Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

COVID-19 DISEASE RECOGNITION USING DISTRIBUTED DATA MINING AND DEEP LEARNING

AHMAD ABADLEH[1],AFNAN AL-SARAIREH[1] , HAMZEH EYAL SALMAN[2], ANAS AL-AKASASBEH[1], SAQER ALJA'AFREH[4], AWNI HAMMOURI[3], RA'FAT AL-MSIE'DEEN[2], AHMAD HASSANAT[1]

[1]Mutah University, Information Technology Faculty, Computer Science Dept., Jordan
[2]Mutah University, Information Technology Faculty, Software Engineering Dept., Jordan
[3]Mutah University, Information Technology Faculty, Data Science Dept., Jordan
[4] Mutah University, Engineering Faculty, Jordan

E-mails`: ahmad_a@mutah.edu.jo, afnansaraireh92@gmail.com , hamzehmu@mutah.edu.jo, anas.alkasasbeh@mutah.edu.jo, eng.saqer-jaa@mutah.edu.jo, hammouri@mutah.edu.j, rafatalmsiedeen@mutah.edu.jo , hasanat@mutah.edu.jo

ABSTRACT

Since the first COVID-19 case was reported at the end of 2019 in China, the COVID-19 virus moved to almost every country with rapidly spread among people. It has a destructive effect on people's health and their daily life. The world health organization (WHO) in April 2020 officially declared the COVID-19 as a pandemic. To date, there is no specific treatment for COVID-19. Therefore, the detection of COVID-19 disease is required to avoid the fast spread of disease and halt its chain. In this article, we suggest an approach to recognize COVID-19 through X-ray images using distributed data mining techniques and Convolutional Neural Networks. To validate the proposed approach, we apply it to a public dataset consisting of 2,905 chest X-ray images for COVID-19 patients in addition to Viral Pneumonia and Normal images. The results show that the suggested approach gives promising results in terms of well-known evaluation metrics in the subject.

Keywords: *COVID-19, Recognition, Classification, Deep Learning, CNN, Distributed Data Mining.*

1. INTRODUCTION

The first COVID-19 case was recorded in December 2019 in Wuhan, China. Although all countries imposed severe restrictions to prohibit the propagation of the virus, they failed to limit the virus spreading. Therefore, the WHO in April 2020 officially declared the COVID-19 disease as a pandemic. This disease causes mild to moderate respiratory illness, cough, shortness of breath, muscle pain, fatigue, sore throat, diarrhea and loss of smell, and abdominal pain [1][2]. Some infected people suffer from developed deadly pneumonia with a death rate of 2%, which may occur due to massive alveolar damage and progressive respiratory failure [3]. To date, there is no specific treatment for COVID-19 other than prevention before contracting the virus, as most countries have taken preventive measures with people who have symptoms of early COVID-19, such as early quarantine of infected people and monitoring the spread of the disease on a wider range in the country.

The standard screening method used to test patients with COVID-19 is PCR test [4]. It is one of the popular methods to detect COVID-19 but it is time-consuming as it requires 7-8 hours to get the results with a positive rate of 63% [5]. Other diagnostic tools may be used to detect COVID-19 such as epidemiological history; positive radiographic images (computed tomography (CT) Chest radiography (CXR)). Although CXR images may aid in early detection of infected cases, various of viral pneumonia images are similar, and they intersect with other infectious and inflammatory lung diseases. In 2021, there has been an increase in the number of X-rays images available from medical cases, particularly from COVID-19 patients. This allows us to examine medical photos and discover all features that may lead to the analysis and recognition of patterns that will aid in the automatic diagnosis of the disease using various data mining approaches.

Data mining is a process, science, and technology to extract the implicit information, discover relationships, and knowledge that can be useful from the mass, incomplete, noise, and random data using data analysis tools[6][7]. Data mining commonly includes four classes of tasks: clustering, classification, regression, and association rules. Recently, the development of data mining and deep learning applications has helped many medical researchers to detect the diseases such as brain tumor detection [8], breast cancer detection [9], object detection from medical images [10], denoising medical images [11], medical image segmentation [12], etc. Moreover, AI techniques have been widely used in different fields such as localization [13][14] [33] and Security [15][16][17], etc.

Due to the current Internet era and cloud computing, data regardless of its domain types (medicine, financial, etc.) is scattered over the world and stored on distributed computing resources in the Internet cloud. Traditional data mining techniques involve integrating this distributed data into a single data warehouse. However, it is well-known that integration multiple sources of data into a single large data warehouse will produce such as poor schema design, lack of integrity constraints, spurious data, semantic problems on data, duplicated data, etc.[18]. In this case, the traditional application of data mining techniques (centralized data mining) results in misleading information and decisions, especially, if this data is medicine data. Distributed data mining techniques mitigate this problem to discover knowledge from distributed data sources without the need to collect them in a single data warehouse.

Our suggestion here is to recognize COVID-19 disease through X-ray images using distributed data mining techniques and Convolutional Neural Networks (CNN). To validate the proposed approach, we apply it to a public dataset consisting of 2,905 chest X-ray images for COVID-19 patients in addition to Normal and Viral Pneumonia images. The findings reveal that our suggestion provides promising results in terms of well-known evaluation metrics in the subject: Precision, Accuracy, Recall, and F1-Score. In addition, we perform an analysis for false-negative and false-positive cases.

