

An Application Review of Artificial Intelligence in Prevention and Cure of COVID-19 Pandemic

Peipeng Yu¹, Zhihua Xia^{1,*}, Jianwei Fei¹ and Sunil Kumar Jha^{1,2}

Abstract: Coronaviruses are a well-known family of viruses that can infect humans or animals. Recently, the new coronavirus (COVID-19) has spread worldwide. All countries in the world are working hard to control the coronavirus disease. However, many countries are faced with a lack of medical equipment and an insufficient number of medical personnel because of the limitations of the medical system, which leads to the mass spread of diseases. As a powerful tool, artificial intelligence (AI) has been successfully applied to solve various complex problems ranging from big data analysis to computer vision. In the process of epidemic control, many algorithms are proposed to solve problems in various fields of medical treatment, which is able to reduce the workload of the medical system. Due to excellent learning ability, AI has played an important role in drug development, epidemic forecast, and clinical diagnosis. This research provides a comprehensive overview of relevant research on AI during the outbreak and helps to develop new and more powerful methods to deal with the current pandemic.

Keywords: Artificial intelligence, COVID-19, medical applications, clinical diagnosis.

1 Introduction

The COVID-19 disease has infected over 3,000,000 people worldwide by 27th April 2020. Apart from China, other countries or regions including South Korea, America, and Europe have reported a rapid increase in the number of COVID-19 cases, implying that this novel coronavirus has posed a global health threat. Under such circumstances, it is necessary to have a comprehensive understanding of the virus that caused the epidemic. At the same time, corresponding measures need to be taken to prevent the virus from spreading continuously. Since the outbreak of COVID-19, the epidemic has spread on a large scale. The most serious problem is that there are not enough doctors and equipment to treat patients, resulting in many patients having to isolate themselves. Moreover, due to the lack of early predictions, it is hard to make effective decisions to control the spread of

¹ Engineering Research Center of Digital Forensics, Ministry of Education, School of Computer and Software, Jiangsu Engineering Center of Network Monitoring, Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing, 210044, China.

² IT Fundamentals and Education Technologies Applications, University of Information Technology and Management in Rzeszow, Rzeszow Voivodeship, 100031, Poland.

* Corresponding Author: Zhihua Xia. Email: xia_zhihua@163.com.

Received: 06 May 2020; Accepted: 15 June 2020.

the epidemic. Traditional drug development methods perform bad in treating COVID-19 and require a lot of time for research. Conditions are extremely severe for the current COVID-19 epidemic.

Artificial intelligence (AI) is a novel technology developed in recent years. It can be used to learn latent patterns in data to accomplish complicated tasks of classification and prediction. AI-based products have been widely used in people's lives ranging from natural language processing, speech recognition, image classification, and data prediction. In the medical field, AI has also been integrated into existing healthcare systems [Szolovits (2019)], such as drug development, medical image recognition, and clinical diagnosis, etc.

So far, researchers have proposed a variety of AI-based complementary medical technologies that can provide doctors with some advice in disease diagnosis [Das, Biswas and Paul (2018)], thereby solving the problem of medical equipment and manpower shortage. At the same time, research related to drug development is also underway. The drugs developed are undergoing further clinical trials and will play a role in future patient treatment. Prediction research on the development of the epidemic has also developed, which provides guidance for policy making. Judging from the current research, AI is able to keep the spread of the virus within reasonable limits.

This article reviews research reports about the prevention and control of epidemics published after the outbreak of COVID-19. Techniques for predicting virus source using AI are introduced in Section 2. In Section 3, the technologies of applying AI to predict the development of the outbreak are introduced. Technologies for drug development using AI and their comparison has been discussed in Section 4. The diagnostic techniques based on AI are presented in Section 5. Finally, the trend of smart medical research and the challenges are discussed in Section 6.

2 Virus source prediction

The outbreak of interpersonal pneumonia caused by COVID-19 poses a serious threat to human health worldwide. It is significant to find out the source of this virus for further prevention and control. After the COVID-19 outbreak, there have been several studies related to the source of this virus. This section highlights some methods related to AI.

Randhawa et al. [Randhawa, Soltysiak, Roz et al. (2020)] used MLDSP and MLDSP-GUI [Randhawa, Hill and Kari (2019)] to classify COVID-19, which is a method of genome analysis using machine learning (ML) and digital signal processing (DSP). The strategy uses more than 5,000 genomes and tests the association at the taxonomic level of the domain and species. Each genomic sequence is mapped into its genomic signal by using a digital representation called Chaos Game Representation (CGR). Feature vectors are constructed from the digital representation and used as input to train six supervised learning-based classification models. In this study, taxonomic classification of the current COVID-19 virus has been confirmed by the two-dimensional genomic feature representation method, which can achieve alignment-free comparative genomics through a machine learning based method. This study proves that an alignment-free method can also accurately predict the classification of new sequences.

Meanwhile, Zhu et al. [Zhu, Guo, Li et al. (2020)] introduced the VHP (virus-host prediction) to predict the potential of infecting humans of COVID-19. To construct the VHP model, as shown in Fig. 1, a Bi-path Convolutional Neural Networks (BiPath-CNN) [Fang, Tan, Wu et al. (2019)] is used, where each viral sequence is represented by a one-hot matrix of its bases and codon separately. In this study, VHP outputs five scores for each host type by inputting viral nucleotide sequences, reflecting the infectivity within each host type. The area under the curve (AUC) of VHP is calculated to evaluate the performance of the VHP, which shows that the VHP has a higher AUC and can make accurate predictions of viral hosts. Based on the prediction, bat coronavirus is thought to have a more similar infection pattern with COVID-19. VHP could play a significant role in public health service and provided strong assistance for taking precautions of novel viruses that have the potential to infect humans.

