

A Survey on Applications of Artificial Intelligence in Fighting Against COVID-19

JIANGUO CHEN, Hunan University, China and University of Toronto, Canada

KENLI LI, Hunan University, China

ZHAOLEI ZHANG, University of Toronto, Canada

KEQIN LI, State University of New York, USA and Hunan University, China

PHILIP S. YU, University of Illinois at Chicago, USA

The COVID-19 pandemic caused by the SARS-CoV-2 virus has spread rapidly worldwide, leading to a global outbreak. Most governments, enterprises, and scientific research institutions are participating in the COVID-19 struggle to curb the spread of the pandemic. As a powerful tool against COVID-19, artificial intelligence (AI) technologies are widely used in combating this pandemic. In this survey, we investigate the main scope and contributions of AI in combating COVID-19 from the aspects of disease detection and diagnosis, virology and pathogenesis, drug and vaccine development, and epidemic and transmission prediction. In addition, we summarize the available data and resources that can be used for AI-based COVID-19 research. Finally, the main challenges and potential directions of AI in fighting against COVID-19 are discussed. Currently, AI mainly focuses on medical image inspection, genomics, drug development, and transmission prediction, and thus AI still has great potential in this field. This survey presents medical and AI researchers with a comprehensive view of the existing and potential applications of AI technology in combating COVID-19 with the goal of inspiring researchers to continue to maximize the advantages of AI and big data to fight COVID-19.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence**; • **Applied computing** → **Life and medical sciences**;

Additional Key Words and Phrases: Artificial intelligence, COVID-19, SARS-CoV-2

This work is partially funded by the National Key R&D Program of China (Grant No. 2020YFB2104000), the National Natural Science Foundation of China (Grant No. 62002110), and the International Postdoctoral Exchange Fellowship Program (Grant No. 20180024). This work is also supported in part by NSF under Grants No. III-1763325, No. III-1909323, and No. SaTC-1930941.

Authors' addresses: J. Chen, Hunan University, College of Computer Science and Electronic Engineering, Changsha, Hunan, China, 410082, University of Toronto, Donnelly Centre for Cellular and Biomolecular Research and Department of Computer Science, Toronto, ON, Canada, 710049; email: jianguo.chen@hnu.edu.cn; K. Li (corresponding author), Hunan University, College of Computer Science and Electronic Engineering, Changsha, Hunan, China, 410082; email: lkl@hnu.edu.cn; Z. Zhang, University of Toronto, Donnelly Centre for Cellular and Biomolecular Research, Department of Molecular Genetics, and Department of Computer Science, Toronto, ON, Canada, 710049; email: zhaolei.zhang@utoronto.ca; K. Li, State University of New York, Department of Computer Science, New Paltz, NY, USA, 12561, Hunan University, College of Computer Science and Electronic Engineering, Changsha, Hunan, China, 410082; email: lik@newpaltz.edu; P. S. Yu, University of Illinois at Chicago, Department of Computer Science, Chicago, IL, USA, 60607; email: psyu@uic.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Association for Computing Machinery.

0360-0300/2021/10-ART158 \$15.00

<https://doi.org/10.1145/3465398>

ACM Reference format:

Jianguo Chen, Kenli Li, Zhaolei Zhang, Keqin Li, and Philip S. Yu. 2021. A Survey on Applications of Artificial Intelligence in Fighting Against COVID-19. *ACM Comput. Surv.* 54, 8, Article 158 (October 2021), 32 pages. <https://doi.org/10.1145/3465398>

1 INTRODUCTION

Severe Acute Respiratory Syndrome Corona-Virus 2 (SARS-CoV-2) is an emerging human infectious coronavirus. Since December 2019, the **coronavirus disease 2019 (COVID-19)** caused by SARS-CoV-2 was first reported in China and later in most countries in the world [9, 52, 210]. The **World Health Organization (WHO)** announced on January 30, 2020, that the outbreak was a **Public Health Emergency of International Concern (PHEIC)**, and confirmed COVID-19 as a pandemic on March 11, 2020. As of February 26, 2021, this disease has been reported in 216 countries or regions around the world and has resulted in serious consequences, including 112,649,371 confirmed COVID-19 cases and 2,501,229 deaths [207].

Most governments, enterprises, and scientific research institutions are fighting COVID-19 from all aspects to curb the spread of the disease [43, 48, 51, 137]. Various stakeholders from different institutions and backgrounds have provided abundant resources and capabilities to support this disease battle [12, 30, 46, 54, 123, 132, 193]. As far as SARS-CoV-2 is concerned, virology, origin and classification, physicochemical properties, receptor interactions, cell entry, genomic variation, and ecology are thoroughly studied [6, 70, 115, 115, 208, 230]. Nucleic acid testing, serologic diagnosis, and medical imaging (i.e., chest X-ray or CT imaging) are the main disease detection and diagnosis methods at present [27, 77, 98, 190, 216]. In terms of pathogenesis, topics such as virus entry and spread, pathological findings, and immune response are the focus [140, 188, 198, 199]. In epidemiology, extensive research has been conducted on the source and spectrum of infection, clinical features, epidemiological characteristics, epidemic prediction, and transmission route tracking [24, 108, 211, 212]. Potential therapeutics of COVID-19 include intensive care, drug development, and vaccine development [32, 111, 143, 192]. In addition, communication prediction and social isolation are currently the main social control methods [1, 90, 150, 224].

Artificial intelligence (AI) is defined as a technology that allows computers to imitate human intelligence to process things, including **Machine Learning (ML)**, knowledge graphs, natural language processing, human-computer interaction, computer vision, biometrics, virtual reality, and augmented reality [23, 107, 182]. ML can be subdivided into traditional ML and **deep learning (DL)**. Traditional ML methods include logistic regression, decision tree, random forest, K-nearest neighbor, Adaboost, K-means clustering, density clustering, hidden Markov models, support vector machine, Naive Bayes, etc [18, 117, 135]. DL is a subset of ML and is a learning method for building deep structural neural networking models [59, 179]. In addition, ML techniques also include transfer learning, active learning, and evolutionary learning. In recent years, AI technology has achieved technological breakthroughs and is widely used in various fields of intelligent medicine, including medical image inspection, disease-assisted diagnosis, surgery, hospital management, and medical big data integration [182]. Moreover, AI is actively explored in emerging fields such as surgical robots, wearable devices, new drug discovery, precision medicine, epidemics prevention and control, and gene sequencing.

Encouragingly, in the short period of time since the outbreak of COVID-19, research in the fields of industry, medical, and science has successfully used advanced AI technologies in the COVID-19 battle and has achieved significant progress. For example, AI supports the diagnosis of COVID-19 through medical image inspections and provides non-invasive detection solutions to prevent medical personnel from contracting infections [4, 23, 107]. AI is used in virology research to

Table 1. Abbreviations and Descriptions of the AI Methods Used in This Survey

Abbreviation	Description	Abbreviation	Description
CL	Chaotic learning	MAE	Modified auto-encoder
CNN	Convolutional neural network	MLP	Multi-layered perceptron
DCNN	Deep CNN	Mol2Vec	Molecular to vectors
DNN	Deep neural network	R-CNN	Regional convolutional network
DT	Decision tree	PNN	Polynomial neural networks [134]
FCN	Fully convolutional network	PR	Polynomial regression [40]
GA	Genetic algorithm	RF	Random forest [18]
GAE	Generative auto-encoder	RL	Reinforcement learning
GANs	Generative adversarial nets	RNN	Recurrent neural network
GAN	Generative adversarial network	SVM	Support vector machine
GRU	Gate recurrent unit	SVR	Support vector regression [11]
KNN	K -nearest neighbor	TL	Transfer learning
LR	Logistic regression	t-SNE	t-distributed stochastic neighbor embedding [117]
LSTM	Long short-term memory [59]	VAE	Variational auto-encoder [179]

analyze the structure of SARS-CoV-2-related proteins and predict new compounds that can be used in drug and vaccine development [19, 129, 231]. In addition, AI has achieved virus source tracking through genomic research, and has successfully discovered the relationship between SARS-CoV-2 and the bat virus, as well as the relationship between SARS-CoV and the **Middle East respiratory syndrome-related coronavirus (MERS-CoV)** [42, 139, 157]. Moreover, AI learns large-scale COVID-19 case data and social media data to construct epidemic transmission models to accurately predict the outbreak time, transmission route, transmission range, and impact of the disease [109, 141, 145, 163]. AI is also widely used in epidemic prevention and social control, such as airport security inspections, patient trajectory tracking, and epidemic visualization [156, 156]. There are some studies and surveys on the COVID-19 epidemic and the related machine learning or artificial intelligence applications. For example, in Reference [175], Shinde et al. summarized various forecasting techniques for COVID-19, including stochastic theory, mathematical models, data science, and machine learning techniques. In Reference [171], Shi et al. provided a survey of AI techniques in imaging data collection, segmentation and diagnosis of COVID-19. However, most of the existing work only focuses on one or more narrow fields of the COVID-19 fight, leaving readers difficult to know the scope of the current research on COVID-19.

In this survey, we present a comprehensive view of the landscape and contributions of AI in combating COVID-19. The main scope of AI in COVID-19 research includes disease detection and diagnosis, virology and pathogenesis, drug and vaccine development, and epidemic and transmission prediction. Note that due to the rapid development of the COVID-19 epidemic, we cited many preprinted references for a comprehensive investigation, which still need to be assessed on their accuracy and quality through peer review. The main scope of AI in combating COVID-19 is summarized in Figure 1.

The rest of the article is structured as follows. Sections 2–5 discuss the four main scope of AI against COVID-19. Section 6 summarizes the available data and resources to support COVID-19 research. Section 7 highlights the challenges and potential directions in this field. Finally, Section 8 presents the conclusion. Table 1 gives the abbreviations and descriptions of the AI methods used in this survey.

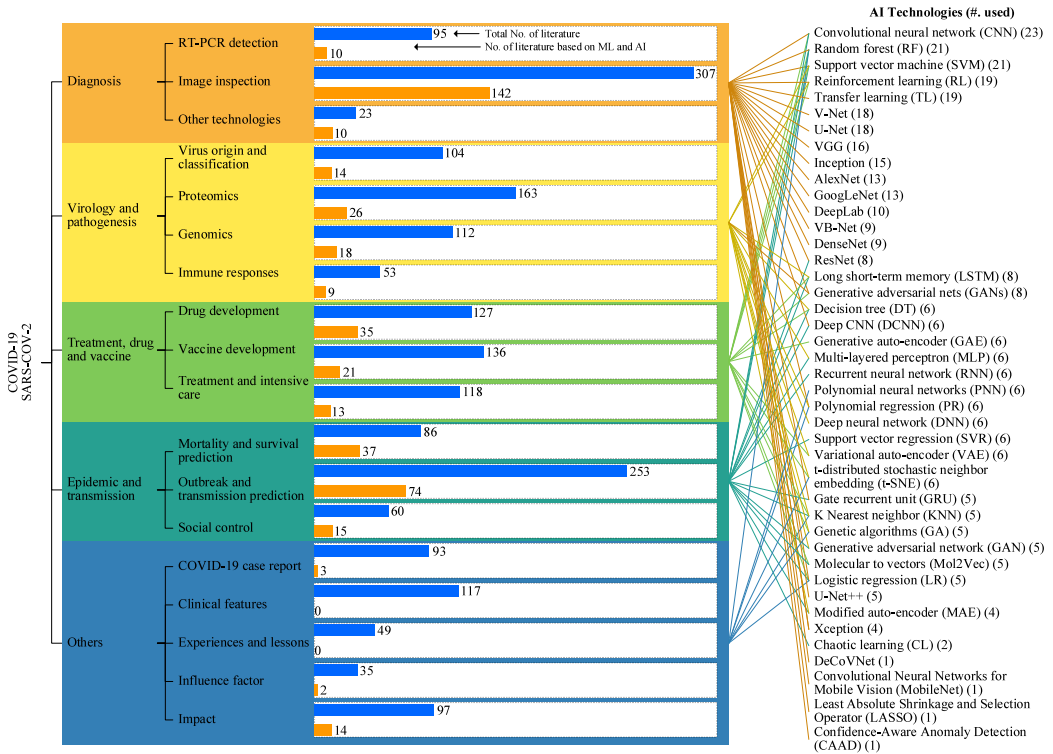


Fig. 1. Main scope of AI technologies in fighting against COVID-19. We collect 2,471 online publications and resources related to COVID-19, SARS-CoV-2, and 2019-nCoV from databases such as Nature, Elsevier, Google Scholar, arxiv, biorxiv, and medRxiv. Then, we filter out 443 papers that explicitly use AI methods. In addition, we list the name of AI technologies used in each field of COVID-19 research.

