and F1, as shown in TABLE III, as it adapts to new datasets and has the optimal cost function. TABLE IV, on the other hand, shows that the DarkNet model has the best detection accuracy with 98% on the X-ray dataset, as well as PRECISION 0.98, RECALL 0.9, and F1 0.94. ResNet ranks in second place in terms of accuracy, with a score of 95%.

TABLES V and VI shows the improved accuracy of applying our ensemble stacking approach to these 6 models on both CT and X-ray dataset.

TABLE V.	SIX CT MODELS ON COVIDX-CT DATASET WITH
	PROPOSED ENSAMBLE STACKING APPROACH

Model Name	Information					
	Precision	Recall	f1	Accuracy	Time	
COVID-Net	1.00	0.99	1.00	99.9%	4 hours	
VGG	0.95	0.94	0.96	96%	6 hours	
ResNet	0.95	0.96	0.94	97.4%	7 hours	
Bayesian	0.94	0.86	0.84	82%	10 hours	
DarkNet	1.00	0.96	0.98	99.4%	5 hours	
DenseNet	0.90	0.84	0.89	86%	7 hours	

TABLE VI.	SIX X-RAY MODELS ON COVIDX DATASET WITH
PROPOSE	D ENSAMBLE STACKING APPROACH

Model Name	Information				
	Precision	Recall	f1	Accuracy	Time
COVID-Net	0.99	0.98	0.99	97%	5 hours
VGG	0.84	0.90	0.85	90%	6 hours
ResNet	0.99	0.96	0.97	98 %	7 hours
Bayesian	0.90	0.89	0.82	76.9%	8 hours
DarkNet	0.99	0.96	0.98	99.7%	6 hours
DenseNet	0.92	0.90	0.93	90%	8 hours

As can be seen in tables the large variation in metrics, particularly in models with low scores when employing a single classifier. On the other hand, one of the major drawbacks of proposed ensemble stacking approach is computation time. That comes from the use of many classifiers. However, it is still acceptable in our application, where accuracy is more valuable than minor increases in computation time

COVID-Net still achieves higher efficiency with an accuracy of 99.9%, as demonstrated in TABLE V in the CT dataset, but both models improve at high rates, as indicated in the table. DarkNet, on the other hand, produced the best results in the X-ray dataset, as shown in TABLE VI, with an accuracy of 99.7% and improved the rates and results for all other models. In order to compare all the indicator results of all models together, we made a result radar chart to represent all the evaluation metrics values, as shown in fig. 5. and fig.

6. It is obvious shown that the further the line is from the center of the circle, the better the performance of the corresponding index. The chart shows that the accuracy, sensitivity and specificity lines corresponding to our results are all close to the periphery of the shape, which means that all performance results are high. And this also shows that our method has higher performance than the other methods.

On the other hand, one of the major drawbacks of proposed ensemble stacking approach is computation time. That comes from the use of many classifiers. However, it is still acceptable in our application, where accuracy is more valuable than minor increases in computation time.



Fig. 5. Radar Chart For Proposed Ensamble Stacking Approach On COVIDX Dataset



Fig. 6. Radar Chart For Proposed Ensamble Stacking Approach On COVIDX-CT Dataset

CONCLUSION

In this paper, we introduce stacked convolutional neural network for automated COVID19 diagnosis using chest CT and X-ray images in this study. The suggested model can learn image discriminating characteristics and extract a wide range of information from chest X-rays and CT scans. The model performed well in terms of accuracy and memory consumption, and we expect to work on a solution to the time consumption issue that hampered the model's performance in the future.

On the CT images of the COVIDX-CT dataset, the COVID-19 detection stacking model obtains a classification accuracy of 99.9%, recall of 0.99, and F1-score of 1.00, while on the X-ray images of the COVIDX dataset, it achieves a classification accuracy of 99.7%, recall of 0.96, and F1-score of 0.98. The findings of our studies demonstrate that the stacked CNN is successful at classifying COVID-19, Normal, and Pneumonia X-ray images. More notably, in the chest radiography classification, the suggested ENSAMBLE STACKING model outperforms.

REFERENCES

[1] W. U. O. COVID-19, 2020. https://www.who.int/emergencies/diseases /novel-coronavirus-2019.

[2] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, Journal of medical virology 92 (4) (2020) 418-423.

[3] J.P. Kanne, B.P. Little, J.H. Chung, B.M. Elicker, L.H. Ketai, Essentials for radiologists on COVID-19: an update-radiology scientific expert panel, Radiology 296 (2) (2020) E113–E114.