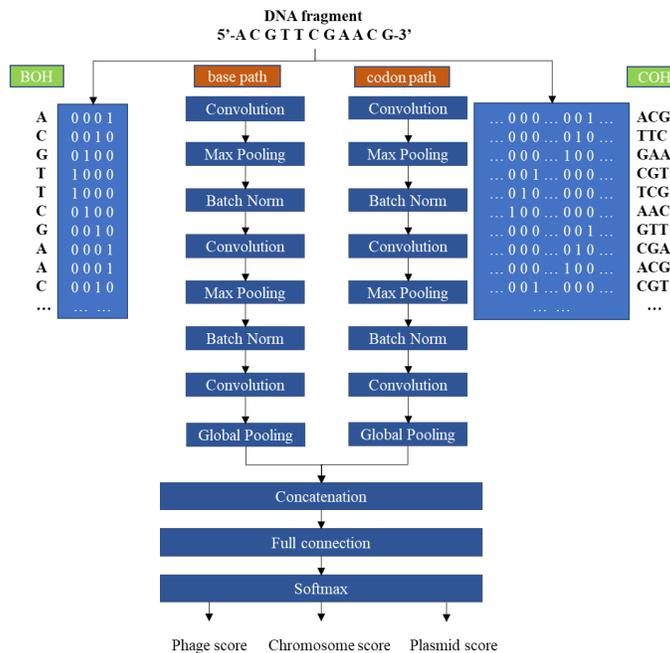


Figure 1: Structure of BiPath-CNN [Zhu, Guo, Li et al. (2020)]

3 Epidemic development forecasting

COVID-19 is a rapidly spreading epidemic that threatens people’s lives. In response to this ongoing public health emergency, Dong et al. [Dong, Du and Gardner (2020)] developed an online dashboard to visualize and track the reported COVID-19 in real-time. The dashboard was first shared publicly on January 22, 2020, showing the location and number of confirmed COVID-19 cases, deaths, and rehabilitations in all affected countries (Fig. 2). Meanwhile, Kastner et al. [Kastner, Wei and Samet (2020)] developed an application that allows users to explore geographic spread through analysis of keyword prevalence in geotagged news articles. To some extent, these tools can provide the public with a user-friendly tool to track the progress of the outbreak. Accurate

prediction of the spread of epidemics is an extremely important and arduous task. Several studies have been published in the last two months.

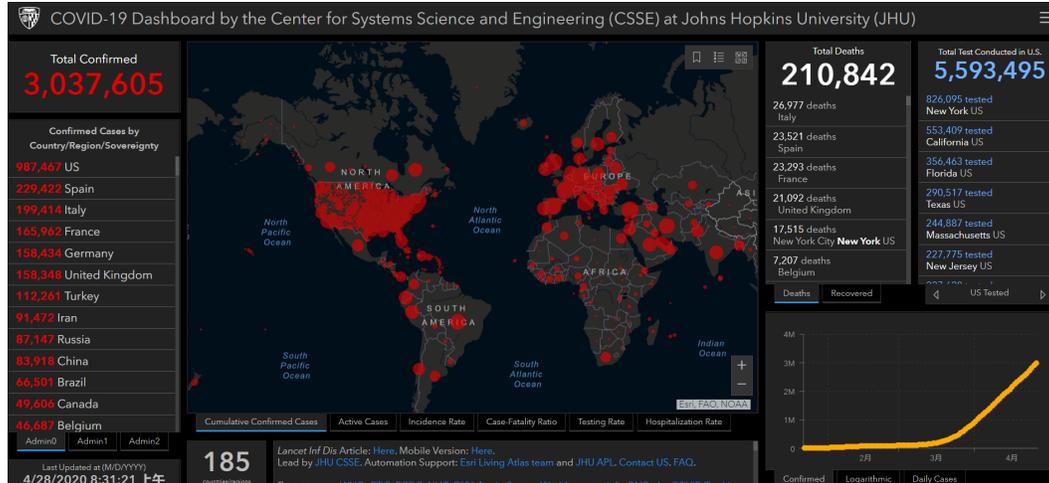


Figure 2: Online dashboard to visualize the reported COVID-19 [Dong, Du and Gardner (2020)]

3.1 Prediction from small dataset

The lack of understanding of new diseases and the high level of uncertainty has contributed to the widespread of the virus. At the beginning of an epidemic, data samples are often scarce. Finding a suitable prediction model with few training samples is a huge challenge for machine learning. Fong et al. [Fong, Li, Dey et al. (2020)] proposed a type of evolutionary neural network called polynomial neural network with corrective feedback (PNN+cf) to predict the outbreak of Wuhan-originated coronavirus. PNN+cf allows additional input variables during the formation of polynomials. To improve accuracy, only statistically significant or so-called relevant additional information should be added. Two additional pieces of information are added, namely lagged data and training errors from past iterations of model training which are known as correction feedback. The feedback is a first-order lag, which looks like a shadow of time series and residuals. The motivation for using residuals (the deviations between the predicted and actual values) is to compensate for the bias that exists in the time series. Experiments show that PNN+cf is superior in yielding a forecasting model with a relatively lowest error.

3.2 Prediction based on auto-encoder

Although existing epidemiological models can be used to estimate the dynamics of transmission, such models require parameters and depend on many assumptions. Meanwhile, the parameters in these models need to be estimated using actual data. At the time of the outbreak, estimated parameters using real-time data are not readily available. To solve this problem, Hu et al. [Hu, Ge, Jin et al. (2020)] developed a modified stacked auto-encoder (MAE) for modeling the transmission dynamics of the COVID-19 in China. By using clustering algorithms, the latent variables in the auto-encoder are used to group

the provinces/cities for investigating the transmission structure. Unlike the classical auto-encoder where the number of nodes in the layers, usually decreases from the input layer to the latent layers. The numbers of the nodes in the input, the first latent layer, the second latent layer, and output layers in the MAE are 8, 32, 4, and 1, respectively. Finally, the trained MAE is applied to predict the future number of confirmed cases of COVID-19 in each province/city. If the data are reliable and there are no second transmissions, this method will be a powerful tool to help public health planning and decision-making.

4 Drug research

COVID-2019 disease has caused a major health threat around the world. Under such circumstances, the drug development for COVID-19 is very important. Due to strong learning ability, AI is widely used in the process of drug development [Fleming (2018)]. Qiao et al. [Qiao, Tran, Shan et al. (2020)] proposed a machine learning model on the immunopeptide panel and then predicted HLA peptides from conserved regions of a virus that are most likely to trigger responses from the person's T cells. Meanwhile, Researchers have proposed a variety of AI-based drug development methods after the outbreak of COVID-19.