2 DISEASE DETECTION AND DIAGNOSIS

The diagnosis of virus infection is an important part of COVID-19 research. The current detection and diagnosis methods used for SARS-CoV-2 virus and COVID-19 disease mainly include nucleic acid testing, serological diagnosis, chest X-ray and CT image inspection, and other noninvasive methods.

2.1 RT-PCR Detection

Benefitting from the advantages of high sensitivity and specificity, real-time **Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)** is the current standard detection technology in diagnosing the SARS-CoV-2 virus and bacterial infections. Using RT-PCR, 9 RNA positives were detected from pharyngeal swabs of patients, indicating that the SARS-CoV-2 virus had spread in communities of Wuhan, China, in early January 2020 [98]. The shedding of the SARS-CoV-2 virus detected in the throat, lungs, and feces suggests multiple routes of virus transmission [208, 216]. However, RT-PCR faces the limitations of complicated sample preparation, low detection efficiency, and high false-negative rate [98, 203, 213].

Isothermal nucleic acid amplification and blood testing methods are also commonly used for rapid screening of SARS-CoV-2 [94, 114, 213]. An ML classification method was used for blood testing to extract important routine hematological and biochemical characteristics and to provide

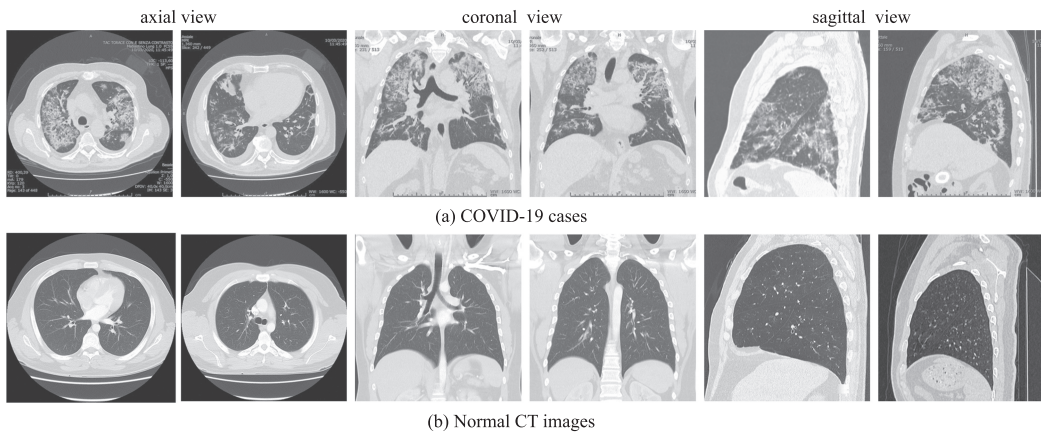


Fig. 2. Examples of lung CT images of normal and COVID-19 cases.

COVID-19 classification. In Reference [213], 105 blood test reports were collected, of which 27 were positive samples from patients with confirmed COVID-19. For comparison, negative samples were collected from patients with ordinary pneumonia, tuberculosis, and lung cancer. Each sample contains 49 feature variables, including 24 routine hematological and 25 biochemical parameters. Next, the authors implemented the RF algorithm [18] on the training samples for feature learning and classification. Based on the extracted 11 key feature variables, they built an RF classifier and tested 253 samples of 169 patients with suspected COVID-19 and obtained an accuracy of 96.97%. Although AI technologies rarely directly participate in RT-PCR and blood testing, the viral load and COVID-19 case data collected in these methods provide important data sources for the subsequent AI-based analysis.

2.2 Medical Image Inspection

Medical imaging inspection is another widely used clinical approach for COVID-19 detection and diagnosis. COVID-19 medical image inspection mainly includes chest X-ray and lung CT imaging. AI technology plays an important role in medical image inspection and has achieved significant results in image acquisition, organ recognition, infection region segmentation, and disease classification. It not only greatly shortens the imaging diagnosis time of radiologists but also improves the accuracy the diagnosis. We will discuss in detail the contributions of AI methods to chest X-ray and lung CT imaging.

2.2.1 CT Image Inspection. CT imaging provides an important basis for the early diagnosis of COVID-19. The CT imaging manifestations of COVID-19 are mainly Ground Glass Opacity (GGO) in the periphery of the subpleural region, and some are consolidated. If a patient's COVID-19 condition improves, then the area will be absorbed and form fibrous stripes [38, 154, 177]. Examples of lung CT images of normal and COVID-19 cases are shown in Figure 2.

The progress of AI-based CT image inspection for COVID-19 usually includes the following steps: Region Of Interest (ROI) segmentation, lung tissue feature extraction, candidate infection region detection, and COVID-19 classification. The representative AI architecture for CT image classification and COVID-19 inspection is shown in Figure 3.

The segmentation of lung organs and ROIs is the foundational step in AI-based image inspection. It depicts the ROIs in lung CT images (i.e., lungs, lung lobes, bronchopulmonary segments, and infected regions or lesions) for further evaluation and quantification. Different DL models (i.e., U-Net, V-Net, and VB-Net) have been used for CT image segmentation [29, 107, 185, 215]. In

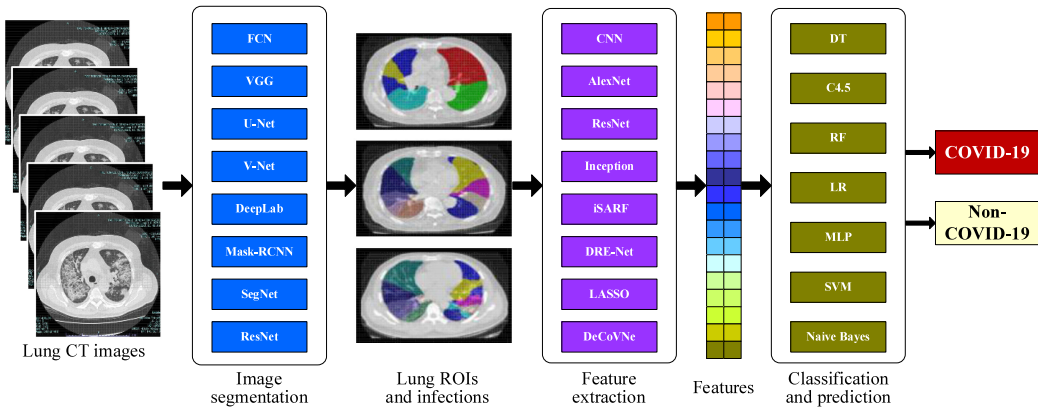


Fig. 3. Representative AI architecture for CT image classification and COVID-19 inspection.

Reference [170], Shan et al. collected 549 CT images from patients with confirmed COVID-19 and proposed an improved segmentation model (called VB-Net) based on the V-Net [110] and ResNet [61] models. In Reference [29], Chen et al. established a DL model based on the U-Net++ structure [232] to extract the ROIs from each CT image and detect the training curve of suspicious lesions. In Reference [215], Xu et al. used a 3D DL model to segment the infection regions from lung CT images. Then, they built a classification model using ResNet and location-attention structures, and divided the segmented regional images into three categories, such as COVID-19, influenza-A viral pneumonia, and normal. In Reference [107], Li et al. used the U-Net segmentation model to extract lung organs as ROIs from each lung CT image. In Reference [185], Tang et al. used the VB-Net model [170] to accurately segment 18 lung regions and infected regions from lung CT images, and further calculated 63 quantitative features.

Focusing on the detection and location of candidate infection regions, different AI methods were proposed in References [58, 79, 173, 203]. In Reference [58], Gozes et al. used commercial software to identify lung nodules and small opacity in the 3D lung volume. Then, they constructed a DL model composed of the U-Net and ResNet structures, where the U-Net module was used to extract the ROI regions, and the ResNet model was used to detect and classify diffuse turbidity and ground glass infiltration. In addition, they compared the CT images of 56 patients with confirmed COVID-19 and 101 non-coronavirus patients, and analyzed the CT features of COVID-19 in detail. In Reference [173], Shi et al. used a V-Net-based CNN model to segment lung organs and infected regions from lung CT images. Then, they used the Least Absolute Shrinkage and Selection Operator (LASSO) method to calculate the best CT morphological features. Finally, based on the best CT morphology and clinical features, the severity of COVID-19 was predicted and evaluated. In Reference [203], Wang et al. collected 195 CT images from 44 patients with COVID-19 and 258 CT images from 55 negative patients. They used the CNN model with the Inception structure [183] to classify randomly selected ROI images and predict COVID-19 disease. In Reference [79], Huang et al. used an InferReadTM CT pneumonia tool based on AI to quantitatively evaluate changes in the lung burden of patients with COVID-19. The tool includes three modules: lung and lobe extraction, pneumonia segmentation, and quantitative analysis. The CT image features of COVID-19 pneumonia are divided into four types: mild, moderate, severe, and critical.

Based on ROI segmentation and candidate infection region detection, the important features of ROIs and infection regions are extracted for COVID-19 classification [149]. In Reference [149], Qi et al. collected 71 CT images from 52 patients with confirmed COVID-19 in 5 hospitals. They used radiobiological methods to extract 1,218 features from each CT image, and then performed LR and

RF methods on these features to distinguish between short-term and long-term hospital stays. In Reference [172], Shi et al. used the VB-Net model [170] to segment lung and infection regions from CT images, and classified them based on 96 features (including 26 volume features, 31 digital features, 32 histogram features, and 7 surface features). Next, they proposed an iSARF method to classify features and predict COVID-19 disease. Comparative experiments show that the iSARF method is superior to LR, SVM, and NN methods. In Reference [229], Zheng et al. proposed a 3D DCNN model (called DeCoVNet) to detect COVID-19 from CT images. The proposed DeCoVNet model includes three components: the first component uses vanilla 3D convolutional layers to extract lung image features, the second component consists of two 3D residual blocks, which perform element conversion on the 3D feature maps, and the third component gradually extracts the information in the 3D feature map through 3D max-pooling, and outputs the probability of COVID-19. In Reference [180], Song et al. collected 1990 CT images, including 777 images from 88 patients with COVID-19, 505 images from 100 patients with bacterial pneumonia, and 708 images from 86 healthy people. They proposed a DRE-Net DL model based on a pre-trained ResNet50 structure and a functional pyramid network. The DRE-Net model extracts the top-K lesion features from each CT image to predict the classification of patients with COVID-19.