Our article is structured as follows. We provide the necessary background in section 2. Related work is listed in section 3. Next, the proposed is detailed in section 4. The experiments are discussed in section 5. Finally, we provide the conclusions in section 6 with future perspectives.

2. BACKGROUND

We present an overview about distributed data mining (DDM) and data quality problems, and Convolutional Neural Networks (CNN), respectively.

2.1 DDM and Data Quality Problems

Data mining is KDD (Knowledge Discovery in Databases) process to extract knowledge from data. The main tasks in data mining include clustering, classification, association rules, and regression. Distributed data mining is a growing topic in data mining that allows to perform parallel and distributed data mining tasks [19]. As our work focuses on the classification task, we explain the distributed data mining through this task.

Generally, distributed data mining consists of two levels of processing [19]: (i) the local level, and (ii) the global level. At the local level, each local site's data is analyzed and classified separately from data from other sources. Then, for each data model of each local site, a set of representatives is created.. At the global level, these representatives of each model are transferred to a central site to form the global model. This global modeling represents global classification that should be sent to each local site to update the classification. Figure 1 displays these levels of processing in distributed data mining.

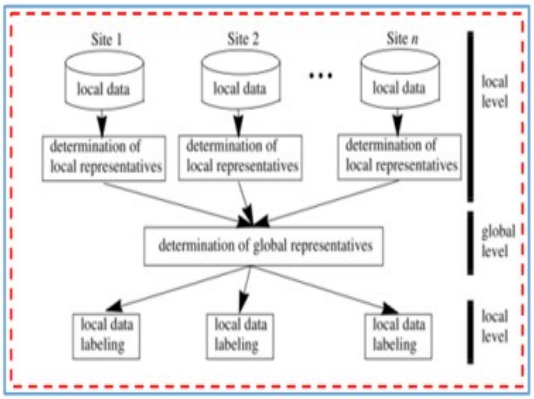


Figure 1:An Overview of Distributed Data Mining Process[19].

Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

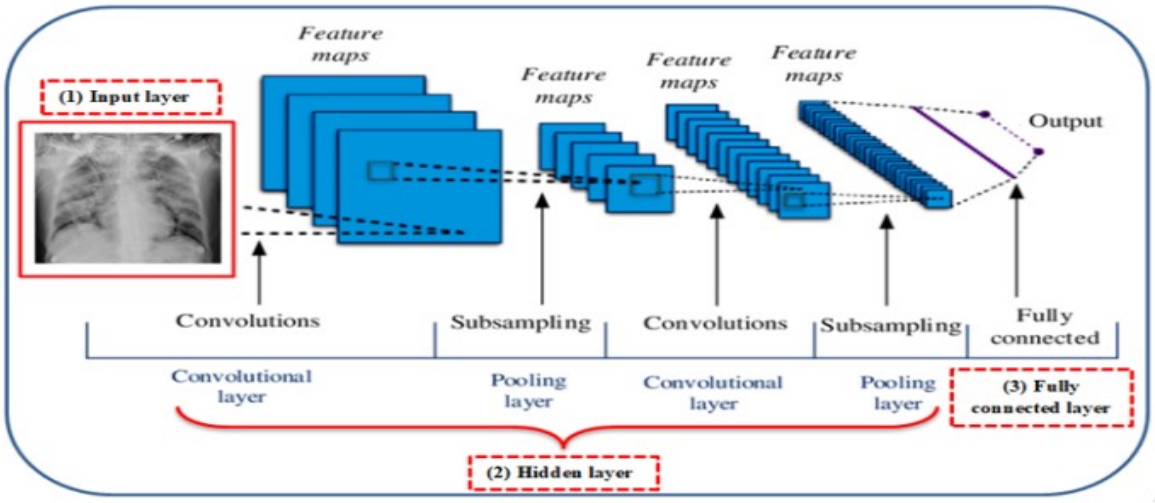


Figure 3. Convolutional Neural Network Architecture.

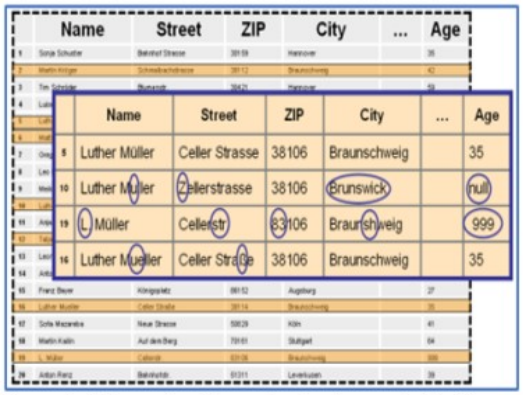


Figure 2: Illustrative Example for Semantic Problems [18]

Data quality problems are mainly classified into two main categories [20]: *multiple-source problems* and *single-source problems*. The problems in these categories are further classified into two levels: instance-based level and schema-based level. Problems of Instance-based level occur due to data entry problems such as misspelling, values missing, data redundancy, etc. Problems of Schema-based level raise due to bad schema design, lack of integrity constraints, differences between the data model and schema model such as uniqueness violation, referential integrity constraints, and missing attributes.

