An H₂O’s Deep Learning-Inspired Model Based on Big Data Analytics for Coronavirus Disease (COVID-19) Diagnosis

Sally Elghamrawy

Abstract The outbreak of coronavirus diseases (COVID-19) has rabidly spread all over the world. The World Health Organization (WHO) has announced that coronavirus COVID-19 is an international pandemic. Big Data analytics tools must handle and analyse the massive amount of big medical data, generated daily, quickly due to the fact that time is very significant issue in healthcare applications. In addition, several deep learning algorithms are used along with big data analysis processes to help in detecting COVID-19 outbreaks and predicting their worldwide spread. Many researchers developed their models to diagnosis COVID-19 using Computed Tomography (CT) or X-ray imaging. This chapter presents a detailed discussion of Deep Learning and Big Data Analytics effects in containment of the disease. In addition, an H₂O’s Deep-Learning-inspired model based on Big Data analytics (DLBD-COV) is proposed for early diagnosis of COVID-19 cases using CT or X-ray images. The proposed diagnosis model is build based on the machine learning framework (H₂O) for scalable processing. The Generative Adversarial Networks (GAN) and the Convolutional Neural Networks (CNNs) are used and their classification results are compared. The experimental results emphasize the superiority of DLBD-COV when using H₂O framework for scalable COVID19 classification. The results obtained, using a dataset with thousands of real data and images, show encouraging performance using the automated feature extraction of deep learning techniques used in DLBD-COV.

Keywords Coronavirus disease (COVID-19) · CT imaging · X-Ray · Big data analytics · H₂O · Deep learning technique · Generative adversarial networks (GAN) · Convolutional neural networks (CNNs)
1 Introduction

The outbreak of an infectious pneumonia produced by the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) has initiated a universal panic. In Mar 28th 2020 [1], there have been 597,458 confirmed cases and 27,370 deaths. To the date (April 28th 2020), there have been 3,064,837 confirmed cases and 211,609 deaths all around the world [2]. It is obvious that in one month, the number of confirmed cases is increased by 5 times. These situation reports, announced by the World Health Organization (WHO), indicates that a rapid spread of the virus is forthcoming and COVID-19 is tremendouslyspreadable between people via respiratory droplets. As a result, it is critical to early detect infected cases for preventing the transmissible of the disease. The early diagnosis of COVID-19 helps the healthcare workers for applying appropriate treatment and quarantine procedures.

A number of researches [3–10] presented several COVID-19 diagnosis models using different Deep Learning techniques with Big Data analytics. These models are used for early discovery of COVID-19 to contribute in its containment. Several studies [5, 8–10] have proved that the chest Computed Tomography CT and X-ray images has high sensitivity over the Reverse Transcription Polymerase Chain Reaction RT-PCR test [7], which gives high false negative results. Accordingly, a CT and X-ray based diagnosis models are critically needed for precisely detecting COVID-19 cases.

In this chapter, an H2O’s Deep-Learning-inspired model based on Big Data analytics (DLBD-COV) is proposed for early diagnosis of COVID-19 using CT or X-ray images. H2O [11] is implemented during the different phases of DLBD-COV for scalable and speedy processing. The Convolutional Neural Networks (CNNs) [12] and the Generative Adversarial Networks (GAN) [13] are used for segmentation and augmentation process of the model. The diagnosis results obtained from these networks are then classified and their classification results are compared.

The rest of the chapter is organized as follows: Sect. 2 shows the main big data analytics frameworks. Section 3 presents the Deep Learning techniques: Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GAN). Section 4 summarizes the main fields of Big Data Analytics in containment of COVID-19. In Sect. 5, a detailed discussion of recent Deep learning techniques used against COVID-19. The proposed DLBD-COV diagnosis model is presented in Sect. 6. Section 7 presented the experimental results of DLBD-COV. Section 8 summarizes the conclusions and presents future work.

Main Contributions of the Proposed DLBD-COV Model

DLBD-COV model is proposed to exploit different deep learning techniques along with big data analytic tools to improve the diagnosis utility. The main advantage of DLBD-COV model is presented as follows:

I. DLBD-COV supports either X-ray or CT images for diagnosing COVID-19 cases.

II. DLBD-COV applies pre-processing techniques to remove noise or missing data.
III. The Generative Adversarial Networks (GAN) is used for X-ray images and the Convolutional Neural Networks (CNNs) is implemented for the CT images.