The lack of large-scale data sets is the main challenge that hinders the implementation of AI-based CT image inspection and affects diagnostic performance. To address these challenges, strategies such as transfer learning, data augmentation, and “Human-In-The-Loop” were used in References [170, 227]. In Reference [227], Zhao et al. provided a public COVID-19 CT scan data set, including 275 COVID-19 cases and 195 non-COVID-19 cases. They used data augmentation and TL methods to alleviate the shortage of training data. In terms of data augmentation, they used transformation operations to expand the training data set, such as random transformation, cropping, and rotation. In terms of TL, they pre-trained the DenseNet model [78] on the chest X-ray data set [204], and then used the pre-trained model to predict COVID-19. In addition, the strategy of “Human-In-The-Loop” was adopted to reduce the workload of radiologists when annotating training samples [170]. The radiologists annotated a small portion of training samples in the first batch of training. Then, they manually corrected the segmentation results in the second batch and used them as annotations for the images. Iterative training is performed in this way to complete the annotation of all training samples.

It is commendable that several works provide open-source code of the designed models and online COVID-19 CT image inspection systems. For example, Li [107], Zheng [229], and Zhao [227] published the proposed DL models on GitHub [81]. In addition, Song et al. provided an online CT diagnosis service [180], and Wang et al. provided a public website for uploading and testing lung CT images [203]. In Reference [29], Chen et al. developed a public online CT diagnostic system, and anyone can upload CT images for self-diagnosis. More detailed information about AI-based CT image segmentation and classification methods is provided in Table 2 and Figure 4.

2.2.2 Chest X-ray Image Inspection. Compared with CT images, chest X-ray (CXR) images are easier to obtain in radiological inspections. Although CXR imaging is a typical imaging method used for the diagnosis of COVID-19, it is generally considered to be less sensitive than CT imaging. Some CXR images of patients with early COVID-19 showed normal characteristics. Radiological signs of COVID-19 CXR images include airspace opacity, GGO, and later mergers. In addition, the distribution of bilateral, peripheral, and lower regions is mainly observed [116, 153]. Examples of CXR images of normal and COVID-19 cases are shown in Figure 5. The AI-based CXR image inspection usually includes steps such as data preprocessing, DL model training, and COVID-19 classification. The representative AI architecture for CXR image classification and COVID-19 inspection is shown in Figure 6.

Table 2. AI-based CT Image Segmentation and Classification Methods for COVID-19 Inspection

Literature	Data sources	Data size	COVID-19 cases	AI methods	ACC/AUC	Sensitivity	Specificity
Chen [29] ¹	private	35,355	20,886	U-Net++	95.24%	100%	94.0%
Gozes [58]	private	157	56	U-Net, ResNet	99.6%	98.2%	92.2%
Huang [79]	private	842	842	U-Net	—	—	—
Li [107] ²	private	4,356	1,325	U-Net, ResNet	96.0%	90.0%	96.0%
Qi [149]	private	52	52	LR, RF	97.0%	100%	75.0%
Shan [170]	private	549	549	V-Net, ResNet	—	—	—
Shi [172]	private	2,685	1,658	RF	87.9%	90.7%	83.3%
Shi [173]	private	196	196	V-Net	89.0%	82.2%	82.8%
Song [180] ³	private	1,990	777	DRE-Net	94.0%	93.0%	—
Tang [185]	private	176	176	RF	87.5%	—	—
Wang [203] ⁴	private	453	195	Inception	73.1%	74.0%	67.0%
Xu [215]	private	618	219	ResNet	86.7%	—	—
Zhao [227] ⁵	[227]	470	275	DenseNet	84.7%	76.2%	—
Zheng [229] ⁶	private	630	630	DeCoVNet	90.1%	90.7%	91.1%

¹<http://121.40.75.149/znyx-ncov/index>. ²<https://github.com/bkong999/COVNet.git>. ³<http://biomed.nscg-gz.cn/server/Ncov2019>. ⁴https://ai.nscg-tj.cn/thai/deploy/public/pneumonia_ct. ⁵<https://github.com/UCSD-AI4H/COVID-CT>. ⁶<https://github.com/sydney0zq/covid-19-detection>.

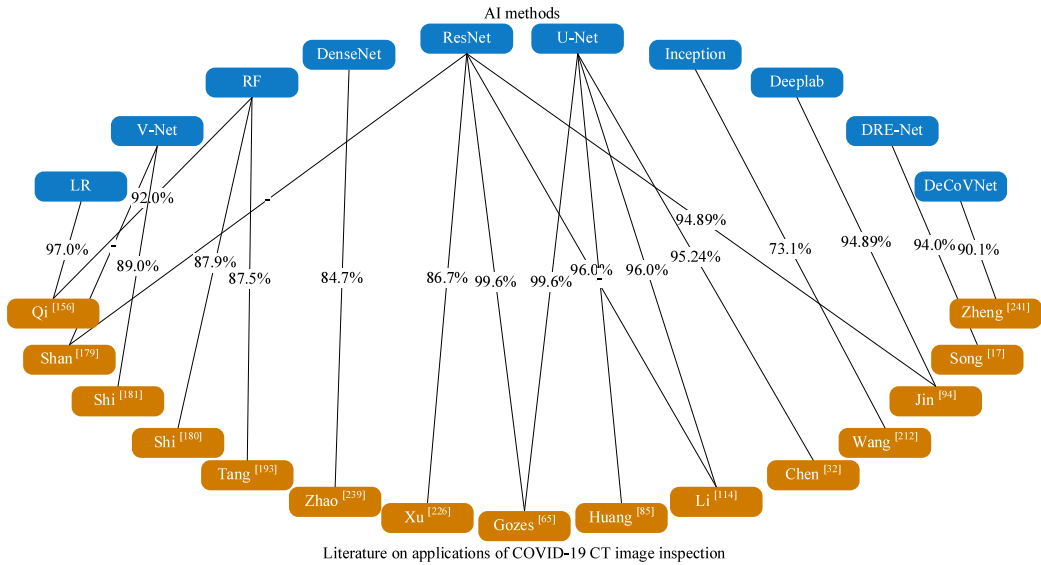


Fig. 4. Relationship between AI methods and applications of COVID-19 CT image inspection.

Unlike CT images, CXR image segmentation is more challenging, because the ribs are projected onto soft tissues, which is confused with image contrast. In this way, most DL models focus on the classification of the entire CXR image, while few works are devoted to segmenting the ROIs and lung organs from CXR images. In Reference [60], Mahdy et al. used a classification method to classify COVID-19 on CXR images through a multilevel threshold and SVM. A multilevel image

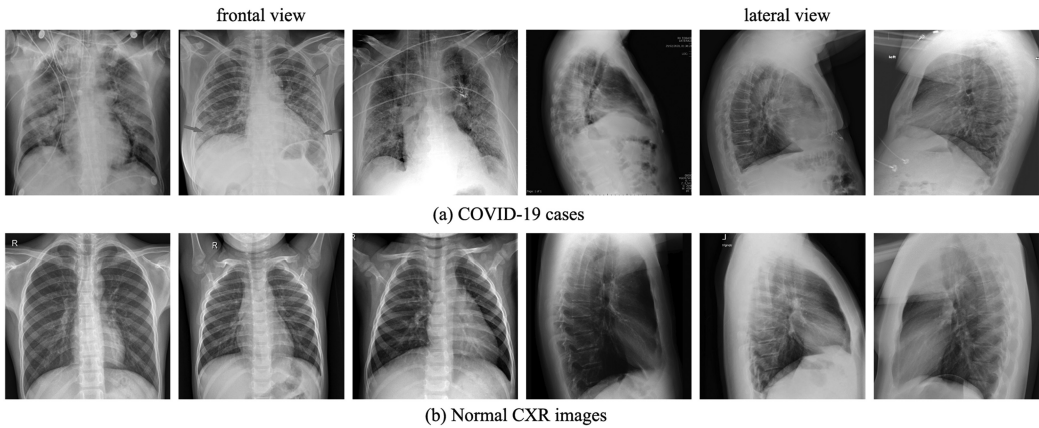


Fig. 5. Examples of chest X-ray images of normal and COVID-19 cases.

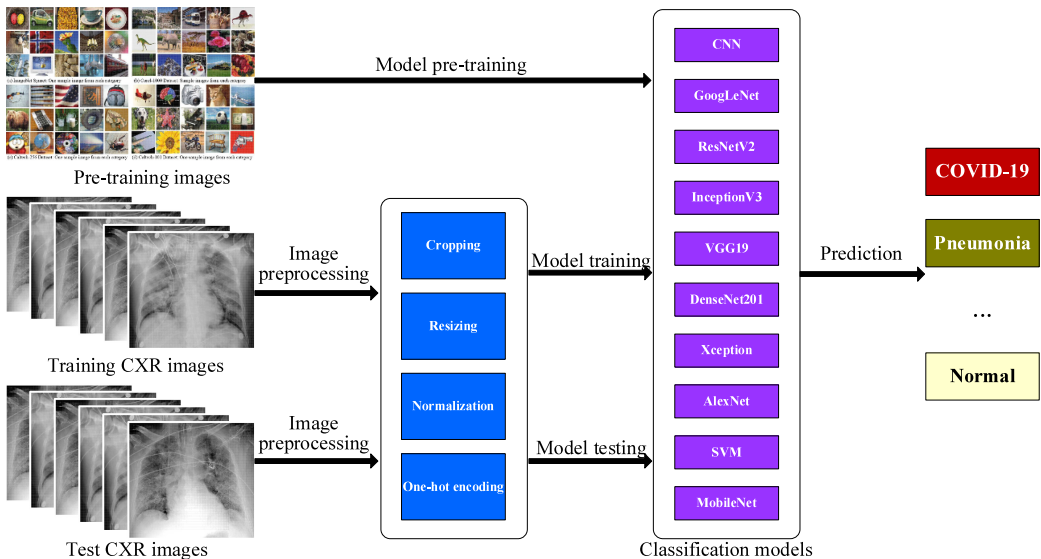


Fig. 6. Representative AI architecture for CXR image classification and COVID-19 inspection.

segmentation threshold was used to segment lung organs from the background, and then the SVM module was used to classify the infected lungs from the uninfected lungs.