In order to apply centralized data mining techniques, several distributed data should be collected into a single repository to build a large data warehouse [18]. Such integration of data raises other data quality problems (for example, semantic problems). This is because the collected data represents different departments of data where each department has its own model and semantic which differs from other departments. Figure 2 shows an example of a semantic problem for data integrated from different sources. For example, the attribute *City* show different names from different languages (English and Germany) to name the same city: *Braunschweig* and *Brunswick*.

2.2 CNN

CNN is a Deep Learning algorithm that takes, as an input, an image and assigns adjustable weights to various features extracted from the input image to assign a label, as output, to this image [21][22]. The main advantage of CNN over traditional machine learning algorithms is that it eliminates the need for manual features extraction from the input image as such extraction is time-consuming.

The general architecture of CNN consists of several layers as displayed in Figure 3: (i) input layer, (2) set of convolutional/pooling layers, and (3) fully connected layer [21]. Firstly in the input layer, CNN takes an image as an input. This image is transformed into an array of pixel values. Then, the purpose of convolutional/pooling layers is to extract features from image parts and analyze them. This is achieved by performing a set of convolution kernels by dividing the image into a set of non-overlap rectangles. Finally, in the fully connected layer, the output of the previous layer is received and turned into a single vector to predict the image class.

3. RELATED WORK

In this section, the most recent and relevant works that are the closest to our work are presented. We divide these works into two categories:

COVID-19 diagnostic methods and identifying data quality problems using data mining methods

3.1 COVID-19 Diagnosis Methods

There is a research body to detect COVID-19 using chest X-ray images. This research work produces binary (normal or infected) or multiple classifications (infected, normal, viral pneumonia, and bacterial pneumonia) for COVID cases. The proposed approaches vary in terms of AI algorithms used. Among algorithms, the most preferred algorithm is CNN [23]. In this section, we restrict our self to present only research works that employ CNN to detect COVID-19 using chest X-ray images.

Chowdhury et al. proposed a CNN-based approach for automatic detection of COVID-19 pneumonia classification (Normal and COVID-19 pneumonia) and (Normal, viral, and COVID- 19 pneumonia) using chest X-ray images [24]. To train their learning model, they used a public database created by merging three public databases from recently published articles. This database contains 423 COVID-19, 1485 viral pneumonia, and 1579 normal chest x-ray images. The classification results showed that their approach can improve the accuracy and speed of diagnosing COVID-19 disease with 98.3% accuracy. For multi-classification, our approach combines deep learning and distributed data mining, which is more accurate than existing approaches like [24].

Apostolopoulos and Mpesiana evaluated the state-of-the-art of CNN architectures in medical image classification [25]. A dataset with X-ray images for patients having Covid-19 disease, common bacterial pneumonia, and normal incidents was used for automatic detection of the COVID-19 disease. The experimental results show that deep learning with CNN using X-ray images help to extract features related to the COVID-19 disease with 96.78% of accuracy.

Narin et al. investigated the result of using five pre-trained CNN-based models (*ResNet101, InceptionV3, ResNet50, InceptionV3*, and *Inception-ResNetV2*) to detect COVID patients automatically using chest X-ray [23]. These models were applied on a dataset with normal, COVID-19,

bacterial, and viral pneumonia cases. Three binary classifications are produced as an output of these models (COVID-19 or normal, COVID-19 or viral pneumonia, COVID-19 or bacterial pneumonia). The results indicate that the pre-trained ResNet50 model gives the highest classification performance with 98% of accuracy.

Zhang et al. develop a deep learning-based model to diagnose COVID-19 disease using chest X-ray images [26]. Their model is applied on 100 chest X-ray images for 70 patients. Their developed model achieves high sensitivity with 96%. Ghoshal and Tucker investigated the results of the dropweights-based Bayesian CNN model to detect COVID-19 patients with chest X-ray images [27]. Sahinbas and Catak evaluated the results of using five pre-trained CNN models (VGG16, VGG19, ResNet, DenseNet, and InceptionV3) to detect COVID-19 patients from non-COVID-19 patients using chest X-ray images [28]. The VGG16 model help to detect COVID-19 with the highest classification performance as 80% accuracy among the other models. Medhi et al. proposed a CNN-based approach to detect COVID-19 disease [29]. Their method is applied on a data set consisting of 150 confirmed COVID-19 cases. Experimentally, their method detects the infected cases with 93% accuracy.

3.2 Using Data Mining Methods for Identifying Data Quality Problems

Applying DDM techniques to detect data quality problems is seldom considered. Identifying data quality problems using DDM is only introduced in the article published by Januzaj et al. In [18], Januzaj et al. proposed to use distributed data mining techniques to detect data quality problems from financial distributed data. According to their proposed approach, the data quality problems can be identified by using a classifier that can use knowledge from sub-clustering. The initial experimental evaluation show that data quality problems can be identified without the need to integrate distributed data into single data warehouse.