IV. The results obtained from the GANs and CNNs are classified and validated using three different classifiers.

V. Using H2O in the DLBD-COV reduces the computational time needed for diagnosis process.

VI. A comparison is made between the results obtained from DLBD-COV when using H2O and Spark. The results obtained showed that H2O proved an improved performance in terms of computational time over Spark. This is due to the availability of supported deep learning (Neural networks) algorithms in H2O.ai [14].

2 Big Data Analytics

Big Data is collected from numerous sources e.g. sensors, social media and digital images and videos. Big data analytics is used to analyze Big Data in order to extract motivating patterns and hidden relationships. The companies can process more date for the same price, which will increase their offer in the market. Therefore, they potentially can increase the total amount of sales. Finally, a large panoply of competitive advantages can be reached by companies [15]. There are thousands of Big data analytics tools that is employed for different tasks [16], as shown in the Table 1.

There are two well-known frameworks for big data analysis, namely H2O [11] and Apache Spark [17]. H2O is considered as the new generation of Machine learning technology. It proved its effective performance over different open source machine learning. It gives users the ability to discover patterns in data using huge numbers of models. H2O supports several types of Deep Learning algorithms [18–20] such as Convolutional Neural Networks (CNNs). In H2O’s Deep learning, the data is divided into partitions and then each partition is analyzed concurrently. Using this method provide H2O the ability to process and analyze all the data instead of neglecting most of it.

| Table 1 | Big data analytics tools and their corresponding tasks |
|---------|-------------------------------------------------------|
| Task    | Big data analytics tools |
| Data storage and management | Hadoop, MongoDB, Cassandra [51] |
| Data cleaning | OpenRefine, DataCleaner |
| Data mining | Rapid Miner, Waikato Environment for Knowledge Analysis (Weka), Oracle data mining, IBM Modeler |
| Data analysis | Qubole, BigML, Statwing. |
| Data visualization | Tableau, Silk, Charito, Plot.ly |
| Data integration | Blockspring, Pentaho |
| Data collection | Import.io |
When the data is well stored and partitioned, an advanced analytics tool can be implemented to analyze this data using techniques of data mining and deep learning [21]. Data mining techniques are used to detect relationships and patterns in data. Deep learning develops a suite of algorithms for analyzing big data.

3 Deep Learning

Deep Learning (DL), hierarchical learning or Deep Neural Network is a branch of Machine Learning (ML) which is based on artificial neural networks. Its main goal is to model high-level abstractions in data [22–24]. DL uses several layers to gradually extract upper level features from the raw input data, as shown in Fig. 1. In this Chapter, different DL techniques are presented which are used for processing COVID19 images. In DL techniques, the lower layers possibly will detect edges, but upper layers will detect the features (signs) diagnosing COVID19.

There are different types of Deep learning architectures concerned with computer vision and Image processing such as Convolutional Neural Networks [12, 25], Generative Adversarial Networks (GAN) [13, 26] and Deconvolutional networks [27].

3.1 Convolutional Neural Networks (CNN)

It is a recent Deep Learning algorithm, first proposed by Badrinarayanan [12]. It’s widely used in the fields of image processing, computer vision and classification. CNN primarily comprise of three kind of layers: convolutional, nonlinear and pooling layers. The convolutional layer is used for extracting the relevant features based on the weights assigned. The nonlinear layer is used to model any nonlinear function on feature maps. The pooling layers reduce the resolution of imaged by providing numerical information of a feature map. CNNs has advantage over other fully connected
neural networks, is that the parameters’ number in each field is reduced due to the ability of nodes in each layer to share weights. ResNet [28], VGGNet [29], GoogLeNet [30] and DenseNet [31] are most common CNN architectures.

3.2 Generative Adversarial Networks (GANs)

The Generative Adversarial Network (GAN) first introduced by Goodfellow [13] on 2014. GAN is a deep learning model, it is considered as a double framework that consists of generator and discriminator networks. The generative network created new objects while the discriminative network evaluates these newly generated objects. For this reason, GANs is widely used in real images generation.