Focusing on the classification of COVID-19 based on CXR images, several studies built AI-based classification models by nesting or combining existing ML and DL models. In Reference [65], Hemdan et al. proposed a DL framework (called COVIDX-Net) to help radiologists automatically diagnose COVID-19 based on CXR images. The proposed framework integrates 7 DCNN models with different structures, such as VGG19 [176], DenseNet201 [78], ResNetV2 [26], InceptionV3, InceptionResNetV2 [182], Xception [35], and MobileNetV2 [162]. Each model is separately trained on CXR images to classify the patient’s status as COVID-19 positive or negative. In Reference [169], Sethy et al. used different DCNN models in a SVM classifier to diagnose COVID-19 based on CXR images. Eleven DCNN models are used as image feature extractors, including AlexNet, GoogLeNet, DenseNet, Inception, ResNet, VGG, XceptionNet, and InceptionResNet. Then, the SVM

classifier classifies the extracted image features to determine whether the input image is COVID-19. In Reference [23], Castiglioni et al. collected CXR data (610 images, including 324 COVID-19 cases) from Lombardy, Italy, and constructed a DCNN model to predict COVID-19. The DCNN model is equipped with 10 CNN models, each of them uses the ResNet-50 structure and has been pre-trained on the ImageNet data set. In Reference [8], Ioannis et al. evaluated the performance of 5 DCNN models for medical image classification, including VGG19, MobileNetV2, Inception, Xception, and InceptionResNetV2. To address the shortcomings of the small-scale COVID-19 data set, each model is pre-trained on the ImageNet data set using the TL strategy. In Reference [130], Narin et al. also used 3 typical DCNN models (i.e., ResNet50, InceptionV3, and InceptionResNetV2) to classify COVID-19 from a small-scale CXR data set (100 images, including 50 COVID-19 cases).

Another group of works constructed specific DL models for COVID-19 classification [225]. For example, Zhang et al. proposed a new DL model, which consists of a backbone network, a classification module, and an anomaly detection module [225]. The backbone network extracts the features of each input CXR image. The classification and anomaly detection modules, respectively, use the extracted features to generate classification scores and scalar anomaly scores. In Reference [201], Wang et al. introduced a COVID-Net DCNN model to identify COVID-19 cases based on CXR images. The COVID-Net model uses a large number of convolutional layers in a projection-expansion-projection design pattern. They collected 13,800 CXR images from 13,725 patients (including 183 COVID-19 patients) to establish a CXR database (called COVIDx) for training COVID-Net. It is commendable that the authors provided an open-source code of the proposed model and the COVIDx database.

Similar to CT images, in CXR image inspection, there is also a lack of large-scale data sets for DL model training. In Reference [119], Maghdid et al., respectively, used CNN and AlexNet [102] models to train CXR and CT images and diagnose COVID-19 cases. Among them, the AlexNet model is pre-trained on the ImageNet data set to perform COVID-19 classification on the data sets in References [36, 116, 133]. Unlike existing TL and image augmentation methods, Afshar et al. designed a capsule network model (named COVID-CAPS) suitable for small-scale CXR data sets [2]. Each layer of the COVID-CAPS model contains multiple capsules, and each capsule represents a specific image instance at a specific position through multiple neurons. The capsule module [67] uses protocol routing to capture alternative models of spatial information and attempts to reach a consensus on the existence of objects. In this way, the protocol uses information from instances and objects to identify the relationship between them without the need for large-scale data sets. More detailed information about AI-based methods for CXR image classification and COVID-19 inspection is shown in Table 3 and Figure 7.

2.3 Other Detection Technologies

In addition to RT-PCR detection and image inspection techniques, some noninvasive measurement methods have also been used for the diagnosis of COVID-19, including cough sound judgment and breathing pattern detection.

(1) Monitoring COVID-19 through AI-based cough sound analysis.

In Reference [166], Schuller et al. discussed the potential application of **computer audition (CA)** and AI in analyzing the cough sounds of patients with COVID-19. They first analyzed the ability of CA to automatically recognize and monitor speech and cough under different semantics (such as breathing, dry and wet coughing or sneezing, speech during colds, eating behaviors, drowsiness, or pain). Then, they applied the CA technology to the diagnosis and treatment of patients with COVID-19. However, due to the lack of available data sets and annotation information, there are few studies on this technology in diagnosing COVID-19. It is encouraging that Orlandic et al. provided a COUGHVID cough dataset for COVID-19 diagnosis, which contains

Table 3. AI-based Methods for CXR Image Classification for COVID-19 Inspection

Literature	Data sources	Data size	COVID-19 cases	AI methods	Accuracy/AUC	Sensitivity	Specificity
Afshar [2] ¹	[36, 131]	100	50	COVID-CAPS	95.70%	90.00%	95.80%
Castiglioni [23]	private	610	324	ResNet	89.00%	78.00%	82.00%
Ioannis [8]	[36, 93, 105]	1,427	224	VGG, MobileNet, Inception, Xception, Inception-ResNet	96.78%	98.66%	96.46%
Mahdy [60]	private	40	25	SVM	97.48%	95.28%	99.70%
Hemdan [65]	[36, 152]	50	25	VGG, DenseNet, ResNet, Inception, InceptionResNet, Xception, MobileNet	83.00%	91.00%	91.00%
Maghdid [119]	[36, 116, 133]	170	60	CNN, AlexNet	94.00%	100%	88.00%
Narin [130]	[36, 131]	100	50	ResNet, Inception, InceptionResNet	98.00%	96.00%	98.00%
Sethy [169]	[36, 131]	50	25	AlexNet, DenseNet, GoogLeNet, Inception, ResNet, VGG, Xception, InceptionResNet, SVM	95.38%	97.29%	93.47%
Wang [201] ²	[20, 36, 131]	13,800	183	COVID-Net	92.6%	87.10%	96.40%
Zhang [225]	[36, 204]	1,531	100	CAAD	95.18%	96.00%	70.65%

¹<https://github.com/ShahinSHH/COVID-CAPS>. ²<https://github.com/lindawangw/COVID-Net>.

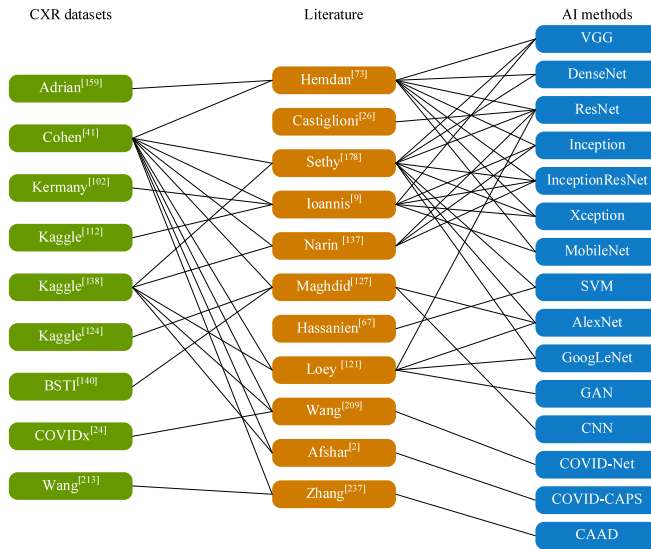


Fig. 7. Relationship among CXR data sets, AI methods, and applications of CXR image inspection.

20,000 crowdsourced cough records [138]. The cough records are labeled by experienced pulmonologists, and medical abnormalities are accurately diagnosed, thus providing a high-quality training dataset for MI/AI methods. Similarly, in Reference [83], Iqbal et al. also discussed an abstract framework that uses the voice recognition function of mobile applications to capture and analyze the cough sounds of suspicious persons to determine whether they are healthy or suffering from a respiratory disease. In Reference [205], Wang et al. analyzed the respiratory patterns of patients with COVID-19 and other breathing patterns of patients with influenza and common cold. In addition, they proposed a respiratory simulation model (called BI-AT-GRU) for COVID-19 diagnosis. The BI-AT-GRU model includes a GRU neural network with a bidirectional and attention mechanism, and can classify six types of clinical respiratory patterns, such as Eupnea,

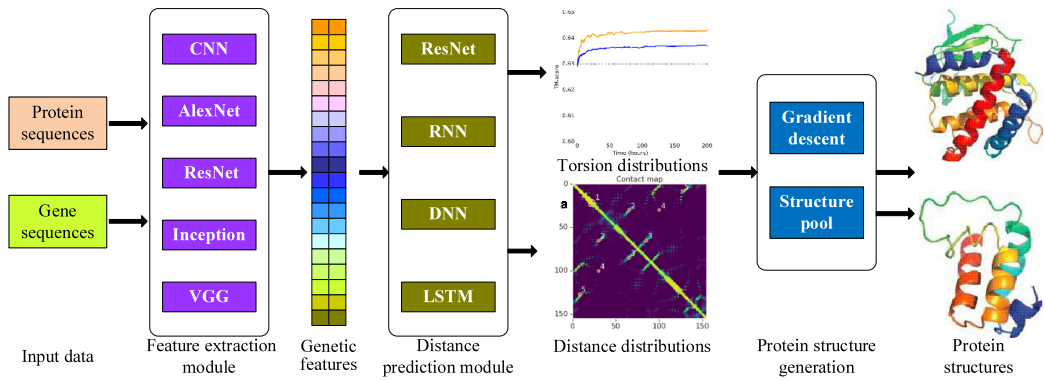


Fig. 8. Representative AI architecture used for protein structure prediction.

Tachypnea, Bradypnea, Biots, Cheyne-Stokes, and Central-Apnea. To facilitate the analysis of COVID-19 cough sounds, we further collect eight groups of human cough sound datasets containing patients with confirmed COVID-19, as shown in Table 3 in Supplementary Material.

(2) COVID-19 diagnosis based on non-invasive measurements.

In Reference [120], Maghdid et al. designed an abstract framework for COVID-19 diagnosis based on smart phone sensors. In the proposed framework, smart phones can be used to collect the disease characteristics of potential patients. For example, the sensor can acquire the patient's voice through the recording function and the patient's body temperature through the fingerprint recognition function. Then, the collected data are submitted to an AI-supported cloud server for disease diagnosis and analysis.

3 VIROLOGY AND PATHOGENESIS

The virology and pathogenesis of SARS-CoV-2 is the most important scientific research in the fields of biology and medicine. Scientists have analyzed the virus characteristics of SARS-CoV-2 through proteomics and genomic studies [6, 70, 115]. In the field of virology, the origin and classification of SARS-CoV-2, physical and chemical properties, receptor interactions, cell entry, and the ecology and genomic variation of SARS-CoV-2 have been studied [115, 208, 230]. We mainly discuss the contribution of AI in the pathological research of SARS-CoV-2 from the perspective of proteomics and genomics.

3.1 Proteomics

Since the advent of SARS-CoV-2, there have been a large number of research achievements in proteomics. Five types of structural proteins of SARS-CoV-2 have been confirmed, including nucleocapsid (N) proteins, envelope (E) proteins, membrane (M) proteins, and spike (S) proteins [139, 198, 228]. Other proteins translated in the host cells essential for virus replication, such as **non-structural protein 5 (NSP5)** and **3C-like protease (3CLpro)**, have also attracted the attention of researchers. In addition, several studies have shown that SARS-CoV-2 uses the human **Angiotensin-Converting Enzyme 2 (ACE2)** to enter the host [70, 230]. In this field, AI techniques are used to predict protein structures and analyze the interaction network between proteins and drugs. The representative AI architecture used for protein structure prediction is shown in Figure 8.

In References [167, 168], Senior et al. used DL models to implement the AlphaFold system for protein structure prediction. The AlphaFold system uses the ResNet model [61] to analyze the

covariance and amino acid residue contacts in the homologous gene sequences, and predict the corresponding protein structures. The AlphaFold system consists of a feature extraction module and a distance prediction neural network. The feature extraction module is responsible for searching for protein sequences similar to the input protein sequences and constructing **multiple sequence alignments (MSA)**. The module simultaneously generates residual position and sequence contour features, and inputs the output of 485 feature parameters into the distance prediction neural network. The distance prediction neural network is a **two-dimensional (2D)** ResNet structure, which is responsible for accurately predicting the distance between all residue pairs in every two protein sequences. The authors added a one-dimensional output layer to the network to predict the accessible surface area, distance map, and secondary structure of each residue. Finally, the generated potential is optimized by gradient descent to generate protein structures.