4.1 Drug repurposing

A large number of clinical trials are required before the drug is put into use, which takes a long time [Elwood (2017)]. In this situation, the screening of old drugs for the COVID-2019 epidemic can help to understand the toxicity characteristics of the virus, which can promote the development of new drugs. Beck et al. [Beck, Shin, Choi et al. (2020)] proposed a pre-trained deep learning-based drug-target interaction model called molecular transformer-drug target interaction (MT-DTI). It can be used as recognition for drugs that act on viral proteins of COVID-19. The natural language processing based on bidirectional encoder representations from transformers (BERT) [Tenney, Das and Pavlick (2019)] framework is the core algorithm of the model. It shows good performance and robust results in diverse drug-target interaction datasets. Simultaneously, Li et al. [Li, Yu, Zhang et al. (2020)] proposed a novel network drug repurposing platform to identify drugs that may be used to treat COVID-19. The genomic sequence of COVID-19 is first analyzed and the AutoSeed pipeline is used to obtain 34 genes related to COVID-19. The obtained genes are then used as seeds to automatically build a molecular network. The network construction is automatically performed by AutoNet [Shang, Zhu, Shen et al. (2018)], which is an algorithm originally developed for the drug discovery cloud platform. Unlike structure-based reuse methods that rely on several known targets, this method can prioritize potential drug targets and existing drugs at a systemic level to deal with this global infectious disease threat. It is hoped that these results will help in the rapid design and implementation of clinical trials that can treat COVID-19.

Finding a way to inhibit the sequence of antibodies to the COVID-19 virus epitope will save thousands of lives. Magar et al. [Magar, Yadav and Farimani (2020)] developed a machine learning model for high throughput screening of synthetic antibodies to discover antibodies that potentially inhibit the COVID-19. Using this model, 18 antibodies are

found to be very effective. Using molecular dynamics simulation [Rapaport and Rapaport (2004)], the stability of the predicted antibody is checked, and 8 stable antibodies that could neutralize COVID-19 are found.

Given the clinical efficacy of traditional Chinese medicine (TCM) in many diseases (even severe acute respiratory syndrome (SARS)), it is considered a promising complementary therapy for the disease. Wang et al. [Wang, Li, Yan et al. (2020)] used an artificial neural network based method to evaluate the TCM prescriptions that are officially recommended in China for novel coronavirus (COVID-19). The ANN model is built to understand the potential relationship between the structure of TCM prescriptions and the possibility of SE. Finally, seven Chinese medicine prescriptions perform well in this study, and it is recommended that these prescriptions should be considered first when treating COVID-19. Meanwhile, coronavirus main protease is also considered as the main therapeutic target, so Zhang et al. [Zhang, Saravanan, Yang et al. (2020)] focused on drug screening based on the modeled COVID-19 main protease structure. The proposed deep learning based method named Deeper-Feature Convolutional Neural Network (DFCNN) can identify protein-ligand interactions with relatively high accuracy. DFCNN can perform virtual screening quickly because no docking or molecular dynamics simulation is required. DFCNN is a dense and fully connected neural network (similar to DenseNet [Iandola, Moskewicz, Karayev et al. (2014)]), but replace the convolutional layer with a fully connected layer). The deeper layers enable the model to learn more abstract features from the data. Compared with many other methods, DFCNN is independent of the docking simulation, and the training dataset includes non-binding decoys. Independent docking simulation makes it very fast, and the inclusion of non-binding decoys during training makes the model more powerful in practical application scenarios.

4.2 Drug generation

Researchers found that the COVID-19 protease has a 96.1% sequence identity with severe acute respiratory syndrome virus (SARS-CoV) [Rapaport and Rapaport (2004)], which means that all potential anti-SARS-CoV chemotherapy drugs are also potential COVID-19 drugs. Consequently, Gao et al. [Gao, Nguyen, Wang et al. (2020)] reported a family of possible 2019-nCoV drugs generated by the generative network complex (GNC) [Grow, Gao, Nguyen et al. (2019)] based on machine intelligence. As shown in Fig. 3, the first component is a generative network, which uses a SMILES string as the input to generate a new molecule. After that, the molecule is inputted into the second component of GNC, namely a deep neural network based on two-dimensional fingerprint (2DFP-DNN) [Grow, Gao, Nguyen et al. (2019)] to reassess its medicinal properties. The next component is the MathPose model, which is used to predict the three-dimensional structural information of the compounds selected by 2DFP-DNN. The biological activity of these compounds is further estimated by a structure-based deep learning model called MathDL. The pharmacological properties predicted by the last part of GNC are used as indicators for selecting promising drug candidates. Finally, Compared with HIV inhibitors, new compounds produced by GNC appear to have better pharmacological properties to treat COVID-19.

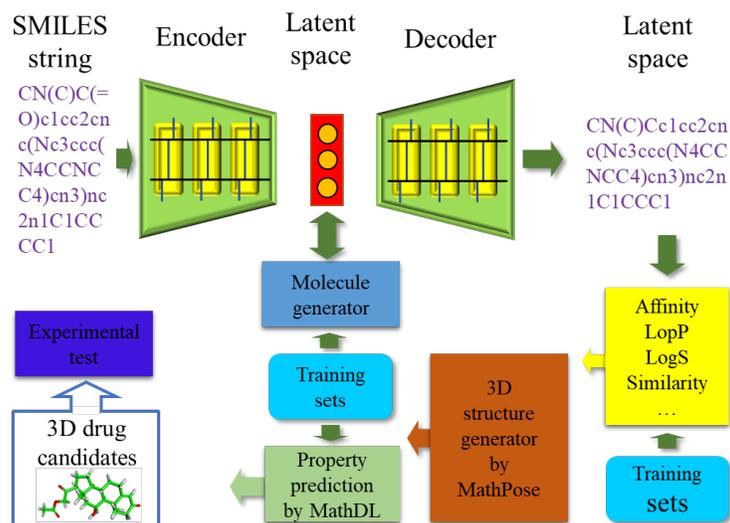


Figure 3: A schematic illustration of the generative network complex [Gao, Nguyen, Wang et al. (2020)]

One of the most important protein targets of COVID-19 is the 3C-like protease, whose crystal structure is known. Subsequently, Zhavoronkov et al. [Zhavoronkov, Aladinskiy, Zhebrak et al. (2020)] used generative approaches to explore potential COVID-2019 3C-like Protease Inhibitors. The drug discovery system consists of three main pipelines: target discovery, small molecule drug discovery, and prediction of clinical trial results. The purpose of this system is to automate the drug discovery process for a variety of human diseases. The small molecule drug discovery pipeline can be used to produce inhibitors of bacterial and viral protein targets. During the generative phase, different ML methods are used by inputting various molecular representations, including string representations and graphics. Each model is optimizing the reward function to explore the chemical space, exploit promising clusters, and generate high-scoring molecules. Results show that the new method has good performance in both cost-effectiveness and time efficiency.