The generator network trains a mapping, with a predefined distribution, to the target distribution of the real objects. The discriminator network tries to differentiate the generated objects from the real ones. GAN can be considered as game between generator and discriminator, in which the discriminator network attempts to reduce the classification error in differentiate between false samples from true ones. On the other hand, the generator network attempts to reduce the loss function [32, 33].

4 Big Data Analytics Against COVID-19

Recent studies used big data analysis tools to decrease the chance of spreading COVID-19 by monitoring COVID-19 disease and detecting the possible areas of infection. It is thus vital to utilize Big Data and intelligent analytics tools for monitoring COVID-19 disease to improve the community health. Many researchers exploit bid data and intelligent tools to monitor COVID-19 disease in different ways, as shown in Fig. 2.

- Data Tracking:

A recent study [34] has presented an approach for tracking passengers’ data traffic from a big-data source. This tracking will show the probability of forecasting [35] COVID-19 and prevent the further spread of novel coronavirus (2019-nCoV). They collect tracking data from the official flight companies and digital big data sources extracted from data from citizens’ mobile phones from the WeChat app, in order to provide information for risk management.

- Data Prediction

Leung et al. [36], at Wuhan city, uses big data storage tool to store and estimate number of travellers from Wuhan by using three different sources: the global flight bookings, the daily number of foreign and the domestic passenger volumes from
and to Wuhan. These data are then used to predict the forecasting of the epidemic in Wuhan based on the number of cases transferred from Wuhan to other cities.

- **Data Analysis:**

  Coronavirus (Covid-19 outbreak has proved the need for proactive containment. All countries must rapidly utilize their resources to save their people lives and their economic stability. Taiwan [37] has been alerted from China epidemics and acted early to contain the crisis. The authors in [37] stated that Taiwan used big data analytic tools to present different approached to identify and contain COVID-19 cases early to protect the public health. They stated that they use big data storage advantage to apply integration between national health insurance database and immigration database. This huge database and its tools are used to dynamically generate alerts of COVID-19 confirmed cases and severe cases during the medical examination of the patient, thus identify and diagnosis Taiwan’s cases and take rapid action.

- **Data Visualization:**

  Zhou et al. [38] used the geographic information systems (GIS) and big data technologies to aggregate different big data sources and quickly create visualization for COVID-19 outbreak information. They stated that the tools used helped in predicting the risk allocation and regional transmissions.

- **Data Diagnosis:**

  An integration between Big Data and artificial intelligence tools is a critical issue in detecting and preventing the spread of novel coronavirus. The authors in [39] investigated the importance of AI techniques with big data tools. They stated that
using the 3 V’s of Big data helped in creating different datasets that are used by many researchers and models. They reviewed a number of models that used AI and big data tools to contain the diseases like GoogleFlue [40] and BlueDot [41] that predicted the COVID-19 outbreak and sent alerts on December 31.

5 Deep Learning Against COVID-19

Due to the rapid development of deep learning technology that has been commonly implemented in the health arena, different studies [42–44] were conducted, based on deep learning technologies, for diagnosing and classifying different diseases like viral pneumonias and organs’ tumours.

Currently, due to the COVID-19 outbreak disaster, many researchers have been motivated to develop models for early diagnosing and detecting COVID-19 as follows: Authors in [3] proposed a 3D deep convolutional neural Network to Detect COVID-19 from CT volume, namely DeCoVNet. But, the algorithm worked in a black-box manner when diagnosing COVID-19, since the algorithm was based on deep learning and its explain ability was still at an early stage. COVNET [4] developed a framework to detect COVID-19 using chest CT and evaluate its performances. The authors proposed a three-dimensional deep learning framework to detect COVID-19 using chest CT. Community acquired pneumonia (CAP) and other non-pneumonia exams were included to test the robustness of the model.

In addition, Yang et al. [5] investigated the diagnostic value and consistency of chest CT as compared with comparison to RT-PCR assay in COVID-19. Their analysis suggests that chest CT should be considered for the COVID-19 screening, comprehensive evaluation, and following-up, especially in epidemic areas with high pre-test probability for disease. Jiang et al. [6] proposed established an early screening model to distinguish COVID-19 pneumonia from Influenza-A viral pneumonia and healthy cases with pulmonary CT images using deep learning techniques. The authors used multiple CNN models to classify CT image datasets and calculate the infection probability of COVID-19. The findings might greatly assist in the early screening of patients with COVID-19 by deep learning technologies. The authors proposed a location-attention mechanism and uses it in the classical ResNet for feature extraction. The authors in [7] constructed a system based on deep learning for identification of viral pneumonia on CT. AIMDP model proposed in [9] utilized different AI techniques to enhance the diagnosis and prediction function of the model.
6 The Proposed H₂O’s Deep-Learning-Inspired Model Based on Big Data Analytics (DLBD-COV) for COVID-19 Diagnosis