Based on References [167, 168], Jumper et al. used the AlphaFold system to predict the structure of SARS-CoV-2 membrane proteins [87]. They published the predicted protein structures such as 3a, Nsp2, Nsp4, Nsp6, and papain-like proteases. Although the structure of these proteins has not been verified by clinical trials, this publication allows researchers to quickly conduct SARA-CoV-2 studies. In Reference [139], Ortega et al. used a computational method to detect changes in the S1 subunit of the spike receptor-binding domain and identify mutations in the SARS-CoV-2 spike protein sequence, which may be beneficial for studying human-to-human transmission. They collected sequences from the **Protein Data Bank (PDB)** [15] and used the SWISS-MODEL software [10] to construct a SARS-CoV-2 spike protein model. Then, the Z-dock software [144] was used to dock the spike protein and ACE2, and a clustering algorithm was used to cluster the docking results. The work indicated that the SARS-CoV-2 spike protein has a higher affinity for human ACE2 receptors.

Another branch of AI-assisted proteomics research involves finding new compounds and drug candidates for the treatment of COVID-19 by building an interactive network and knowledge graph between proteins and drugs. Please see Section 4 for details.

3.2 Genomics

Genomics is mainly used in SARS-CoV-2 to analyze the origin of SARS-CoV-2, vaccine development, and PT-PCR detection. Various AI algorithms are applied to compare the similarity of gene sequences, gene fragments, and miRNA predictions [42, 157]. In Reference [157], Randhawa et al. used different ML methods to analyze the pathogen sequence of COVID-19, and identified the inherent characteristics of the viral genomes, thereby rapidly classifying new pathogens. They collected the complete reference genome of the COVID virus from NCBI [46], the bat β -coronavirus from GISAID [54], and all available virus sequences from Virus-Host DB [127]. By using chaotic game representation, each genomic sequence was mapped to the corresponding genomic signal in a discrete digital sequence [86]. In addition, the amplitude spectrum of these genomic signals was calculated by using a discrete Fourier transform. On this basis, they used six ML classification models to train the above sequence distance matrix and compared their performance. Finally, they conducted the trained ML models on 29 COVID-19 sequences to classify COVID-19 pathogens. The results of this work support the hypothesis that COVID-19 originated in bats and its classification as a β -coronavirus.

In addition, the amplitude spectrum of these genomic signals is calculated by using a discrete Fourier transform. On this basis, they used 6 ML classification models (such as linear discriminant, linear SVM, quadratic SVM, fine KNN, subspace discriminant, and subspace KNN) to train the sequence distance matrix and compared their performance. Finally, they performed a well-trained ML model on 29 COVID-19 sequences to classify COVID-19 pathogens. The results of

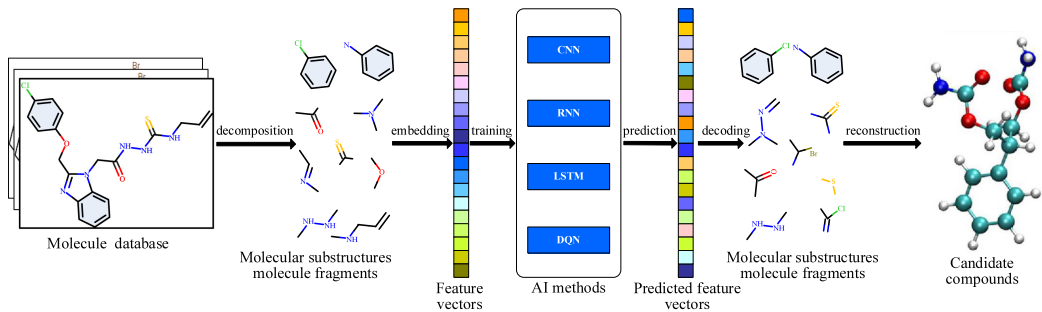


Fig. 9. Representative AI architecture used for new drug-like compound discover.

this work support the following hypothesis: COVID-19 originated in bats and is classified as type β -coronavirus.

In Reference [42], based on three ML methods (such as DT, Naive Bayes, and RF), Demirci et al. performed miRNA prediction on the SARS-CoV-2 genome, and identified miRNA-like hairpins and microRNA-mediated SARS-CoV-2 infection interactions. They collected the complete COVID-19 genome from NCBI [46] and human-mature miRNA sequences from miRBase [101]. The genomic sequences are transcribed and divided into multiple overlapping fragments, which are folded into a secondary structure to extract the hairpin structure. On this basis, the authors used the three ML methods to predict the category of each hairpin and determine the similarity between the hairpins and human miRNA. They searched for mature miRNA targets in human and SARS-CoV-2 genes, and analyzed the potential interactions between SARS-CoV-2 miRNAs and human genes and between human miRNAs and SARS-CoV-2 genes. Finally, the gene ontology of SARS-CoV-2 miRNA targets in human genes was analyzed, and the PANTHER classification system was used to evaluate the similarity between SARS-CoV-2 miRNA candidates and the mature miRNAs of any known organism [126].

In Reference [125], Metsky et al. used genomic and AI technologies to design nucleic acid detection assays and improve the current SARS-CoV-2 RT-PCR test. They developed a CRISPR tool that uses enzymes to edit the genome by cutting specific genetic code chains, and used different ML methods to predict the diversity of the target genome. The authors designed an RT-PCR test method through the CRISPR tool, which can effectively detect 67 respiratory viruses, including SARS-CoV-2.

4 DRUG AND VACCINE DEVELOPMENT

Based on proteomics and genomics research, a variety of drug and vaccine development programs have been proposed for SARS-CoV-2 and COVID-19. The application of AI in the development of new drugs and vaccines is one of the main contributions in smart medicine, and it plays an important role in the battle against COVID-19.

4.1 Drug Development

In the field of drug development, AI technologies can screen existing drug candidates for COVID-19 by analyzing the interaction between existing drugs and COVID-19 protein targets. In addition, AI technologies can also help discover new drug-like compounds against COVID-19 by constructing new molecular structures that inhibit proteases at the molecular level. The representative AI architecture used for new drug-like compound discover is shown in Figure 9.

Drug development can be divided into small-molecule drug discovery and biological product development. The discovery of small-molecule drugs mainly focused on chemically

synthesized small molecule active substances, which can be made into small-molecule drugs through chemical reactions between different organic and inorganic compounds. One group of AI-based drug development focuses on discovering new drug-like compounds at the molecular level. In References [14, 174], Beck et al. proposed a DL-based **molecule transformer-drug target interaction (MT-DTI)** model to predict potential drug candidates for COVID-19. The MT-DTI model uses the SMILES strings and amino-acid sequences to predict the target proteins with a 3D crystal structure. The authors collected the amino-acid sequences of 3C-like proteases and related antiviral drugs and drug targets from NCBI [46], **Drug Target Common (DTC)** [186], and BindingDB [113] databases. In addition, they used a molecular docking and virtual screening tool (AutoDock Vina [194]) to predict the binding affinity between 3,410 drugs and SARS-CoV-2 3CLpro. The experimental results provided 6 potential drugs, such as Remdesivir, Atazanavir, Efavirenz, Ritonavir, Dolutegravir, and Kaletra (lopinavir/ritonavir). Note that Remdesivir shows promising in clinical trials. In Reference [129], Moskal et al. used AI methods to analyze the molecular similarity between anti-COVID-19 drugs (termed “parents”) and drugs involving similar indications to screen out second-generation drugs (termed “progeny”) of COVID-19. They first used the Mol2Vec [84] method to convert the molecular structure of the parent drugs into a high-dimensional vector space, treated the drug molecule as a “sentence,” and mapped its molecular substructure to a “word.” Then, they used the VAE [179] model to generate SMILES strings with 3D shape and pharmacodynamic properties to the given seed molecule [56]. In addition, CNN, LSTM, and MLP models are used to generate the corresponding SMILES strings and molecules. The authors selected 71 parent drugs from the literature as seed molecules, and 4456 drugs from ZINC [233] and ChEMBL as candidate progeny drugs [49].

In Reference [19], Bung et al. committed to the development of new chemical entities of SARS-CoV-2 3CLpro based on DL technology. They constructed an RL-based RNN model to classify protease inhibitor molecules and obtained a smaller subset that favored the chemical space. Then, they collected 2515 protease inhibitor molecules in the SMILES format from the ChEMBL database as training data, where each SMILES string is regarded as a time series, and each position or symbol is regarded as a time point. The output of small molecules is docked to the 3CLpro structure with minimal energy and ranked according to the virtual screening score obtained by selecting anti-SARS-CoV-2 candidates [194]. In Reference [184], Tang et al. analyzed 3CLpro with a 3D structure similar to SARS-CoV, and evaluated it as an attractive target for the development of anti-COVID-19 drugs. They proposed an advanced deep Q learning network (called ADQN-FBDD) to generate potential lead compounds of SARS-CoV-2 3CLpro. They collected 284 molecules reported as SARS-CoV-2 3CLpro inhibitors. An improved BRICS algorithm [41] was used to split these molecules to obtain the target fragment library of SARS-CoV-2 3CLpro. Then, the proposed ADQN-FBDD model trains each target fragment and predicts the corresponding molecules and lead compounds. Through the proposed **Structure-based Optimization Policy (SBOP)**, they finally obtained 47 derivatives that inhibit SARS-CoV-2 3CLpro from these lead compounds, which are regarded as potential anti-SARS-CoV-2 drugs.

Another group of studies focuses on screening candidate biological products of COVID-19. Biological products are protein products with therapeutic effects, which mainly bind to specific cell receptors involved in the disease process. Biological products are prepared from microbial cells (such as genetically modified bacteria, yeast, or mammalian cell strains) through a biotechnology process. In Reference [73], Hu et al. established a multi-task DL model to predict the possible binding between potential drugs and SARS-CoV-2 protein targets, to select drugs that can be used for SARS-CoV-2. They first collected 8 SARS-CoV-2 viral proteins from GHDDI [51] as potential targets. The proposed DL model is based on the AtomNet model [181, 197] and includes a shared

Table 4. AI-based COVID-19 Drug Development Methods

Literature	AI methods	Role of AI methods	COVID-19 targets	Potential drugs
Tang [184] ¹	RL, DQN	predict molecules and lead compounds for each target fragment	3CLpro	47 compounds
Zhavoronkov [228] ²	28 ML	generate new molecular structures for 3CLpro	3CLpro	100 molecules
Bung [19]	RNN, RL	classify protease inhibitor molecules	3CLpro	31 compounds
Hofmarcher [71] ³	ChemAI	predict inhibitory effects of molecules on COVID-19 proteases	3CLpro, PLpro	30,000 molecules
Kadioglu [89]	NN, Naive bayes	construct drug likelihood prediction model	spike protein, . . .	3 drugs
Hu [73]	DL	predict binding between drugs and protein targets	3CLpro, RdRp, . . .	10 drugs
Beck [14, 174]	MT-DTI	predict binding affinity between drugs and protein targets	3CLpro, RdRp, helicase, . . .	6 drugs
Moskal [129]	VAE, CNN, LSTM, MLP	generate SMILES strings and molecules	—	110 drugs

¹<https://github.com/tbwxmu/2019-nCov>. ²<https://www.insilico.com/ncov-sprint>. ³<https://github.com/ml-jku/sars-cov-inhibitors-chemai>.

layer to learn the joint representation of all tasks and a task processing layer to perform specific tasks. By fine-tuning the DL model using a coronavirus-specific data set, the model can predict the possible binding between the drugs and the protein targets and output the binding affinity score. Based on existing studies, RdRp, 3CLpro, and papain-like protease have been confirmed as the three principal targets of SARS-CoV-2 [53, 140, 198]. Based on the prediction results [106, 202], the authors selected the top 10 potential drugs with a high likelihood of inhibiting each target. In Reference [89], Kadioglu et al. used **High-Performance Computing (HPC)**, virtual drug screening, molecular docking, and ML technologies to identify SARS-CoV-2 drug candidates. After performing virtual drug screening and molecular docking, two supervised ML models (such as NN and Naive Bayes) are used to analyze clinical drugs and test compounds to construct corresponding drug likelihood prediction models. Several approved drugs were selected as SARS-CoV-2 drug candidates, including drugs for **hepatitis C virus (HCV)**, enveloped ssRNA virus, and other infectious diseases.