5 Clinical diagnosis

After the outbreak of the new coronavirus (COVID-19), special attention is needed due to its future growth and possible global threat. In addition to clinical procedures and treatments, several algorithms based on machine learning (ML) are used to analyze data and decision-making processes [Sammut and Webb (2017)]. It means that AI-driven tools can help to diagnose the COVID-19.

5.1 Detection based on abnormal respiratory patterns

According to the latest clinical studies, COVID-19 patients breathe differently than ordinary patients with respiratory diseases. Shortness of breath is an important symptom for COVID-19. Therefore, Wang et al. [Wang, Hu, Li et al. (2020)] proposed a COVID-19 diagnostic method based on respiratory features. The network architecture is shown in

Fig. 4. It consists of the input layer, BI-GRU layer [Wang, Xu, Zhou et al. (2018)], attention layer, and output layer. The input layer is used to input simulation data (training phase) or depth data (testing phase) at each point of the respiratory waveform. Bi-directional and attention are added in GRU, which corresponds to the attention layer in the BI-GRU layer and BI-AT-GRU, respectively. Finally, the obtained BI-AT-GRU specific to respiratory pattern classification yields excellent performance and outperforms the existing state-of-the-art models. This research can be applied to the diagnosis of COVID-19 patients to reduce the detection workload of the current medical system.

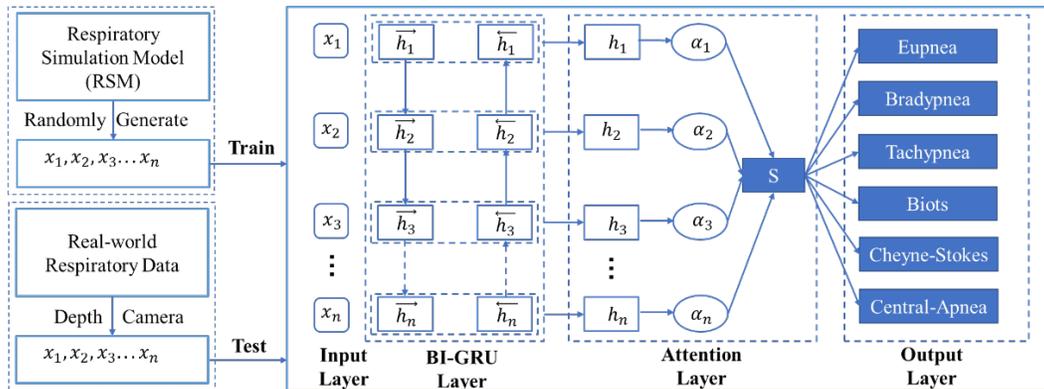


Figure 4: BI-AT-GRU for classification [Wang, Hu, Li et al. (2020)]

5.2 COVID-19 detection using smartphones

There are many mechanisms for detecting COVID-19, including clinical analysis of chest CT images and blood test results. However, such equipment is expensive, and it takes time to install and use them. Rao et al. [Rao and Vazquez (2020)] proposed a new machine learning algorithm to improve possible quicker case identifications of COVID-19 by using a mobile phone-based web survey. It will also help to check the spread of susceptible populations. Meanwhile, Maghdid et al. [Maghdid, Ghafoor, Sadiq et al. (2020)] proposed a new framework for detecting COVID-19 using smartphone sensors. The proposal provides a low-cost solution because most radiologists already have smartphones that can be used for different purposes in daily life. Today's smartphones are powerful and contain a large number of sensors, including temperature sensors, inertial sensors, color sensors, and humidity sensors. As shown in Fig. 5, the framework designed to support AI can read signal measurements from smartphone sensors [Fogel and Kvedar (2018)], which will help to predict the severity of pneumonia and the outcome of the disease.

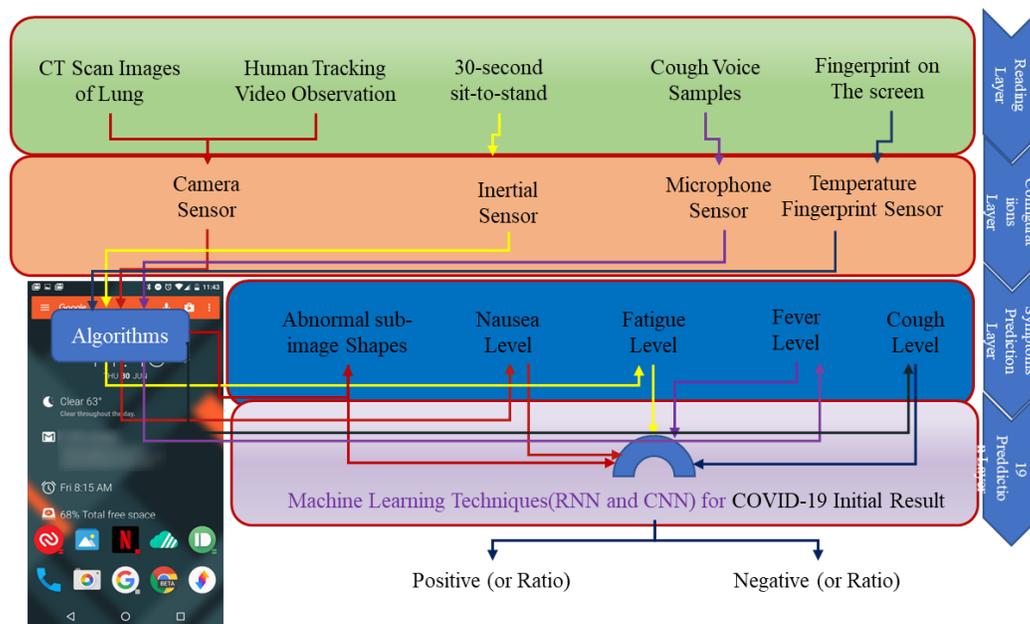


Figure 5: General diagram of the framework for predicting disease COVID-19 [Rao and Vazquez (2020)]

5.3 Using CT images to screen for COVID-19

Radiography is also the main diagnostic tool of COVID-19. Although typical CT images are helpful for early screening of suspicious cases, images of various viral pneumonia are very similar and overlap with other infectious lung diseases. Therefore, it is difficult for radiologists to distinguish COVID-19 from other viral pneumonia.