A novel Deep-Learning-inspired model (DLBD-COV) for COVID-19 diagnosis based on Big Data analytics tools is introduced in this section. The model consists of four main layers based on H₂O architecture to deal with large scale data collected, as shown in Fig. 3. DLBD-COV is used for early COVID-19 diagnosis using either X-ray or CT images based on Deep learning techniques.

At the beginning, the COVID-19 datasets are collected. As a starting point, in the pre-processing layer, the X-ray and CT dataset are loaded from the HDFS, of the H₂O Architecture, and stored as RDD object. An RDD wrapper is created using RDD method by the H₂O architecture. The RDD is used to improve the distributed parallel computation and to provide H₂O the ability to implement the iterative algorithms with effective fault tolerance. In this layer the test and train datasets are splitted for both X-ray and CT images.

Then, the dataset analyzer phase handles the noise and corrupted files for each dataset, separately, by the stability manager module. The DLBD-COV model will detect the type of the image loaded either it’s an X-ray or CT image using the Class detector. Finally, the number of images in each class is detected. The classes in the training dataset are stored in the images/classes repository. If the detected image belongs to the X-Ray class, then the GAN layer will handle these types of images. And if the detected image is in the CT class, then the CNN Layer will handle it, as shown in Fig. 3.

In general, H₂O’s Deep Learning is implemented for feedforward and multi-Layer neural networks models. In Deep learning techniques, the networks may consist of huge number of hidden, ReLU and Tanh Layers. Each layer comprises of number of nodes and neurons and thousands of features and parameters. These parameters are trained by nodes with the parallel computing feature of H₂O to parallelize the DL algorithms processes, which leads to reducing the overhead and computational time.

The H₂O’s Generative Adversarial Network layer is used for segmentation and for generating new X-ray images. Many researches [10, 13, 26] proved the superiority of using GAN with X-ray images. The generated data by GAN provides the model with large scale data for training the neural networks. This will lead to a significant improvement in the deep learning performance for classifying the new COVID-19 cases. In this layer, the generative network provides additional X-ray images that will be stored in generated images repository, as shown in Fig. 3. The data collector mixes the generated images with the original training data. The dataset is then augmented based on the generated dataset. Finally, the discriminator network detect each image whether its a real image or generated one. The results obtained will be evaluated in the classification layer by testing the accuracy of the generated data.

On the other hand, the CT images are handled by the H₂O’s Convolutional Neural Networks (CNNs). COVID-19 CT images contains specific features that uniquely identifies this pneumonia. CNNs applies number of filters to detect the COVID19
The Proposed H₂O’s deep-Learning-inspired model based on Big Data analytics (DLBD-COV)

Fig. 3
associated features from image. In the proposed model, the CNNs contains four main layers: Max Pool, convolution, Pooling and dense layers [9]. In the first pooling layer, the features in the CT image is reduces. Then, the relevant features are extracted using predefined weights in the convolution layer. The resolution of images is minimized using statistical data in the second pooling layer. Finally, the features with identical patters are grouped together in the dense layer. The output from H2O’s CNNs and H2O’s GAN’s are then delivered to the classification layer to diagnosis COVID-19 cases based on the features selected. The classifier picker is used to select the most appropriate classifiers, based on the delivered feature and type of image, from different classifiers: Support Vector Machine(SVM), Naive Bayes (NB) and Random Forest (RF). After classifying the data, the model is trained and validated in this layer.

The DLBD-COV model is conducted through the following stages, as shown in Fig. 4:

I. The dataset of X-ray or CT images for COVID-19 confirmed and suspected cases is loaded.

II. Applying pre-processing techniques to remove missing data.

Fig. 4 The Flow diagram for the DLBD-COV Model
III. The type of images is detected to determine the most appropriate Deep learning technique that best handle this type of image.