[4] 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, J. Med. Syst. 44 (8) (2020).

[5] D. Das, K.C. Santosh, U. Pal, Truncated inception net: COVID-19 outbreak screening using chest X-rays, Phys. Eng. Sci. Med. (2020).

[6] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding,

A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, arXiv preprint arXiv:1711.05225 (2017).

[7] X.L.A.D. Zhu, COVID-XPERT: an AI Powered Population Screening of COVID-19 Cases Using Chest Radiography Images, arXiv:2004.03042v1 [eess.IV], APRIL 8, 2020.

[8] C. Butt, J. Gill, D. Chun, B. A. Babu, Deep learning system to screen oronavirus disease 2019 pneumonia, Applied Intelligence (2020) 1.

[9] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision andpattern recognition, 2016.

[10] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015.

[11] C. Liu, Y. Cao, M. Alcantara, B. Liu, M. Brunette, J. Peinado, W. Curioso, Tx-cnn: Detecting tuberculosis in chest x-ray images using convolutional neural network, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017.

[12] C. Butt, J. Gill, D. Chun, B. A. Babu, Deep learning system to screen oronavirus disease 2019 pneumonia, Applied Intelligence (2020) 1.

[13] P.K. Sethy, S.K. Behera, P.K. Ratha, P. Biswas, Detection of coronavirus disease (COVID-19) based on deep features and support vector machine, Int. J. Math. Eng. Manage. Sci. 5 (4) (2020) 643–651.

[14] R. Kumar, et al., Accurate prediction of COVID-19 using chest X Ray images through deep feature learning model with SMOTE and machine learning classifiers, medRxiv (April) (2020) 1–10.

[15] A. Makris, I. Kontopoulos, K. Tserpes, COVID-19 detection from chest X-ray images using deep learning and convolutional neural networks, medRxiv (December) (2020).

[16] E. Dawod, N. Mahmoud, A. Elsisi, (2021). 'Investigation of Deep Convolutional Neural Network (CNN) approaches' accuracy for the detection of COVID-19', IJCI. International Journal of Computers and Information, 8(1), 2021.

[17] Q. Liang, W. Zhu, W. Sun, Z. Yu, Y. Wang, D. Zhang, In-line inspection solution for codes on complex backgrounds for the plastic container industry, Measurement: J. Int. Measurement Confed. 148 (2019).

[18] E.E.-D. Hemdan, M.A. Shouman, M.E. Karar, COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images, arXiv preprint arXiv: 2003.11055. 2020 Mar 24, 2020.

[19] L. Breiman, Stacked regressions, Machine learning 24 (1) (1996) 49-64.

[20] T.G Dietterich, : Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Models, pp. 1–15 (2000)

[21] A. Goyal, N. Sardana, : Empirical analysis of ensemble machine learning techniques for bug Triaging. In: 2019 Twelfth International Conference on Contemporary Computing (IC3), pp. 1–6 (2019)

[22] L. Deng, D.Yu, J. Platt, : Scalable stacking and learning for building deep architectures. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE (2012).

[23] B. Pavlyshenko, "Using Stacking Approaches for Machine Learning Models," 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 2018.

[24] GitHub. Lindawangg/COVID-Net. [online] Available at: https://github.com/lindawangg/COVID-Net [Accessed 7 February 2021].

[25] GitHub. Haydengunraj/COVIDNet-CT. [online] Available at: https://github.com/haydengunraj/COVIDNet-CT [Accessed 7 February 2021].

[26] L. Wang, A. Wong, COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases From Chest X-Ray Images, arXiv: 2003.09871v2 [eess.IV], 30 Mar, 2020.

[27] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays, Comput. Methods Programs Biomed. 196 (2020).

[28] H. Panwar, P.K. Gupta, M.K. Siddiqui, R. Morales-Menendez, V. Singh, Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet, Chaos Solitons Fractals 138 (2020).

[29] X. Li, L. Ding, L. Wang and F. Cao, "FPGA accelerates deep residual learning for image recognition," 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, 2017.

[30] L.Wang, Z.Lin and A.Wong, COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Scientific Reports, 10(1), 2020.

[31] A. Narin, C. Kaya, Z. Pamuk, Department of Biomedical Engineering, Zonguldak Bulent Ecevit University, 67100, Zonguldak, Turkey, arXiv preprint arXiv:

2003.10849.,2020