5.3.1 Infection classification

Currently, detection using CT images is still an image classification task. Research on medical images has been carried out for many years. This paper has provided a summary of typical approaches in Tab. 1. To distinguish COVID -19, Narin et al. [Narin, Kaya and Pamuk (2020)] built a deep convolutional neural network (CNN) based ResNet50, InceptionV3, and Inception-ResNetV2 models to classify COVID-19 chest X-ray images into normal and COVID-19 categories. The problems of insufficient data and training time are overcome by implementing transfer learning techniques using ImageNet data. Likewise, Wang et al. [Wang, Kang, Ma et al. (2020)] proposed a new diagnostic algorithm based on the radiographic changes in CT images. The results demonstrate that the extraction of radiographic features using deep learning methods is of great value for COVID-19 diagnosis.

Aiming to establish a model for early screening of COVID-19 pneumonia, Xu et al. [Xu, Jiang, Ma et al. (2020)] used multiple CNN models to classify the CT image dataset and calculated the infection probability of COVID-19. Fig. 6 shows the entire process of generating the COVID-19 diagnostic report. First, images are pre-processed to extract

effective lung regions. Secondly, the 3D CNN [Kamnitsas, Ledig, Newcombe et al. (2017)] model is used to segment multiple candidate image cubes. Thirdly, the classification model is used to classify all image cubes. Image patches from the same cube vote for the overall type of candidate. Finally, a Noisy-or Bayesian function [Oniško, Druzdel and Wasyluk (2001)] is used to calculate the overall analysis report of the CT sample.

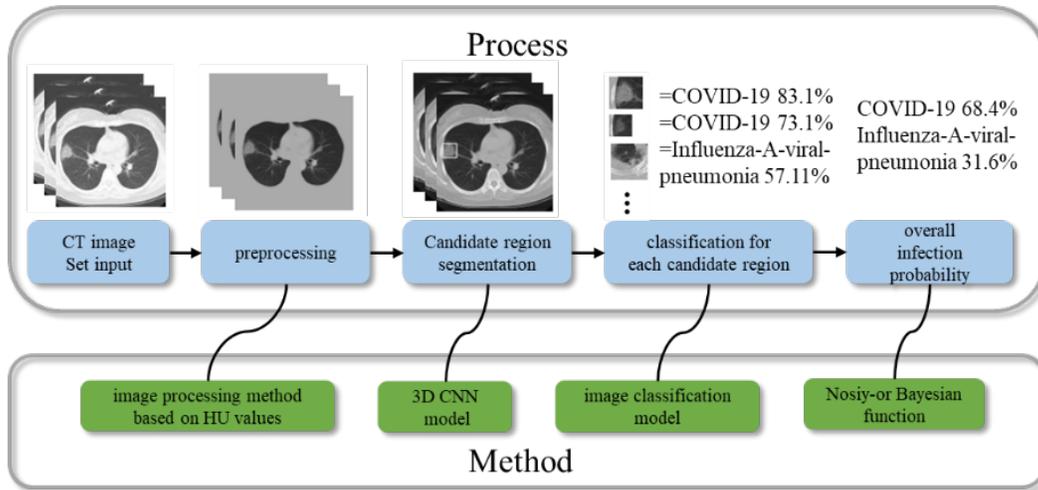


Figure 6: Process of generating diagnostic reports [Xu, Jiang, Ma et al. (2020)]

In early studies, patients were found to have abnormalities on chest radiography images. Barstugan et al. [Barstugan, Ozkaya and Ozturk (2020)] proposed an early detection of coronavirus (COVID-19) by machine learning methods. To detect COVID-19 infection, various feature extraction methods are used to improve classification performance. Thereafter, SVM is employed to classify the extracted features. The best classification accuracy of 99.68% is obtained, which shows the huge potential of the algorithm in infection detection. Wang et al. [Wang and Wong (2020)] introduced a deep convolutional neural network called COVID-Net, which is specifically designed to detect COVID-19 cases from chest X-ray images. The COVID-Net network architecture makes extensive use of lightweight residual projection-expansion-projection-extension (PEPX) design patterns. Besides, how to predict COVID-Net in an interpretable manner is also the focus of the study. Interpretability can deepen the understanding of the key factors related to COVID cases, which can help clinicians to perform better screening.

In addition to detection tasks, the diagnostic uncertainty in the report is a challenging but inevitable task for radiologists. Based on the publicly available COVID-19 chest X-ray dataset, Ghoshal et al. [Ghoshal and Tucker (2020)] studied the dropweights-based Bayesian convolutional neural network (BCNN) to estimate the uncertainty in deep learning solutions. It improves human-machine integration and prove that the accuracy of uncertainty prediction is closely related to the accuracy of prediction. It is believed that the emergence of uncertain deep learning solutions will lead to wider clinical applications of AI.

Table 1: Infection classification algorithms

Methods	Classifiers	Dealing with	Datasets
[Narin, Kaya and Pamuk (2020)]	transfer-learning model	chest radiography images	Online
[Wang, Kang, Ma et al. (2020)]	transfer-learning model	CT images	Collected
[Xu, Jiang, Ma et al. (2020)]	multiple CNN models	CT images	Collected
[Barstugan, Ozkaya and Ozturk (2020)]	SVM	CT images	Collected
[Wang and Wong (2020)]	COVID-Net	chest radiography images	COVIDx

5.3.2 Segment the main areas of pneumonia

Unlike simple image classification tasks, image segmentation can accurately locate the main areas of pneumonia and help doctors to carry out the targeted treatment. Chen et al. [Chen, Wu, Zhang et al.] proposed a novel and powerful network architecture called UNet++ for medical image segmentation, which accelerates the diagnosis and isolation time of COVID-19 patients, thus helping to control the epidemic. Likewise, Gaál et al. [Gaál, Maga and Lukács (2020)] proposed to use the state-of-the-art fully convolutional neural network in combination with an adversarial model to produce an accurate organ segmentation mask on chest X-rays. Song et al. [Song, Zheng, Li et al. (2020)] developed a novel deep learning based CT diagnostic system called DeepPneumonia to help doctors to detect COVID-19 and identify the main areas of pneumonia. The fully automated lung CT diagnosis system includes three main steps. First, the main area of the lung is extracted to avoid noise caused by different lung contours. Then, a detailed relation extraction neural network (DRE-Net) is designed to extract the top-K details from CT images to obtain image-level predictions. The built model is based on pre-trained ResNet-50, on which the feature pyramid network (FPN) [Lin, Dollár, Girshick et al. (2017)] is added to extract the top-K details from each image. The attention module is used to understand the importance of every detail. By using FPN and attention modules, the proposed model can not only detect the most important part of the image but also interpret the output through the neural network. Finally, predictions are summarized to achieve precise diagnoses.