IV. The GAN is used for diagnosing the X-ray images.

V. Implementing the CNNs for the CT images.

VI. The results obtained from the GANs and CNNs are classified and validated using different classifiers.

VII. Using H$_2$O in the DLBD-COV reduces the computational time needed for the diagnosis process.

7 The Experimental Results

A number of experiments were conducted to validate the proposed DLBD-COV’s effectiveness. The model was implemented in MATLAB R2019a and are executed on a windows 10 PC with Intel(R) Core (TM) i7 CPU, with 16 GB RAM and 2.81 GHz clock speed. In the following experiments, a 10-fold cross validation is used. The stated results were taken as averages of the ten partitions. The CT scans chest dataset were collected from different resources [45–47]. The X-ray Images were collected from [48–50].

To validate the effectiveness of proposed model over other deep learning models, the accuracy, precision, recall(Sensitivity) and the computational time were used as the evaluations metrics, where:

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

where TP, TN, FP and FN are true positive, true negative, false positive, and false negative respectively. The significant measures of the performance are: True Positive Rate (TPR), True Negative rate (TNR) and, Positive Predictive Value (PPV), defined as follows,

$$\text{Recall(Sensitivity)} = \text{TPR} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \text{TNR} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{Precision} = \text{PPV} = \frac{TP}{TP + FP} \quad (4)$$
7.1 Experiment 1: Compare Between GAN and CNN Implemented in DLBD-COV

In order to validate the choice of choosing CNNs in diagnosis CT images and GAN for diagnosing X-ray images, the overall precision, recall and accuracy of DLBD-COV model is tested when diagnosing CT images only using GAN and CNN, as shown in Fig. 5. In addition, the overall precision, recall and accuracy of DLBD-COV model is tested when diagnosing X-ray images only using GAN and CNN, as shown in Fig. 6. The results obtained proved that the accuracy of diagnosing CT
images is better when using CNN while the accuracy of diagnosing X-ray images is better when using GAN.

7.2 Experiment Two: Evaluate the Overall Accuracy of DLBD-COV Model

In this experiment, the overall accuracy of DLBD-COV model is tested using three different classifiers as mention in the previous section. The results obtained are compared to the result obtained from recent deep learning models, namely, DeConNet [3] and ReNet + [6], as shown in Fig. 7.

It is obviously shown from Fig. 7 that the proposed DLBD-COV model achieved a superior accuracy over DeConNet and ReNet +. To perform this comparison, the same threshold values, classifiers and datasets are used for different models.

7.3 Experiment Three: Evaluate the Computational Time of DLBD-COV Model

This experiment main goal is to test the computational time needed for DLBD-COV model for diagnosing COVID-19 images and compare the time taken by DLBD-COV with the time needed for other deep learning models, DeConNet and ReNet +, for diagnosing, as shown in Fig. 8. In addition, three different classifiers are used in testing, namely SVM, NB and RF.

The results obtained proved that using $H_2O$ for parallel processing along with GAN and CNN fasten the diagnosing process compared to other models when using the three classifiers.
In addition, in order to validate that H2O is the optimum choice to be used in DLBD-COV model to deal with large scale data, this experiment has been conducted. A comparison is made, in this experiment, between the results obtained from DLBD-COV when using H2O and Spark, as shown in Table 2. The results obtained showed that H2O proved an improved performance in terms of computational speed over Spark. This is due to the availability of supported deep learning (Neural networks) algorithms in H2O.ai.

**Table 2** The computational time of DLBD-COV Model when using H2O and SPARK

| Dataset type        | No. of maps | DLBD-COV based spark time (s) | DLBD-COV based H2O time (s) |
|---------------------|-------------|-------------------------------|-----------------------------|
| X-Ray               | 512         | 31.8436                       | 13.0986                      |
|                     | 1024        | 35.5862                       | 13.1081                      |
|                     | 2048        | 46.6194                       | 13.1275                      |
| CT images           | 512         | 31.7255                       | 13.0876                      |
|                     | 1024        | 36.1147                       | 13.1024                      |
|                     | 2048        | 42.0057                       | 13.1371                      |
| Both X-Ray and CT images | 512     | 34.264                        | 13.2942                      |
|                     | 1024        | 36.7934                       | 13.3566                      |
|                     | 2048        | 48.424                        | 13.2739                      |
8 Conclusion