A survey for methods and techniques to deal with data quality problems is presented in [30]. Eshref and Visar proposed an approach based on traditional (centralized) data mining techniques such as clustering, sub-clustering, and classification to identify quality data problems during the integration process of distributed data. Different approaches are proposed to detect outliers as rule-based approaches [31][32]. Machine learning and deep learning approaches are used for different purposes such as classification, detection, and others [34-39].

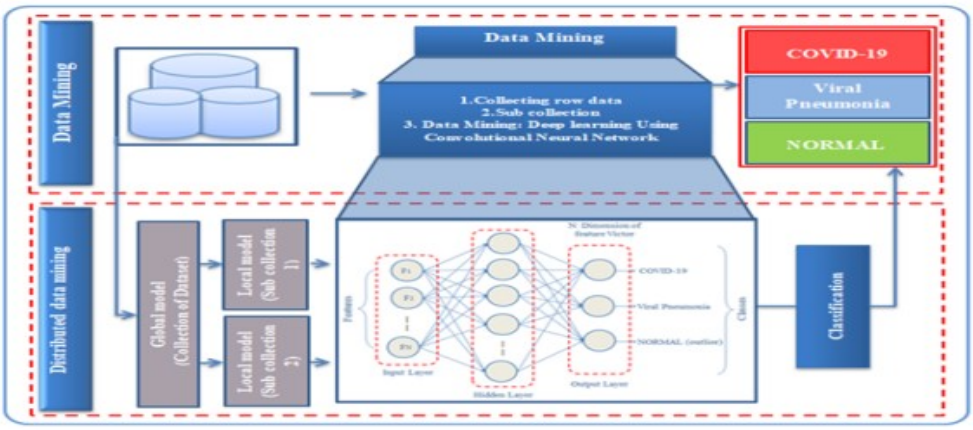


Figure 4. An Overview of COVID-19 classification using centralized and distributed data mining.

Our approach presented in this article differs from the existing works presented above. Firstly, our proposed approach investigates the results of using a combination between distributed data mining and CNN against using centralized data mining to avoid data quality problems occur during the integration process. Secondly, the proposed approach is a multi-classification of COVID-19 disease with three classes (*normal, COVID-19, and viral pneumonia*).

4. THE PROPOSED APPROACH

In this section, we firstly provide an overview of centralized and distributed data mining systems. Then, we detail the application of distributed data mining.

4.1 An Overview of The Proposed Approach

Figure 4 shows a centralized data mining system versus distributed system for COVID-19 classification. Both systems use CNN as a deep learning algorithm. As displayed in the figure, centralized data mining resides on the top while distributed data mining resides on the bottom. In distributed one, we divide the dataset of interest into subsets to simulate the distributed situation. We collect the data locally on each sub-collection site to be locally analyzed. Then, we classify the distributed data on each site into 3 classes: infected with COVID-19, infected with Viral Pneumonia, and Normal. This is performed by applying a uniform classifier to each site. In order to train the classification model, we utilize the global model as training data. Then, we use the labeled local data as a test data.

4.2. Applying DDM

This section presents in detail our proposal for COVID-19 classification based on distributed data mining. Our proposed framework consists of three main phases: image preprocessing, building a global model for training, and building two sub-collections for testing. Sequentially, we detail these phases.

4.3. Image Preprocessing

In the first phase, we need to set a standard size for all chest X-ray images because not all images have the same size. So all images are standardized to (224×224) pixels. Then they are converted to gray-scale images. Finally, all images are converted to matrix format to recognize them by the CNN algorithm.

4.4 Building Global Model for Training

In the second phase, our dataset of interest that contains 2,905 chest X-ray image is used as training data to build a global model. The input layer contains (224×224=50,176) neurons in the CNN network corresponding to the training images. The hidden layers consists of a convolutional layer with three convolutional filters (feature maps) with a pixel kernel window applied over the input patch. A maximum pooling layer with (2×2) sub-sampling ratios is the next layer. The *Rectified Linear Unit* (ReLu) activation function is employed in the fully connected layer, whereas the *Softmax* activation function is used in the output layer. The mathematical equation of these functions as follows:

$$ReLU(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } x \geq 0, \end{cases} \quad (1)$$

$$softmax(x_i) = \frac{\exp(x_i)}{\sum_{y=1}^{m} \exp(x_y)} \quad (2)$$

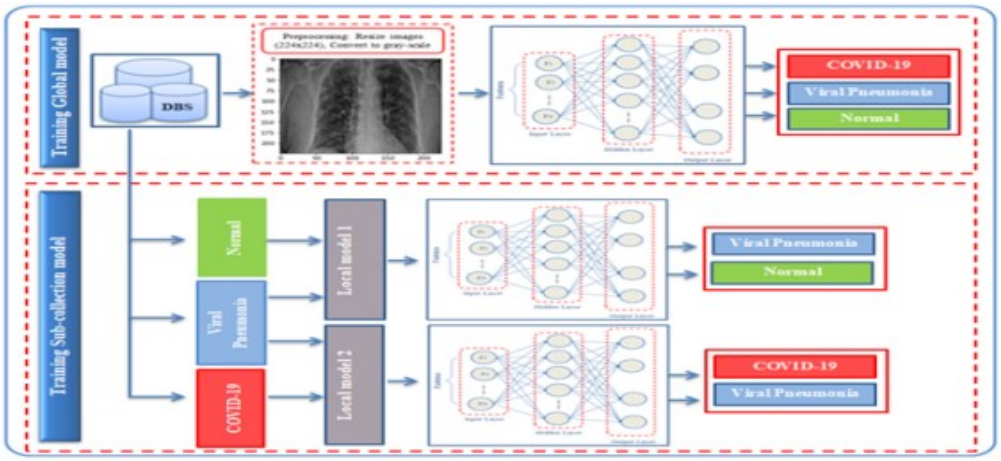


Figure 5. Steps of building global Model for training.