To develop a deep learning based system for automatically segmenting infected areas in chest CT scans, Shan et al. [Shan, Gao, Wang et al. (2020)] employed the VB-Net neural network to segment the COVID-19 infection area in CT scans. VB-Net is an improved 3D convolutional neural network that combines V-Net [Milletari, Navab and Ahmadi (2016)] with a bottleneck structure. VB-Net contains two paths. The first is the compression path, which includes down-sampling and convolution operations to extract

global image features. The second is the expansion path, which includes up-sampling and convolution operations to integrate fine-grained image features. Compared with the previous V-Net, VB-Net is faster because the bottleneck structure has been integrated into VB-Net. By reducing and combining feature mapping channels, not only the model size and inference time are greatly reduced, but also the cross-channel features are effectively fused through a convolution, which makes the VB-Net more suitable for processing large 3D data than traditional V-Net.

Enough labeled data are not available during the outbreak of COVID-19, which makes the training of supervised algorithms a problem. Zheng et al. [Zheng, Deng, Fu et al. (2020)] developed a deep learning software system based on weak learning. For each case, the lung region is segmented using a pre-trained UNet [Li, Chen, Qi et al. (2018)]. Then the segmented 3D lung region is inputted into a 3D deep neural network (DeCoVNet) to predict the possibility of COVID-19 infection. As shown in Fig. 7, DeCoVNet uses the CT volume and its 3D lung mask as input and consists of a stem, two 3D ResBlocks and a classifier. The weakly-supervised deep learning model can accurately predict the probability of COVID-19 infection in chest CT volume without requiring extensive manual annotations training. The easy-to-train high-performance deep learning algorithm provides a method for quick identification of COVID-19 patients, which is highly beneficial for controlling the outbreak.

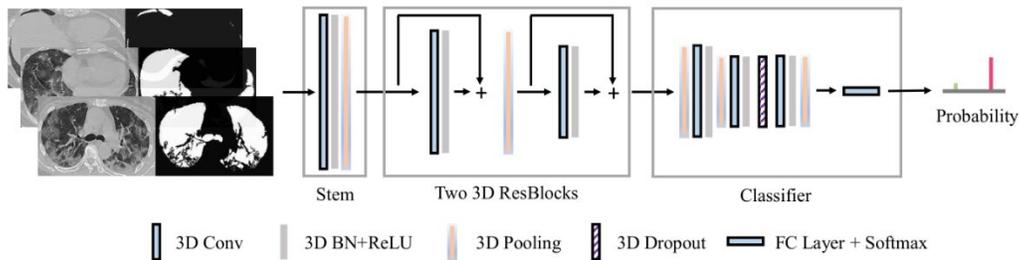


Figure 7: Architecture of the proposed DeCoVNet [Zheng, Deng, Fu et al. (2020)]

5.4 Clinical severity prediction

COVID-19 has spread to every inhabited continent. Given the large increase in the number of cases, there is an urgent need to enhance clinical skills to identify a few cases from the large number of mild cases that will develop into severe cases. Jiang et al. [Jiang, Coffee, Bari et al. (2020)] developed a tool with AI capabilities that can predict patients who may be at risk of more severe illness. In this study, machine learning algorithms are used to extract rules from experience to identify features with the most accurate predictive power. Feature engineering is used to extract features that play an important role in the feature set, which can be used to obtain higher predictability. Information gain, Gini coefficient, and chi-square statistics are used to measure the importance of features. Finally, a total of 11 features are selected as highly predictive features. By using decision trees, random forests, and support vector machines for classification, the algorithm shows good detection performance, demonstrating the importance of these 11 features in disease detection. Similarly, Yan et al. [Yan, Zhang,

Goncalves et al. (2020)] used the XGBoost machine learning method to build a predictive model for the early identification of critically ill patients. As shown in Fig. 8, after preprocessing and segmenting the data, the selected features are classified according to the importance of clinical features using the XGBoost model. Using this model, three key clinical features are extracted from more than 300 features. The extracted features (LDH, Hs-CRP, and Lymphocytes) can accurately predict the survival rate, and the accuracy exceeds 90%.

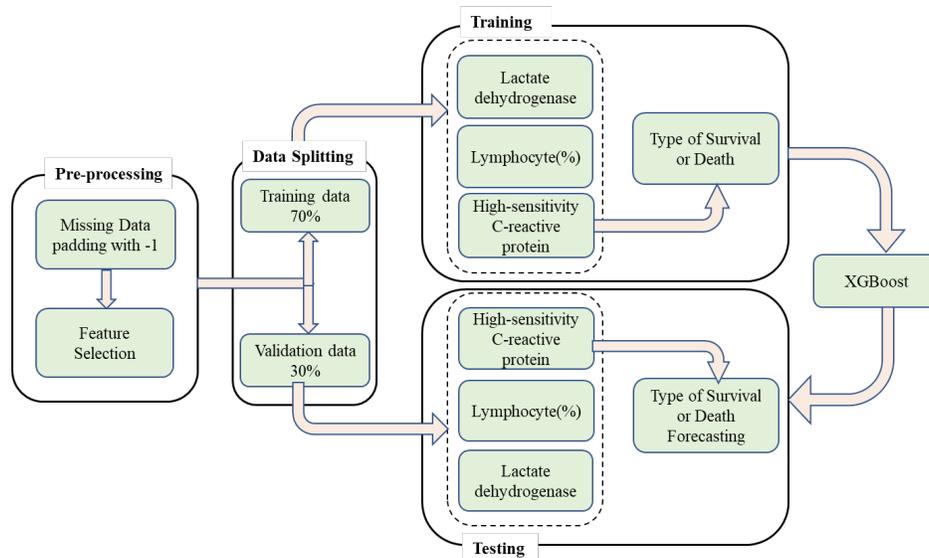


Figure 8: Flowchart of the XGBoost machine learning algorithm [Yan, Zhang, Goncalves et al. (2020)]

6 Discussions and future research directions

COVID-19 is firstly encountered in the Wuhan region in China, posing a threat to public health, trade, and the world economy. The virus exhibits partially similar behavior to other viral pneumonia. As a result, the situation is difficult to control due to the rapid rate of transmission of the virus.