WHO organization announced that COVID-19 disease as a pandemic. It leads to thousands of death in short time. Fast and accurate diagnosis of COVID-19 shows a crucial role in its containment. In this context, an H2O’s Deep-Learning-inspired model based on Big Data analytics (DLBD-COV) is proposed in this chapter for early diagnosis of COVID-19 cases. The DLBD-COV model supports two types of COVID-19’s lung-region images: CT and X-ray images. H2O is used for scalable processing that leads to speed the diagnosing process. The Generative Adversarial Network (GAN) had proved to perform better in diagnosing x-ray images, and the Convolutional Neural Network (CNN) are used for CT images. And their classification results are compared using three different classifiers: SVM, NB, RF. A classifier picker is implemented to select from these three classifiers, the most suitable classifier based on the lowest classification error obtained.

Number of experiments, using thousands of real and generated images, were performed to validate DLBD-COV performance. The obtained results proved that using H2O framework speedup the diagnosing process which time is a critical issue in COVID-19 containment. In addition, the DLBD-COV achieves high accuracy, precision and recall compared to other diagnosing model. As a future work, a plan is made to use genetic algorithm (GA) for optimizing the parameters to improve the classifier performance.

References

1. WHO: Coronavirus disease 2019 (COVID-19) Situation Report—66. 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200326-sitrep-66-covid-19.pdf?sfvrsn=81b94e61_2. Accessed 27 Mar 2020
2. WHO: Coronavirus disease 2019 (COVID-19) Situation Report—66. 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200427-sitrep-98-covid-19.pdf?sfvrsn=90323472_4. Accessed 27 Apr 2020
3. Zheng, C., Deng, X., Fu, Q., & Zhou, Q.: Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label, pp. 1–13 (2020)
4. Hou, H., Lv, W., Tao, Q., Hospital, T., Company, J.T., Ai, T., Hospital, T., Wuhan, T., Hospital, T. (2019). Press In Pr. (2019)
5. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L.: Correlation of chest CT and RT-PCR testing in Coronavirus Disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 2019, 200642 (2020). https://doi.org/10.1148/radiol.2020200642
6. Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Chen, Y., Su, J., Lang, G., Li, Y., Zhao, H., Xu, K., Ruan, L., Wu, W.: Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia, pp. 1–29 (2020). http://arxiv.org/abs/2002.09334
7. Corman, V.M., Landt, O., Kaiser, M., et al.: Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro. Surveill. 25(3) (2020). https://doi.org/10.2807/1560-7917.es.2020.25.3.2000045
8. Lei, J., Li, J., Li, X., Qi, X.: CT Imaging of the 2019 Novel Coronavirus (2019-nCoV) Pneumonia. Radiology 200236 (2020)
9. ELGhamrawy, S.M.: Diagnosis and Prediction Model for COVID19 Patients Response to Treatment based on Convolutional Neural Networks and Whale Optimization Algorithm Using CT Images. medRxiv. 2020 Jan 1

10. Khalifa, N.E., Taha, M.H., Hassanien, A.E., Elghamrawy, S.: Detection of Coronavirus (COVID-19) Associated Pneumonia based on Generative Adversarial Networks and a Fine-Tuned Deep Transfer Learning Model using Chest X-ray Dataset. arXiv preprint arXiv:2004.01184. 2020 Apr 2

11. https://www.h2o.ai/blog/h2o-architecture/

12. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv preprint arXiv:1511.00561 (2015)

13. Goodfellow, J., Pouget-Abadie, M., Mirza, B., Xu, D., Warde-Farley, S., Ozair, Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)

14. http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science.html. Accessed 27 Apr 2020

15. Almeida, F.: Benefits, Challenges and Tools of Big Data Management, pp. 12–20 (2017)

16. Abdel-Hamid, N.B., ElGhamrawy, S., Desouky, A.E. et al.: A dynamic spark-based classification framework for Imbalanced big data. J Grid Computing 16, 607–626 (2018). https://doi.org/10.1007/s10723-018-9465-z

17. Ed-daoudy, A., Maalmi, K.: Application of machine learning model on streaming health data event in real-time to predict health status using spark. In: 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT) (2018)

18. Tripathi, R.; Kumari, V.; Patel, S.; Singh, Y.; Varadwaj, P.: Prediction of IncRNA using deep learning approach. In: International Conference on Advances in Biotechnology (BioTech). Proceedings, pp. 138–142. Global Science and Technology Forum, Singapore (2015)