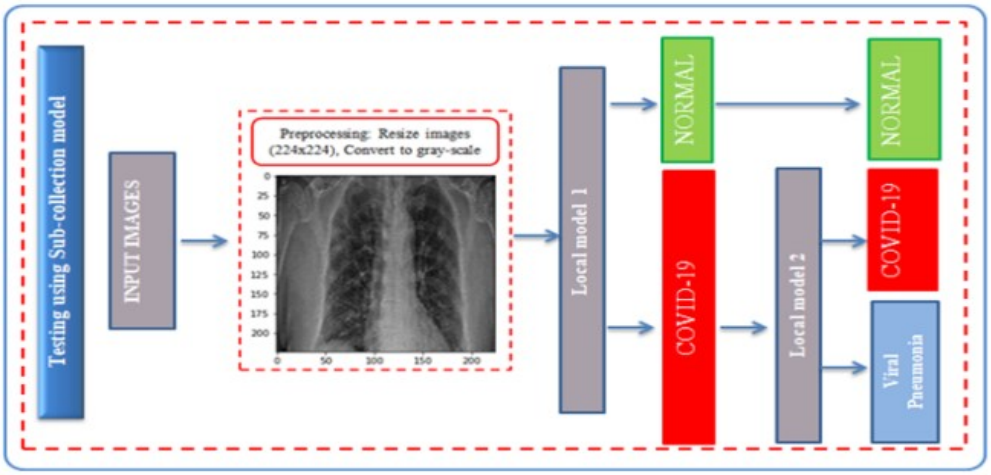


Figure 6. Two sub-collection for testing.

where x and m refer to input data and number of classes, respectively.

Similarly, in the two local models, we used *ReLu* activation function in the fully connected layer while we used a *sigmoid* function in the output layer which means whatever the input, the output ranging between 0 and 1. Every neuron will be scaled to a value between 0 and 1. The training steps of this phase are summarized in Figure 5.

$$Sigmoid(\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{x})} \quad (3)$$

4.5 Building Two Sub-Collections for Testing

For testing, we use the two generated local models to test the input image. Firstly, the input image is tested using the first local model to classify it as either a COVID-19 case or a normal (not Uninfected) case. Secondly, if the classification result from the first local model is COVID-19, the same image is taken as input to the second local model to classify it as either viral pneumonia or only COVID-19. All these steps are summarized in Figure 6.

5. RESULTS AND EVALUATION

we describe in this section the dataset used to validate our proposal. Then, we assess the performance of the global model and local data models, respectively.

5.1 Dataset Description and Performance Metrics

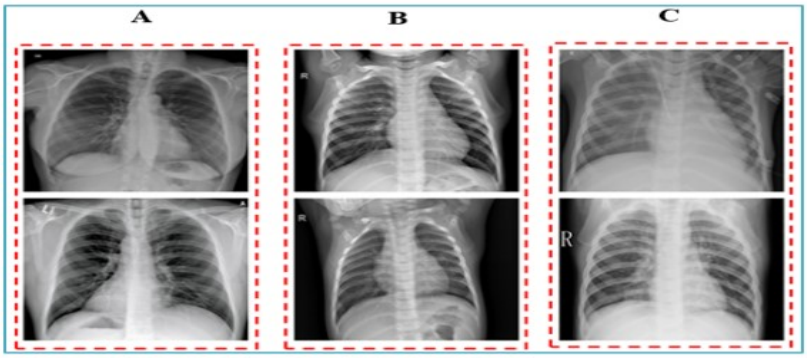


Figure 7. Chest X-ray samples from the dataset. (A) normal image,(B) COVID-19 image, (C) viral pneumonia image.

Table 1: Evaluation metrics results for the global model.

Class Name	Avg. Accuracy	Precision	Recall	F1-score
COVID-19	97.43%	99%	97%	98%
Normal		99%	99%	99%
Viral Pneumonia		98%	99%	99%

To asses the effectiveness of our proposed approach, we have applied it to a COVID-19 public dataset of chest X-ray images [24]. Our assessment process considers 2,905 chest X-ray images, 219 of COVID-19 images, 1,345 of viral pneumonia images and, 1,341 of normal images. Figure 7 displays samples from the dataset, classified as COVID-19 positive cases images in the first column, normal images in the second column, and viral pneumonia in the third column.

We have implemented our proposal using a *TensorFlow* machine learning package presented by Google which is a Python-embedded open-source library to develop and train ML models. The model was implemented on a PC Pentium i5 3.2 GHz and with 8 GB RAM.We have used four well-known evaluation metrics to asses the effectiveness of our proposal. These metrics are:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + TN} \quad (5)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$F1 - \text{score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Where FP, TP, FN, and TN in Equations 4-7 refer to the number of false positive, true positive, false negative and true negative, respectively.