As a powerful tool, AI has been used to solve the problems encountered in current medical systems. It has played an important role in the medical field. However, this study believes that still some issues need to be addressed regarding the application of artificial intelligence algorithms in the medical field.

Firstly, the diagnosis needs to provide sufficient evidence to prove its correctness in the field of traditional medicine. However, AI-based algorithms are black-box models, which means that researchers cannot explore why the models output such results. Even if the model prediction accuracy is high enough, a sufficient interpretable basis must be provided the final judgment. One way to make AI systems more interpretable is to predict different types of medical data and measurement results that characterize a patient's condition while predicting clinical outcomes (disease/health) [Wei, Poirion, Bodini et al. (2019)]. However, the

complexity of the algorithm is greatly increased due to this operation. Therefore, exploring the interpretability of AI is a very meaningful work in the medical field.

Secondly, the lack of available labeled data is a problem for researchers. Despite the abundance of medical data, there is still a lack of available labeled data during the outbreak. The amount of tagged data currently collected is insufficient to support accurate predictions by machine learning models. Therefore, expanding existing datasets or using a small number of samples in model training is a must-select strategy currently. However, due to a lack of sufficient labeled data, the accuracy of supervised algorithms is often not sufficient to support effective predictions. In this condition, weakly supervised models can play a role in such scenarios. It can break the limitation of labeling data while ensuring the accuracy of the algorithm.

Thirdly, the generalization of algorithms is a common problem of AI algorithms. The algorithms listed in the present study can achieve good results on their own datasets, but the effectiveness for realistic data remains to be verified. For now, the generalization of the algorithm can be improved by increasing the amount of data in the training dataset or constraining the model. Generalization is a research direction that is needed to promote and utilize the advances and advantages of AI.

Finally, AI-based algorithms using multi-modal data should be noticed. At present, AI-based algorithms are mostly processed using single modal data. For example, in the clinical diagnosis area, the algorithm usually uses CT images for prediction. However, the clinical data of patients contains multiple categories that are underutilized. By extracting multi-modal features for prediction, a more accurate judgment can be given. Thus, algorithms using multi-modal data are a research direction in the medical field.

7 Conclusion

In the past few months, COVID-19 has spread throughout the world and poses a serious threat to people's health. Intelligent medical technology has played an important role in the fight against COVID-19. Applications in virus source prediction, epidemic development forecasting, drug research, and clinical diagnosis are reviewed in this article, covering the entire application using the AI algorithm during the COVID-19 pandemic. The described algorithms prove the important role of AI in combating epidemics. While substantial progress has been made, there is still much room to improve existing algorithms, especially in terms of interpretability and generalization. Multimodal data is also considered as an effective method that can improve the performance of the algorithm. These technologies promise to further improve our ability to control the epidemic.

Acknowledgement: Thanks Zhentian Zhang and Wenhao Zhou for his help in completing this article. The author is also very grateful to the reviewers.

Funding Statement: This work is supported in part by the Jiangsu Basic Research Programs-Natural Science Foundation under Grant Numbers BK20181407, in part by the National Natural Science Foundation of China under Grant Numbers U1936118, 61672294, in part by Six peak talent project of Jiangsu Province (R2016L13), Qinglan Project of Jiangsu Province, and "333" project of Jiangsu Province, in part by the

National Natural Science Foundation of China under Grant Numbers U1836208, 61702276, 61772283, 61602253, and 61601236, in part by National Key R&D Program of China under Grant 2018YFB1003205, in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund, in part by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China. Zhihua Xia is supported by BK21+ program from the Ministry of Education of Korea.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Barstugan, M.; Ozkaya, U.; Ozturk, S.** (2020): Coronavirus (COVID-19) classification using CT images by machine learning methods. arXiv: 2003.09424.
- Beck, B. R.; Shin, B.; Choi, Y.; Park, S.; Kang, K.** (2020): Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), wuhan, china through a drug-target interaction deep learning model. BioRxiv: 2020.2001.2031.929547.
- Chen, J.; Wu, L.; Zhang, J.; Zhang, L.; Gong, D. et al.** (2020): Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. MedRxiv: 2020.02.25.20021568.
- Das, S.; Biswas, S.; Paul, A.; Dey, A.** (2018): AI doctor: an intelligent approach for medical diagnosis. *Industry Interactive Innovations in Science, Engineering and Technology*, vol. 11, pp. 173-183.
- Dong, E.; Du, H.; Gardner, L.** (2020): An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533-534.
- Elwood, M.** (2017). *Critical Appraisal of Epidemiological Studies and Clinical Trials*. Oxford University Press.
- Fang, Z.; Tan, J.; Wu, S.; Li, M.; Xu, C. et al.** (2019): PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience*, vol. 8, no. 6, giz066.
- Fleming, N.** (2018): How artificial intelligence is changing drug discovery. *Nature*, vol. 557, no. 7706, pp. S55.
- Fogel, A. L.; Kvedar, J. C.** (2018): Artificial intelligence powers digital medicine. *NPJ Digital Medicine*, vol. 1, no. 1, pp. 1-4.
- Fong, S. J.; Li, G.; Dey, N.; Crespo, R. G.; Herrera-Viedma, E.** (2020): Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak. arXiv: 2003.10776.
- Gaál, G.; Maga, B.; Lukács, A.** (2020): Attention U-Net based adversarial architectures for chest X-ray lung segmentation. arXiv: 2003.10304.
- Gao, K.; Nguyen, D. D.; Wang, R.; Wei, G. W.** (2020): Machine intelligence design of 2019-nCoV drugs. BioRxiv: 2020.2001.2030.927889.