19. Candel, A., Parmar, V., LeDell, E., Arora, A.: Deep learning with h2o (2015)

20. Mehmood, R., Alam, F., Albogami, N.N., Katib, I., Albeshri, A., Altowaijri, S.M.: Utilearn: a personalised ubiquitous teaching and learning system for smart societies. IEEE Access 5, 2615–2635 (2017)

21. Dey, N., Hassanien, A.E., Bhatt, C., Ashour, A., Satapathy, S.C. (eds.): (2018). Internet of Things and Big Data Analytics Toward Next-Generation Intelligence, pp. 3–549. Springer, Berlin, For COVID-19

22. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Comput. 18(7), 1527–1554 (2006)

23. Bengio, Y.: Learning deep architectures for AI. Foundations Trends® Mach. Learn. 2(1), 1–127 (2009)

24. Lan, K., Wang, D.T., Fong, S., Liu, L.S., Wong, K.K., Dey, N.: A survey of data mining and deep learning in bioinformatics. J. Med. Syst. 42(8), 139 (2018)

25. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

26. Radford, A., Metz, L., Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434. 19 Nov 2015

27. Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R.: Deconvolutional networks. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 13 June 2010, pp. 2528–2535, IEEE

28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

29. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556 (2014)

30. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
31. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
32. Radford, A., Metz, L., Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434 (2015)
33. Mirza, M., Osindero, S.: Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784 (2014)
34. Wu, J.T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 395(10225), 689–697 (2020)
35. Shinde, G.R., Kalamkar, A.B., Mahalle, P.N., Dey, N., Chaki, J., Hassainian, A.: Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art. TechRxiv (2020). Preprint. https://doi.org/10.36227/techrxiv, 12101547, v1
36. Zhao, X., Liu, X., Li, X.: Tracking the Spread of Novel Coronavirus (2019-nCoV) Based on Big Data. medRxiv (2020)
37. Wang, C.J., Ng, C.Y., Brook, R.H.: Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA (2020)
38. Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Z., Wang, J., Yuan, W., Zhu, Y., Song, C.: COVID-19: challenges to GIS with big data. Geogr. Sustain. (2020)
39. Long, J.B., Ehrenfeld, J.M.: The Role of Augmented Intelligence (AI) in Detecting and Preventing the Spread of Novel Coronavirus (2020)
40. Lazer, D., Kennedy, R.: What We Can Learn from the Epic Failure of Google Flu Trends: WIRED (2020). https://www.wired.com/2015/10/canlearn-epic-failure-google-flu-trends/. Published 2015. Accessed 31 Jan
41. Niller, E.: An AI Epidemiologist Sent the First Warnings of the Wuhan Virus: WIRED (2020). https://www.wired.com/story/aiepidemiologist-wuhan-public-health-warnings/. Published 2020. Accessed 31 Jan
42. Salehinejad, H., Valaee, S., Dowdell, T., Colak, E., Barfett, J.: Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 15 Apr 2018, pp. 990–994, IEEE
43. Chen, C., Dou, Q., Chen, H., Heng, P.A.: Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation. In: International Workshop on Machine Learning in Medical Imaging, 16 Sept 2018, pp. 143–151. Springer, Cham
44. Madani, A., Moradi, M., Karargyris, A., Syeda-Mahmood, T.: Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 4 Apr 2018, pp. 1038–1042, IEEE
45. https://github.com/ieee8023/covid-chestxray-dataset. Accessed 27 Apr 2020
46. https://github.com/UCSD-AI4H/COVID-CT. Accessed 27 Apr 2020
47. https://www.sirm.org/en. Accessed 27 Apr 2020
48. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. Accessed 27 Apr 2020
49. Kermany, D., Zhang, K., Goldbaum, M. (2018). Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images. Mendeley Data, v3. http://dx.doi.org/10.17632/ rscbjhr9sj.3
50. https://www.kaggle.com/bachrr/covid-chest-xray. Accessed 27 Apr 2020
51. Elghamrawy, S.M., Hassanien, A.E.: A partitioning framework for Cassandra NoSQL database using Rendezvous hashing. J. Supercomput. 73, 4444–4465 (2017). https://doi.org/10.1007/s11227-017-2027-5