Facing the known COVID-19 protease target 3CLpro, Zhavoronkov et al. designed a small-molecule drug-discovery pipeline to produce 3CLpro inhibitors, and used 3CLpro's crystal structure, homology modelling and co-crystallized fragments to generate 3CLpro molecules [228]. They collected the crystal structure of COVID-19 3CLpro from Reference [219] and constructed a homology model. At the same time, molecules activity to various proteases were extracted from References [49, 82] and constituted a protease peptidomimetic data set with 5,891 compounds. Then, they used 28 ML methods (such as GAE, GAN, and GA) and RL strategies to separately train the input data sets (e.g., crystal structure, homology model, and co-crystal ligands), and generated new molecular structures with a high score. In Reference [71], Hofmarcher et al. used the ChemAI DL model [147] based on the SmilesLSTM structure [69] to test the resistance of the molecules to COVID-19 protease. They collected 3.6 million molecules from ChEMBL [49], ZINC [233], and PubChem [95] databases and formed a training data set. Then, the ChemAI model is trained on the data set in a multi-task parallel training way, where the output neurons of the model represent the biological effects of the input molecules. The authors used the ChemAI model to predict the inhibitory effects of these molecules on the 3CLpro and PLpro proteases of COVID-19. These molecules have a binding, inhibitory, and toxic effect on the targets. A list of COVID-19 drug development methods based on AI technology is provided in Table 4.

4.2 Vaccine Development

Currently, there are three types of COVID-19 vaccine candidates, such as whole virus vaccines, recombinant protein subunit vaccines, and nucleic acid vaccines [34, 226]. AI technology has participated in the design and development of COVID-19 vaccine. Compared with the explicit applications in other fields, AI technology is usually used in the sub-processes of vaccine development in an implicit manner.

The AI algorithms of netMHC and netMHCpan are used to develop COVID-19 vaccines for epitope prediction [66, 88, 206]. In Reference [66], Herst et al. obtained the SARS-CoV-2 protein sequences from GenBank and used the MSA algorithm to trim the nucleocapsid phosphoprotein sequences into possible peptide sequences. On this basis, they used netMHC and netMHCpan AI algorithms to train and predict peptide sequences [7, 88]. The pan variant of netMHC integrates the *in vitro* objects of 215 HLAs for prediction. Finally, they used the average value of the ANN, SVM, netMHC and netMHCpan methods to calculate vaccine candidates. In Reference [206], Ward et al. downloaded the SARS-CoV-2 nucleotide sequences from NCBI [46] and GISAID [54] databases, and generated a consensus sequence for each SARS-CoV-2 protein. The sequences can be used as references for prediction, specificity, and epitope mapping analysis. Next, the authors used different epitope prediction tools to predict B cell epitopes and map them to the amino acid sequences of each gene. On this basis, they used the AI-based netMHCpan algorithm to predict HLA-1 peptides and obtained a total of 2,915 alleles in all peptide lengths. The BLASTp tool [5] was used to map the short amino acid epitope sequences to the canonical sequences of SARS-CoV-2 proteins. Finally, the authors provided an online tool that provides functions of SARS-CoV-2 genetic variation analysis, epitope prediction, coronavirus homology analysis, and candidate proteome analysis.

In Reference [136], Ong et al. used ML and Reverse Vaccinology (RV) methods to predict and evaluate potential vaccines for COVID-19. They used RV to analyze the bioinformatics of pathogen genomes to identify promising vaccine candidates. They obtained the SARS-CoV-2 sequences and all the proteins of six known human coronavirus strains from the NCBI [46] and UniProt [13] databases. Then, they used Vaxign and Vaxign-ML [63, 135] to analyze the complete proteome of the coronavirus and predicted its service biological characteristics. Next, they used LR, SVM, KNN, RF, and XGBoost methods to improve the Vaxign-ML model, and predicted the protein levels of all SARS-CoV-2 proteins. The nsp3 protein was selected for phylogenetic analysis, and the immunogenicity of nsp3 was evaluated by predicting T cell MHC-I and MHC-II and linear B-cell epitopes.

In Reference [151], Qiao et al. used DL to predict the patient's neoantigen mutation and identified the best T-cell epitope for the peptide-based COVID-19 vaccines. They first sequenced the diseased cells in the patient's blood and extracted six **human leukocyte antigen (HLA)** types and **T-cell receptor (TCR)** sequences. Then, they proposed the DeepNovo model to train patient's immune peptides and to identify the best T-cell epitope set based on a person's HLA alleles and immune peptide group information. The DeepNovo model uses LSTM and RNN structures to capture sequence patterns in peptides or proteins, and predicts HLA peptides from conserved regions of the virus, thereby predicting new mutant antigens in patients. In addition, they used the IEDB tool [196] to predict the immunogenicity of 177 peptides. They suggested designing an epitope-based COVID-19 vaccine specifically for each person based on his/her HLA alleles.

The prediction of immune stimulation ability is an important part of vaccine design [155, 159]. Different ML methods and **position-specific scoring matrices (PSSM)** are usually used to predict epitope and immune interactions, thereby predicting the production of adaptive immunity in the target host. In Reference [155], Rahman et al. used immuno-informatics and comparative genomic methods to design a multi-epitope peptide vaccine against SARS-CoV-2, which combines

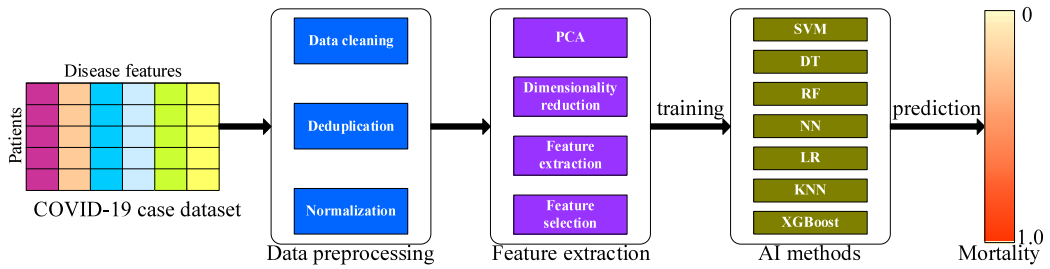


Fig. 10. Representative AI architecture used for patient mortality and survival rate prediction.

the epitopes of S, M, and E proteins. They used the Ellipro antibody epitope prediction tool [80] to predict linear B-cell epitopes on the S protein. In addition, Sarkar et al. studied the epitope-based vaccine design against COVID-19 and used the SVM method to predict the toxicity of the selected epitopes [164]. In Reference [146], Prachar et al. used 19 epitope-HLA combination prediction tools including IEDB, ANN, and PSSM algorithms to predict and verify 174 SARS-CoV-2 epitopes.

5 EPIDEMIC AND TRANSMISSION PREDICTION

Thanks to the developed information and multimedia technology, the outbreak and spread of COVID-19 were reported timely and accurately. The number of suspected, confirmed, cured, and dead COVID-19 cases in each country/region is announced in real-time. In addition, passenger travel trajectories and related big data are shared for scientific research. Based on a wealth of data, numerous researchers have participated in the prediction, spread, and tracking of the COVID-19 outbreak.

5.1 Patient Mortality and Survival Rate Prediction

Researchers collected clinical COVID-19 case data and used different AI methods to extract important features and to predict the mortality and survival rate of patients with COVID-19. The representative AI architecture used for patient mortality and survival rate prediction is shown in Figure 10.

In Reference [145], Pourhomayoun et al. used six AI methods to predict the mortality of patients with COVID-19. They used public data of patients with COVID-19 from 76 countries/regions around the world [214], and counted 112 features, including 80 medical annotations and disease features and 32 features from the patients' demographic and physiological data. Based on the filtering method and wrapper method, 42 best features were extracted, such as demographic features, general medical information, and patient symptoms. On this basis, six AI methods (such as SVM, NN, RF, DT, LR, and KNN) are used to predict the mortality of patients with COVID-19. In Reference [165], Sarkar et al. used the RF model to analyze the records of 433 patients with COVID-19 from Kaggle [39], and identified the important features and their impact on mortality. Experimental results show that patients over 62 years of age have a higher risk of death. In References [217, 218], Yan et al. analyzed a blood sample data set of 404 patients with COVID-19 in Wuhan, China, and used the XGBoost classification method [33] to select three important biomarkers to predict the survival rate of individual patients. Experimental results with an accuracy of 90% indicate that higher LDH levels seem to play an important role in distinguishing the most critical COVID-19 cases.