- Ghoshal, B.; Tucker, A.** (2020): Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. arXiv: 2003.10769.
- Grow, C.; Gao, K.; Nguyen, D. D.; Wei, G.-W.** (2019): Generative network complex (GNC) for drug discovery. arXiv: 1910.14650.
- Hu, Z.; Ge, Q.; Jin, L.; Xiong, M.** (2020): Artificial intelligence forecasting of COVID-19 in china. arXiv: 2002.07112.
- Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrell, T. et al.** (2014): Densenet: implementing efficient convnet descriptor pyramids. arXiv: 1404.1869.
- Jiang, X.; Coffee, M.; Bari, A.; Wang, J.; Jiang, X. et al.** (2020): Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua*, vol. 63, pp. 537-551.
- Kamnitsas, K.; Ledig, C.; Newcombe, V. F.; Simpson, J. P.; Kane, A. D. et al.** (2017): Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, vol. 36, pp. 61-78.
- Kastner, J.; Wei, H.; Samet, H.** (2020): Viewing the progression of the novel corona virus (COVID-19) with NewsStand. arXiv: 2003.00107.
- Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C. W. et al.** (2018): H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663-2674.
- Li, X.; Yu, J.; Zhang, Z.; Ren, J.; Peluffo, A. E. et al.** (2020): Network bioinformatics analysis provides insight into drug repurposing for COVID-2019. Preprints: 2020030286.
- Lin, T. Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B. et al.** (2017): Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Magar, R.; Yadav, P.; Farimani, A. B.** (2020): Potential neutralizing antibodies discovered for novel corona virus using machine learning. arXiv: 2003.08447.
- Maghdid, H. S.; Ghafoor, K. Z.; Sadiq, A. S.; Curran, K.; Rabie, K.** (2020): A novel ai-enabled framework to diagnose coronavirus COVID 19 using smartphone embedded sensors: Design study. arXiv: 2003.07434.
- Milletari, F.; Navab, N.; Ahmadi, S. A.** (2016): V-net: fully convolutional neural networks for volumetric medical image segmentation. *Fourth International Conference on 3D Vision*.
- Narin, A.; Kaya, C.; Pamuk, Z.** (2020): Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. arXiv: 2003.10849.
- Oniško, A.; Druzdzel, M. J.; Wasyluk, H.** (2001): Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. *International Journal of Approximate Reasoning*, vol. 27, no. 2, pp. 165-182.
- Qiao, R.; Tran, N. H.; Shan, B.; Ghodsi, A.; Li, M.** (2020): Personalized workflow to identify optimal T-cell epitopes for peptide-based vaccines against COVID-19. arXiv: 2003.10650.

Randhawa, G. S.; Hill, K. A.; Kari, L. (2019): MLDSP-GUI: an alignment-free standalone tool with an interactive graphical user interface for DNA sequence comparison and analysis. *BioRxiv*: 745406.

Randhawa, G. S.; Soltysiak, M. P. M.; Roz, H. E.; de Souza, C. P. E.; Hill, K. A. et al. (2020): Machine learning-based analysis of genomes suggests associations between Wuhan 2019-nCoV and bat betacoronaviruses. *BioRxiv*: 2020.2002.2003.932350.

Rao, A. S. S.; Vazquez, J. A. (2020): Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. *Infection Control & Hospital Epidemiology*, pp. 1-18.

Rapaport, D. C.; Rapaport, D. C. R. (2004): *The Art of Molecular Dynamics Simulation*. Cambridge University Press.

Sammut, C.; Webb, G. I. (2017): *Encyclopedia of Machine Learning and Data Mining*. Springer Publishing Company, Incorporated.

Shan, F.; Gao, Y.; Wang, J.; Shi, W.; Shi, N. et al. (2020): Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv*: 2003.04655.

Shang, J.; Zhu, Q.; Shen, J.; Wang, X.; Gu, X. et al. (2018): AutoNet: automated network construction and exploration system from domain-specific corpora. *KDD*.

Song, Y.; Zheng, S.; Li, L.; Zhang, X.; Zhang, X. et al. (2020): Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *MedRxiv*: 2020.2002.2023.20026930.

Szolovits, P. (2019): *Artificial Intelligence in Medicine*. Routledge.

Tenney, I.; Das, D.; Pavlick, E. (2019): Bert rediscovers the classical NLP pipeline. *arXiv*: 1905.05950.

Wang, L.; Wong, A. (2020): COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. *arXiv*: 2003.09871.

Wang, Q.; Xu, C.; Zhou, Y.; Ruan, T.; Gao, D. et al. (2018): An attention-based Bi-GRU-CapsNet model for hypernymy detection between compound entities. *IEEE International Conference on Bioinformatics and Biomedicine*.

Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M. et al. (2020): A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *MedRxiv*: 2020.02.14.20023028.

Wang, Y.; Hu, M.; Li, Q.; Zhang, X. P.; Zhai, G. et al. (2020): Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. *arXiv*: 2002.05534.

Wang, Z.; Li, L.; Yan, J.; Yao, Y. (2020): Evaluating the traditional chinese medicine (TCM) officially recommended in China for COVID-19 using ontology-based side-effect prediction framework (OSPF) and deep learning. *Preprints*: 2020020230.

Wei, W.; Poirion, E.; Bodini, B.; Durrleman, S.; Colliot, O. et al. (2019): Fluid-attenuated inversion recovery MRI synthesis from multisequence MRI using three-

dimensional fully convolutional networks for multiple sclerosis. *Journal of Medical Imaging*, vol. 6, no. 1, 014005.

Xu, X.; Jiang, X.; Ma, C.; Du, P.; Li, X. et al. (2020): Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv: 2002.09334.

Yan, L.; Zhang, H. T.; Goncalves, J.; Xiao, Y.; Wang, M. et al. (2020): A machine learning-based model for survival prediction in patients with severe COVID-19 infection. MedRxiv: 2020.2002.2027.20028027.

Zhang, H.; Saravanan, K. M.; Yang, Y.; Hossain, M. T.; Li, J. et al. (2020): Deep learning based drug screening for novel coronavirus 2019-nCov. Preprints: 2020020061.

Zhavoronkov, A.; Aladinskiy, V.; Zhebrak, A.; Zagribelnyy, B.; Terentiev, V. et al. (2020): Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *Insilico Medicine Hong Kong Ltd A*, vol. 307, pp. E1.

Zheng, C.; Deng, X.; Fu, Q.; Zhou, Q.; Feng, J. et al. (2020): Deep learning-based detection for COVID-19 from chest CT using weak label. MedRxiv: 2020.2003.2012.20027185.

Zhu, H.; Guo, Q.; Li, M.; Wang, C.; Fang, Z. et al. (2020): Host and infectivity prediction of wuhan 2019 novel coronavirus using deep learning algorithm. BioRxiv: 2020.01.21.914044.