5.2. Global Training Model Performance

To train the CNN model for building the global model, the dataset is splitted into two parts: training and validation datasets with 30%. The total number of samples is 2,905. However, the model is trained with 2,033 samples and validated with 872 training. The dataset is performed up to a maximum of 25 epochs using the CNN method with a batch size of 32. The performance results of the global model are displayed in Table 1 for each class. As can be seen, the proposed approach yields a high performance in the global model as the average accuracy is 97.43% after 25 epoch. Additionally, the proposed approach achieves (98% - 99%), (97% - 99%), and (98% - 99%) Precision, Recall, F1-score, respectively, for all classes.

Figure 8 displays the accuracy of training and validation data over different epochs. It explains that the global model provides the best results at epoch 25. The given model is the most suitable when training the given data due to its ability to get close to all the points with a minimum validation loss and training loss. In addition, it shows a maximum value of training and testing accuracy.

5.3. Local Data Models Performance

Table 2: Evaluation metrics results for local data models.

Model Name	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score
Local Model One	97.55%	98.58%	98.43%	98.05%
Local Model Two	98.04%	98.14%	96.80%	97.47%

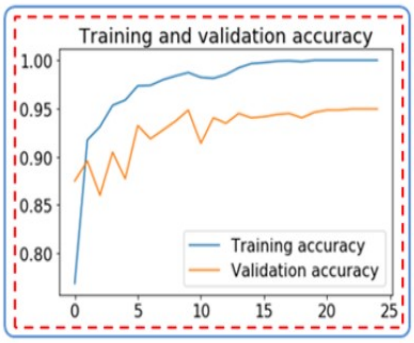


Figure 8. Accuracy results of training and validation data for the global model .

Similar to the global model performance, to train the CNN model for building the local models, the dataset is divided into two parts: training and validation datasets with 30%. In the local data model number one, the total number of the samples that are used is 2,686 where the model has been trained with 1,880 sample and validated with 806 sample. In the local data model number two, the total number of samples that are used is 1,564 where the model has been trained with 1,094 sample and validated with 407 sample. In both data models, the dataset has been performed up to a maximum of 25 epochs using the CNN algorithm with a batch size of 32. The accuracy of the entire system increases as the number of epochs increases; therefore, if the number of epochs is not sufficient, then the accuracy will decrease.

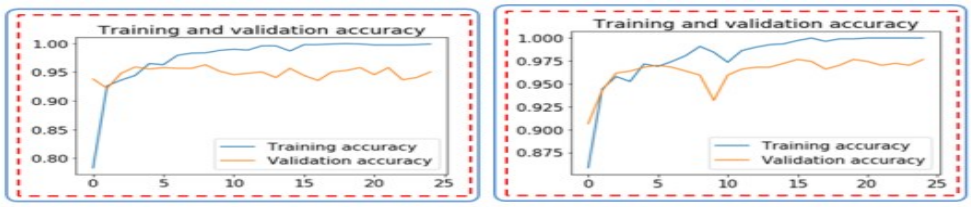
Table2 displays the results obtained by the generated local models using CNN. As shown, the CNN in the local model number successes in distinguishing viral pneumonia and the normal chest X-ray images with high average Accuracy, Precision, Recall, and F1-score (97.55%, 98.58%, 98.43%, 98.05%, respectively). Moreover, the CNN algorithm in the local model number two successes in distinguishing viral pneumonia and the CONID-19 chest X-ray images with high average Accuracy, Precision, Recall, and F1-score (98.04%, 98.14%, 96.80%, 97.47%, respectively). Based on the results shown in this table, we can conclude that the proposed approach is reliable to classify the provided chest X-ray images into viral pneumonia, normal, or COVID-19. This is because that the appropriate features are extracted from the given images.

Figure 9 illustrates the training and validation accuracy obtained by both the local model one and the local model two at different epochs. As shown, 25 epochs of training data revealed that the models fitted well. All points are covered by the models, and the training accuracy is maximized.

Moreover, Figure 10 represents the 2-class confusion matrix to further describe the performance of both local model one and local model two for classification chest X-ray images. The figure contains information about the actual and prediction classifications used to evaluate the performance of the classifier. As can be seen, the local model one produces 1,320 true negative instances, 19 false-positive instances, 21 false-negative instances, and 1,326 true positive instances. For local model two, the model produces 1,341 true negative instances, 4 false-positive instances, 7 false-negative instances, and 212 true positive instances. According to these numerical values, the designed models perform well and give a high number of correct predictions.