5.2 Outbreak and Transmission Prediction

BlueDot [17] and Metabiota [124] are two AI companies that made accurate predictions for the COVID-19 outbreak. BlueDot collected large-scale heterogeneous data from various sources, such

Table 5. AI Methods used for COVID-19 Outbreak and Transmission Prediction

Literature	Data sources	Methods	Country/region
Huang [76]	Yang [220], WHO [207]	CNN, LSTM, MLP, GRU	China
Hu [74, 75]	The Paper [142], WHO [207]	MAE, clustering	China
Yang [222]	Baidu [12]	SEIR, LSTM	China
Fong [44, 45]	NHC [132]	SVM, PNN	China
Ai [3]	WHO [47, 207]	ANFIS, FPA	China, USA
Rizk [160]	WHO [207]	ISACL-MFNN	USA, Italy, Spain
Giuliani [55]	Italy [137]	EMTMGL	Italy
Marini [121, 122]	Swiss population	Enerpol	Switzerland
Lai [103]	IATA [118], Worldpop [209]	ML	Global
Punn [148]	JHU CSSE [43]	SVR, PR, DNN, LSTM, RNN	Global
Lamos [104]	MediaCloud [123], PHE [57], ECDC [48]	Transfer learning	Global

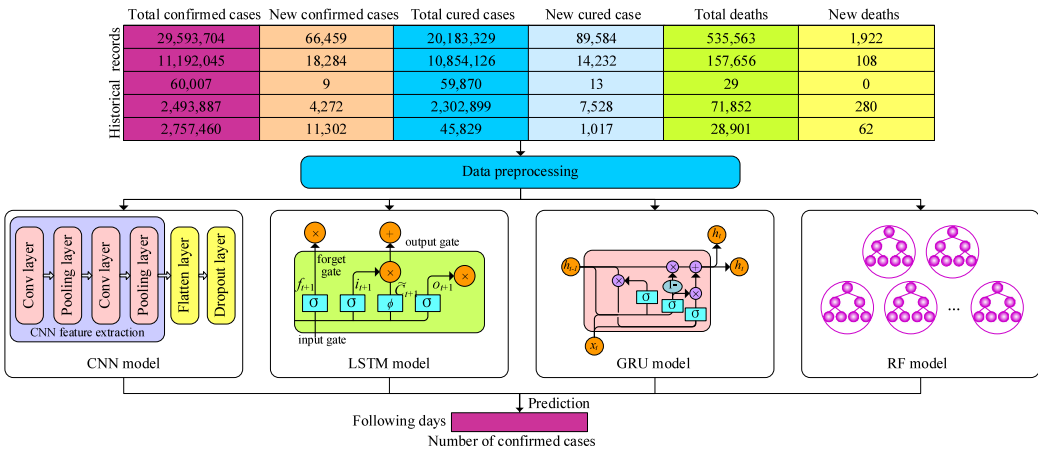


Fig. 11. Representative AI architecture used for COVID-19 outbreak prediction.

as news reports, global ticketing data, animal diseases, global infectious disease alerts, and real-time climate conditions. Then, it used filtering tools to narrow its focus; used various ML and **Natural Language Processing (NLP)** techniques to detect, mark, and display the potential risk frequency of COVID-19; and finally predicted the outbreak time of transmission. It is worth mentioning that 9 days before the official announcement of the COVID-19 outbreak, BlueDot has accurately predicted the epidemic of COVID-19 and cities with a high risk of virus outbreaks. Metabiota collected large-scale data from social and nonsocial sources (such as biology, socio-economic, political, and environmental data), and used technologies such as AI, ML, big data, and NLP to accurately predict the outbreak, spread, and intervention measures of COVID-19. More AI-based methods used for COVID-19 outbreak and transmission prediction are shown in Table 5.

Although the source of the COVID-19 epidemic has not yet been identified, it was first reported in Wuhan, China. Therefore, the outbreak and spread of COVID-19 in China have received extensive attention. In Reference [76], Huang et al. used 4 DL models (such as CNN, LSTM, GRU, and MLP) to train and predict COVID-19 cases from seven severely epidemic cities in China. The input of these DL models are the features of COVID-19 cases, including the number of confirmed cases, cured cases, and deaths. Based on the input of the previous 5 days, each model can predict the number of COVID-19 cases in the following few days. The representative AI architecture used for the COVID-19 outbreak prediction is shown in Figure 11.

In References [74, 75], Hu et al. used AI methods such as MAE and clustering algorithms to predict the number of confirmed COVID-19 cases in different provinces and cities in China. They divided 34 provinces and cities in China into nine groups based on the prediction results, and further predicted the spread of COVID-19 between provinces and cities. In Reference [222], Yang et al. used the SEIR model [92] and the LSTM model to predict COVID-19 in China. The population migration data and the latest COVID-19 epidemiological data from Baidu [12] were input into the SEIR model to derive the epidemic curves. In addition, they used the 2003 SARS data to pre-train the LSTM model to predict COVID-19 in the following few days, in which epidemiological parameters (such as the transmission, incubation, recovery probability, and the number of deaths) were selected as input features. Both SEIR and LSTM models predict the daily peak of infection in the first week of February will be 4,000. In References [44, 45], Fong et al. obtained early COVID-19 epidemiological data from NHC [132]. Then, they used traditional time series data analysis methods (such as ARIMA, Exponential, and Holt-Winters), ML methods (such as KR, SVM, and DT), and AI methods (such as PNN) to analyze and predict future outbreaks.

In addition to China, the outbreak and spread of COVID-19 in other countries (including the United States, Italy, Spain, Iran, and Switzerland) have also received widespread attention. In Reference [3], Ai et al. proposed an improved ANFIS method [85] to predict the number of COVID-19 cases. The proposed system connects fuzzy logic and neural networks, and uses an enhanced flower pollination algorithm [221] for model parameter optimization and model training. In Reference [160], Rizk et al. proposed an improved **Multi-layer Feed-forward Neural Network (ISACL-MFNN)** model, which uses an **internal search algorithm (ISA)** to optimize model parameters and a CL strategy to enhance ISA performance. From the official COVID-19 data set reported by the WHO [207], data from January 22, 2020, to April 3, 2020, in the United States, Italy, and Spain were collected to train the ISACL-MFNN model and to predict the confirmed cases within the next 10 days. In Reference [55], Giuliani et al. collected the number of infected people in various provinces of Italy [137], and used the EMTMGL model to simulate and predict the spatial and temporal distribution of COVID-19 infection in Italy. They collected daily epidemic data and saved them in a time series data format, and then used LR and LSTM models to make predictions, thereby obtaining the outbreak and spread trend of COVID-19 in Iran. In References [121, 122], Marini et al. developed an agent-based AI platform, which accepts the entire Swiss population as input data to simulate and predict the spread of COVID-19 in Switzerland. It simulates people's daily trajectories by calibrating micro-census data, and effectively predicts individual contacts and possible transmission routes.

Many studies have likewise focus on predicting the spread of COVID-19 worldwide. They collected a large amount of travel data, mobile phone data, and social media data, and used AI methods to accurately predict the potential spread and transmission route of COVID-19. In Reference [103], Lai et al. collected a large amount of travel and mobile phone data from Reference [209], and constructed a corresponding model to predict the risk of COVID-19 transmission in different countries. On this basis, they established an air travel network model between domestic cities and cities in other countries to predict risk cities at home and abroad. In Reference [148], Punn et al. used 2 ML models (e.g., SVR [11] and PR [40]) and 3 DL regression models (e.g., DNN, LSTM [59], and RNN) to predict real-time COVID-19 cases. In Reference [104], Lampos et al. used an automatic crawling tool to obtain daily confirmed COVID-19 case data and related articles from online media, such as MediaCloud [123], Public Health England (PHE) [57], and European Centre for Disease Prevention and Control (ECDC) [48]. They used the TL strategy to transfer the COVID-19 model of the disease-spreading country that are still in the early stage of the epidemic curve, achieving the epidemic prediction in the target country. In addition, companies such as Microsoft Bing [16], Google

[193], and Baidu [12] have aggregated multiple available data sources and developed a COVID-19 global tracking system to provide a visual tracking interface.

In addition to AI methods, various methods based on statistics and epidemiology are also used to predict the outbreak and spread of COVID-19. In Reference [62], He et al. collected the highest viral load in the pharyngeal swabs of 94 patients with confirmed COVID-19. They fitted a generalized additive model with identification links and smooth spline curves to analyze their overall trend. A gamma distribution was fitted to the transmission pair data to evaluate the serial interval distribution. The results of statistical analysis indicated that the patients with confirmed COVID-19 have reached the peak of virus shedding before or during symptom onset, and some kinds of transmission may occur before the initial symptoms appear. In Reference [200], Wang et al. determined a set of technical indicators that reflect the infection status of COVID-19, such as the number of infection cases in the hospital, daily infection rate, and daily cured rate. Next, they proposed a calculation method based on statistical theory to quantify the iconic characteristics of each period and predict the turning point in the development of the epidemic. In addition, numerous studies based on the **Susceptible-Infected-Recovered (SIR)** and SEIR models have studied the spread of COVID-19 from an epidemiological perspective. Please see References [21, 112, 161, 178, 187, 191, 223] for more information.

5.3 Social Control

When COVID-19 appeared, most countries in the world adopted different forms of social control, social alienation, school closures, and blockade measures to prevent the spread of the epidemic [195]. AI technologies have been widely used in epidemic control and social management, including individual temperature detection, video tracking, contact tracking, intelligent robots, and so on. Many countries have used smart devices equipped with AI to detect suspicious persons in public transportation places, such as airports and train stations. For example, infrared cameras are used to scan high temperatures in a crowd, and different AI methods perform efficient analysis to detect whether an individual is wearing a mask in real time. In addition, DL-based video tracking technology was used to detect and track suspicious COVID-19 patients in public places [28]. Moreover, at the entrances and exits of cities, the identity information of each passing person was collected. Then, AI-based systems are used to efficiently query the travel history and trajectory of each passing individual to check whether they are from areas seriously affected by COVID-19 [31, 189].

AI technology is also used for contact tracking of patients with COVID-19 [91]. For each patient with confirmed COVID-19, personal data (such as mobile phone positioning data, consumption records, and travel records) can be integrated to identify potential transmission trajectories [50]. In addition, when people are in a state of social isolation, mobile phone positioning and AI frameworks can assist the government better understand the status of individuals [158]. Intelligent robots are used to perform on-site disinfection and product transfer, and mobile phone positioning functions are used to detect and track the distribution and flow of personnel. During the COVID-19 pandemic, Bluetooth technology was used in contact tracing APPs and deployed in many countries [72]. Moreover, various countries/regions such as China, Canada, Singapore, Australia, France, Germany, and the United Kingdom, have deployed different contact tracing systems to effectively track public contacts and locate persons suspected of COVID-19 [97]. For example, in China, a system of QR code was used to track the infected or potentially infected people [100]. The Canadian government has published several Apps for COVID-19 contact tracing, such as the COVID Alert App, Canada COVID-19 self-assessment tool, and ArriveCAN [22]. In Singapore, a real-time locating system (TraceTogether App) was developed to track public contacts during the COVID-19 pandemic [68].

Another group of studies focuses on the impact of various social control strategies on the spread of COVID-19. In Reference [64], Hellewell et al. proposed a random propagation model to analyze the success rate of different social control approaches to prevent the spread of COVID-19. By quantifying parameters such as contact tracking and quarantined cases, the model analyzes the maximum number of cases tracked every week and measures the feasibility of public health work. In Reference [96], Kissler et al. established a mathematical model to assess the impact of interventions on the prevalence of COVID-19 in the United States, and recommended the use of intermittent grooming measures to maintain effective control of COVID-19. Koo et al. [99] applied the population and personal behavior data of Singapore to the influenza epidemic simulation model and assessed different social isolation strategies for the dynamic transmission of COVID-19. Similarly, Chang et al. [25] constructed a simulation model to evaluate the propagation and control of COVID-19 in Australia. The model analyzes the propagation characteristics of COVID-19 and the impact of different control strategies on the propagation results.

6 DATA AND RESOURCES

The implementation and performance improvement of AI greatly depends on a large-scale available data and resources. Therefore, we compiled available public resources that can be used for COVID-19 disease diagnosis, virology research, drug and vaccine development, and epidemic and transmission prediction. Three types of data and resources are summarized, including medical images, biological data, and informatics resources. Due to space limitation, the detailed description of data and resources is provided in the supplementary file.

7 CHALLENGES AND FUTURE DIRECTIONS

Although AI technologies have been used in fighting against COVID-19 and many studies have been published, we note that the application and contribution of AI in this work is still relatively limited. We summarize the main challenges currently faced by AI against COVID-19 and provide the corresponding suggestions.