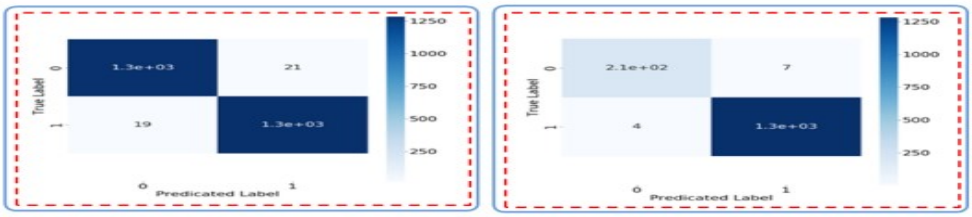
6. CONCLUSION

We have suggested a deep learning-based approach with distributed data mining to diagnose COVID-19 disease using chest X-ray images. Our suggestion has been applied on a public dataset of chest X-ray COVID-19 images containing 2,905 images (219 of COVID-19 images, 1,345of viral pneumonia images, and 1,341 of normal images. The findings reveal that the effectiveness of our proposal for COVID-19 diagnosis with high average accuracy (97.43%) for global model, and (97.55% and 98.04%) for two local models respectively using a completely automated computer program. The proposed approach is a multi-classification approach where we classify



(a) Local data model one. **(b)** Local data model two.

Figure 9. Accuracy results of training and validation data for local models.



(a) Local data model one. **(b)** Local data model two.

Figure 10. Confusion matrix of local models.

several diseases using the same deep learning approach. Moreover, the proposed approach utilizes distributed data mining and deep learning to avoid the problem of data quantity. However, our approach needs to be enhanced to support different data mining and deep learning techniques. As a future work, we plan to investigate the results of using other deep learning algorithms with more distributed datasets.

REERENCES:

[1] Jadhao, V.; Bodhe, R.; Mahajan, H.; Jadhav, V.; Patil, K.; Baheti, N.; Kale, N. A Novel Coronavirus (nCOV- 2019): A Pandemic Severe Respiratory Tract Infections by SARS COV-2 in Human. Journal of Drug Delivery and Therapeutics 2020, 10, 271–279.

[2] Shatnawi, M.; Shatnawi, A.; AlShara, Z.; Husari, G. Symptoms-Based Fuzzy-Logic Approach for COVID-19 Diagnosis. International Journal of Advanced Computer Science and Applications 2021, 12. doi:10.14569/IJACSA.2021.0120457.

[3] Xu, Z.; Shi, L.; Wang, Y.; Zhang, J.; Huang, L.; Zhang, C.; Liu, S.; Zhao, P.; Liu, H.; Zhu, L.; Tai, Y.; Bai, C.; Gao, T.; Song, J.; Xia, P.; Dong, J.; Zhao, J.; Wang, F.S. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. The Lancet Respiratory Medicine 2020, 8, 420–422. doi:https://doi.org/10.1016/S2213-2600(20)30076-X.

[4] Corman, V.M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D.K.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M.L.; Mulders, D.G.; Haagmans, B.L.; van der Veer, B.; van den Brink, S.; Wijsman, L.; Goderski, G.; Romette, J.L.; Ellis, J.; Zambon, M.; Peiris, M.; Goossens, H.; Reusken, C.; Koopmans, M.P.; Drosten, C. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance 2020, 25. doi:https://doi.org/10.2807/1560917.ES.2020.25.3.2000045.

[5] Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA 2020, 323, 1843–1844, [https://jamanetwork.com/journals/jama/articlepdf/2762997/jama_wang_2020_ld_200018.pdf]. doi: 10.1001/jama.2020.3786.

[6] Agrawal, R.; Imielinski, T.; Swami, A. Database mining: a performance perspective. IEEE Transactions on Knowledge and Data Engineering 1993, 5, 914–925. doi:10.1109/69.250074.

[7] Rahman, N. Data Mining Techniques and Applications: A Ten-Year Update. Int. J. Strateg. Inf. Technol. Appl. 2018, 9, 78–97. doi:10.4018/IJSITA.2018010104.

[8] Özyurt, F.; Sert, E.; Avci, E.; Dogantekin, E. Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy. Measurement 2019, 147, 106830. doi: https://doi.org/10.1016/j.measurement.2019.07.058.

[9] Masud, M.; Hossain, M.S.; Alhumyani, H.; Alshamrani, S.S.; Cheikhrouhou, O.; Ibrahim, S.; Muhammad, G.; Rashed, A.E.E.; Gupta, B.B. Pre-Trained Convolutional Neural Networks for Breast Cancer Detection Using Ultrasound Images. ACM Trans. Internet Technol. 2021, 21. doi:10.1145/3418355.

[10] Li, Z.; Dong, M.; Wen, S.; Hu, X.; Zhou, P.; Zeng, Z. CLU-CNNs: Object detection for medical images. Neurocomputing 2019, 350, 53–59. doi:https://doi.org/10.1016/j.neucom.2019.04.028.

[11] Ezer, B.A.; Zvi, Devir, E.; Dahan, Tal, K.; Rosman, G. Denoising medical images, 2013. Patent No. US8605970B2. 12.

[12] Mittal, M.; Arora, M.; Pandey, T.; Goyal, L.M. Image segmentation using deep learning techniques in medical images. In Advancement of machine intelligence in interactive medical image analysis; Springer, 2020; pp. 41–63.

[13] Alabadleh, A.; Aljaafreh, S.; Aljaafreh, A.; Alawasa, K. A RSS-based localization method using HMM-based error correction. Journal of Location Based Services 2018, 12, 273–285, [https://doi.org/10.1080/17489725.2018.1535140]. doi:10.1080/17489725.2018.1535140.

[14] Abadleh, A.; Han, S.; Hyun, S.J.; Lee, B.; Kim, M. Construction of indoor floor plan and localization. Wireless Networks 2016, 22, 175–191.

[15] Hamadaqa, E.; Abadleh, A.; Mars, A.; Adi, W. Highly Secured Implantable Medical Devices. 2018 International Conference on Innovations in Information Technology (IIT), 2018, pp. 7–12. doi:10.1109/INNOVATIONS.2018.8605968.

[16] Mars, A.; Abadleh, A.; Adi, W. Operator and Manufacturer Independent D2D Private Link for Future 5G Networks. IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2019, pp. 1–6.

[17] Mulhem, S.; Abadleh, A.; Adi, W. Accelerometer-Based Joint User-Device Clone-Resistant Identity. 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2018, pp. 230–237. doi:10.1109/WorldS4.2018.8611476.

[18] Januzaj, E.; Januzaj, V.; Mandl, P. An Application of Distributed Data Mining to Identify Data Quality Problems. Proceedings of the 21st International Conference on Information Integration and Web-Based Applications and Services; Association for Computing Machinery: New York, NY, USA, 2019; iiWAS2019, p. 418–422. doi:10.1145/3366030.3366103.

[19] Januzaj, E.; Kriegel, H.P.; Pfeifle, M. Scalable Density-Based Distributed Clustering. Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases; Springer-Verlag: Berlin, Heidelberg, 2004; PKDD '04, p. 231–244.

[20] Rahm, E.; Do, H.H. Data Cleaning: Problems and Current Approaches. IEEE Data Eng. Bull. 2000, 23, 3–13.

[21] Bhandare, A.; Bhide, M.V.; Gokhale, P.; Chandavarkar, R. Applications of Convolutional Neural Networks. International Journal of Computer Science and Information Technologies 2016, 7, 2206–2215.

[22] Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Transactions on Neural Networks and Learning Systems 2021, pp. 1–21. doi:10.1109/TNNLS.2021.3084827.

[23] Narin, A.; Kaya, C.; Pamuk, Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Analysis and Applications 2021, pp. 1 – 14.

[24] Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Mahbub, Z.B.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Emadi, N.A.; Reaz, M.B.I.; Islam, M.T. Can AI Help in Screening Viral and COVID-19 Pneumonia? IEEE Access 2020, 8, 132665–132676. doi:10.1109/ACCESS.2020.3010287.

[25] Apostolopoulos, I.D.; Bessiana, T. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine 2020, pp. 1 – 6.

[26] Zhang, J.; Xie, Y.; Li, Y.; Shen, C.; Xia, Y. COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection. ArXiv 2020, abs/2003.12338.

[27] Ghoshal, B.; Tucker, A. Estimating Uncertainty and Interpretability in Deep Learning for Coronavirus (COVID-19) Detection. ArXiv 2020, abs/2003.10769.

[28] Sahinbas, K.; Catak, F.O. Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images. In Data Science for COVID-19; Kose, U.; Gupta, D.; de Albuquerque, V.H.C.; Khanna, A., Eds.; Academic Press, 2021; pp. 451–466. doi:https://doi.org/10.1016/B978-0-12-824536-1.00003-4.

[29] Medhi, K.; Jamil, M.; Hussain, M.I. Automatic Detection of COVID-19 Infection from Chest X-ray using Deep Learning. medRxiv 2020. doi:10.1101/2020.05.10.20097063.
[30] Batini, C.; Scannapieco, M. Data Quality: Concepts, Methodologies and Techniques (Data-Centric Systems and Applications); Springer-Verlag: Berlin, Heidelberg, 2006.
[31] Hipp, J.; Güntzer, U.; Grimmer, U. Data Quality Mining – Making a Virtue of Necessity. IN PROCEEDINGS OF THE 6TH ACM SIGMOD WORKSHOP ON RESEARCH ISSUES IN DATA MINING AND KNOWLEDGE DISCOVERY (DMKD 2001, 2001, pp. 52–57.
[32] Li, X.; Shi, Y.; Li, J.; Zhang, P. Data Mining Consulting Improve Data Quality. Data Sci. J. 2007, 6, 658–666.
[33] Abadleh, Ahmad, et al. "Noise segmentation for step detection and distance estimation using smartphone sensor data." Wireless Networks 27.4 (2021): 2337-2346.
[34] Hassanat, Ahmad BA. "On identifying terrorists using their victory signs." Data Science Journal 17 (2018).
[35] Hassanat, Ahmad. "Greedy algorithms for approximating the diameter of machine learning datasets in multidimensional euclidean space: Experimental results." (2018).
[36] Hassanat, Ahmad. "Furthest-pair-based decision trees: Experimental results on big data classification." Information 9.11 (2018): 284.
[37] Hassanat, Ahmad B., et al. "Classification and gender recognition from veiled-faces." International Journal of Biometrics 9.4 (2017): 347-364.
[38] Tarawneh, Ahmad S., et al. "Invoice classification using deep features and machine learning techniques." 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). IEEE, 2019.
[39] Hassanat, Ahmad BA. "Two-point-based binary search trees for accelerating big data classification using KNN." PloS one 13.11 (2018): e0207772.