7.1 Challenges

At present, the application of AI in COVID-19 research mainly faces five challenges:

- Lack of available large-scale training data. Most AI methods rely on large-scale annotated training data, including medical images and various biological data. However, due to the rapid outbreak of COVID-19, there is insufficient data available for AI. In addition, annotating training samples is very time-consuming and requires professional medical personnel.
- Data imbalance between positive and negative samples. There is a serious proportional imbalance between positive COVID-19 samples and negative samples (data resources without COVID-19). For data sets of medical images, biology samples, and cough sounds, negative samples are relatively easy to obtain from historical publications. However, there are limited positive COVID-19 samples can be collected, which further affects the accuracy and robustness of AI algorithms in COVID-19 diagnosis.
- Massive noisy data and rumors. Relying on the developed mobile Internet and social media, massive noise information and fake news about COVID-19 have been published on various online media without rigorous review. However, AI algorithms seem to be powerless in judging and filtering the noise and erroneous data. This problem limits the accuracy and application of AI algorithms, especially in epidemic prediction and transmission analysis.
- Limited knowledge in the intersection of computer science and medicine. Many AI scientists are from computer science, but the application of AI in the COVID-19 battle requires

in-depth cooperation in computer science, medical imaging, bioinformatics, virology, and many other disciplines. Therefore, it is crucial to coordinate the cooperative work of researchers from different fields and integrate the knowledge of multiple subjects to jointly respond to COVID-19.

- Data privacy and human rights protection. In the era of big data and AI, the cost of obtaining personal privacy data is very low. Faced with public health issues such as COVID-19, many governments hope to obtain various types of personal information, including mobile phone positioning data, personal travel trajectory data, and patient disease data. How to effectively protect personal privacy and human rights during information acquisition and AI-based processing is an issue worthy of discussion and attention.

7.2 Future Directions

In addition to the applications investigated in this survey, AI can also contribute to the COVID-19 battle from the following 11 potential directions:

- (1) Non-contact disease detection. In CXR and CT image detection, the use of non-contact automatic image acquisition can effectively avoid the risk of infection between radiologists and patients during the COVID-19 pandemic. AI can be used for patient posture positioning, standard section acquisition of CXR and CT images, and movement of camera equipment.
- (2) Few-shot learning and transfer learning. Focusing on the serious imbalance between the positive COVID-19 samples and negative samples, few-shot Learning, meta-learning, and transfer learning have the potential to be further applied in COVID-19 research [37, 128]. After learning a large amount of data in a certain category, few-shot Learning algorithms can quickly detect a new category and obtain high accuracy by learning only a small number of samples. Transfer learning relaxes the assumption that training data must be independent and identically distributed (i.i.d) with test data, and effectively solve the problem of insufficient training data and imbalanced sample categories.
- (3) Remote video diagnosis. AI and NLP technologies can be used to develop remote video diagnosis systems and chat robot systems, and provide the public with COVID-19 disease consultation and preliminary diagnosis.
- (4) Patient prognosis management. In addition to long-term tracking and management of patients with COVID-19, AI technology (such as intelligent image and video analysis) can also be used to automatically monitor patient behavior during follow-up monitoring and prognostic management.
- (5) Biological research. In the field of biological research, AI can be used to discover the protein structures and features of viruses by accurately analyzing biomedical information, such as large-scale protein structures, gene sequences, and viral trajectories.
- (6) Drug and vaccine development. AI can be used not only to discover potential drugs and vaccines but also to simulate the interaction between drugs and proteins and between vaccines and receptors, thereby predicting the potential responses of patients with COVID-19 with different constitutions to drugs and vaccines.
- (7) Identification and filtering of fake news. AI can be used to reduce and eliminate fake news and noise data on online social media platforms to provide reliable, correct, and scientific information about the COVID-19 pandemic.
- (8) Impact simulation and evaluation. Various simulation models can use AI to analyze the impact of different social control strategies on disease transmission. Then, they can be used to explore more effective and scientific disease prevention and social control methods.

- (9) Patient contact tracking. By constructing social relationship networks and knowledge graphs, AI can identify and track the trajectories of people in close contact with patients with COVID-19, to accurately predict and control the potential spread of the disease.
- (10) Intelligent robots. Intelligent robots are expected to be used in applications such as disinfection and cleaning in public places, product distribution, and patient care.
- (11) Intelligent Internet of Things. AI is expected to be combined with the Internet of Things to be deployed in customs, airports, railway stations, bus stations, and business centers. In this case, we can quickly identify suspicious COVID-19 virus and patients through intelligent monitoring of the environment and personnel.

8 CONCLUSIONS

In this survey, we investigated the main scope and contributions of AI in combating COVID-19. Compared with the pandemic of SARS-CoV in 2003 and MERS-CoV in 2012, AI technologies have been successfully applied to almost every corner of the COVID-19 battle. The application of AI in COVID-19 research can be summarized in four aspects, such as disease detection and diagnosis, virological research, drug and vaccine development, and epidemic and transmission prediction. Among them, medical image analysis, drug discovery, and epidemic prediction are the main battlefields of AI against COVID-19. We also summarized the currently available data and resources for AI-based COVID-19 research, including medical imaging data, biological data, and informatics resources. Finally, we highlighted the main challenges and potential directions in this field. This survey provides medical and AI researchers with a comprehensive view of the current and potential contributions of AI in combating COVID-19, with the goal of inspiring them to continue to maximize the advantages of AI and big data to combat this pandemic.

REFERENCES

- [1] T. Abel and D. McQueen. 2020. The COVID-19 pandemic calls for spatial distancing and social closeness: Not for social distancing. *Int. J. Public Health* 65, 3 (2020), 231–231.
- [2] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi. 2020. COVID-caps: A capsule network-based framework for identification of COVID-19 cases from X-ray images. *Pattern Recogn. Lett.* 138 (2020), 638–643.
- [3] M. Al, A. Ewees, and H. Fan. 2020. Optimization method for forecasting confirmed cases of COVID-19 in China. *J. Clin. Med.* 9, 3 (2020), 674.
- [4] Z. Allam, G. Dey, and D. Jones. 2020. Artificial intelligence (AI) provided early detection of the coronavirus (COVID-19) in China and will influence future urban health policy internationally. *AI* 1, 2 (2020), 156–165.
- [5] S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman. 1990. Basic local alignment search tool. *J. Molec. Biol.* 215, 3 (1990), 403–410.
- [6] K. Andersen, A. Rambaut, W. Lipkin, E. Holmes, and R. Garry. 2020. The proximal origin of SARS-CoV-2. *Nature Med.* 26, 4 (2020), 450–452.
- [7] M. Andreatta and M. Nielsen. 2016. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* 32, 4 (2016), 511–517.
- [8] I. Apostolopoulos and T. Mpesiana. 2020. COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* 43, 2 (2020), 635–640.
- [9] Y. Arabi, S. Murthy, and S. Webb. 2020. COVID-19: A novel coronavirus and a novel challenge for critical care. *Intens. Care Med.* 46, 5 (2020), 833–836.
- [10] K. Arnold, L. Bordoli, and J. Kopp. 2006. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22, 2 (2006), 195–201.
- [11] M. Awad and R. Khanna. 2015. Support vector regression. In *Efficient Learning Machines*. Apress, 67–80.
- [12] Baidu. 2020. Real-time COVID-19 data. Retrieved from <https://voice.baidu.com/act/newpneumonia>.
- [13] A. Bairoch, R. Apweiler, C. Wu, and W. Barker. 2005. The universal protein resource (UniProt). *Nucleic Acids Res.* 33, 1 (2005), D154–D159.
- [14] B. Beck, B. Shin, Y. Choi, and S. Park. 2020. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* 18 (2020), 784–790.

- [15] H. Berman, K. Henrick, and H. Nakamura. 2003. Announcing the worldwide protein data bank. *Nature Struct. Molec. Biol.* 10, 12 (2003), 980–980.
- [16] Microsoft Bing. 2020. COVID-19 tracker). Retrieved from <https://bing.com/covid>.
- [17] BlueDot. 2020. An AI epidemiologist sent the first warnings of the Wuhan virus. Retrieved from <https://bluedot.global>.
- [18] L. Breiman. 2001. Random forests. *Mach. Learn.* 45, 1 (2001), 5–32.
- [19] N. Bung, S. Krishnan, G. Bulusu, and A. Roy. 2020. De novo design of new chemical entities (NCEs) for SARS-CoV-2 using artificial intelligence.
- [20] C. Audrey. 2020. COVID-19 chest X-ray dataset initiative. Retrieved from <https://github.com/agchung/Figure1-COVID-chestxray-dataset>.
- [21] D. Caccavo. 2020. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. medRxiv, 2020.
- [22] Canada. 2020. Digital government response to COVID-19. Retrieved from <https://www.canada.ca/en/government/system/digital-government>.
- [23] I. Castiglioni, D. Ippolito, and M. Interlenghi. 2020. Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: A first experience from Lombardy, Italy. MedRxiv, 2020.
- [24] J. Chan, S. Yuan, K. Kok, and K. To. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395, 10223 (2020), 514–523.
- [25] S. Chang, N. Harding, C. Zachreson, and O. Cliff. 2020. Modelling transmission and control of the COVID-19 pandemic in Australia. *Nature Commun.* 11, 1 (2020), 1–13.
- [26] T. Chebet, Y. Li, N. Sam, and Y. Liu. 2019. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161 (2019), 272–279.
- [27] H. Chen, J. Guo, C. Wang, F. Luo, and X. Yu. 2020. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. *Lancet* 395, 10226 (2020), 809–815.
- [28] J. Chen, K. Li, Q. Deng, K. Li, and Y. Philip. 2019. Distributed deep learning model for intelligent video surveillance systems with edge computing. *IEEE Trans. Industr. Info.* 99, 1 (2019), 1–12.
- [29] J. Chen, L. Wu, and J. Zhang. 2020. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. *Sci. Rep.* 10, 1 (2020), 1–11.
- [30] Q. Chen, A. Allot, and Z. Lu. 2020. Keep up with the latest coronavirus research. *Nature* 579, 7798 (2020), 193–193.
- [31] S. Chen, J. Yang, W. Yang, and C. Wang. 2020. COVID-19 control in China during mass population movements at New Year. *Lancet* 395, 10226 (2020), 764–766.
- [32] S. Chen, Z. Zhang, J. Yang, J. Wang, and X. Zhai. 2020. Fangcang shelter hospitals: A novel concept for responding to public health emergencies. *Lancet* 395, 10232 (2020), 1305–1314.
- [33] T. Chen and C. Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*. 785–794.
- [34] W. Chen, U. Strych, P. Hotez, and M. Bottazzi. 2020. The SARS-CoV-2 vaccine pipeline: An overview *Current Tropical Medicine Reports*. 7, 2 (2020), 61–64.
- [35] F. Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR'17)*. 1251–1258.
- [36] J. Cohen, P. Morrison, and L. Dao. 2020. COVID-19 image data collection: Prospective predictions are the future. Retrieved from <https://github.com/ieee8023/covid-chestxray-dataset>.
- [37] Joseph Paul Cohen, Lan Dao, Karsten Roth, Paul Morrison, Yoshua Bengio, Almas F. Abbasi, Beiyi Shen, Hoshmand Kochi Mahsa, Marzyeh Ghassemi, Haifang Li, et al. 2020. Predicting COVID-19 pneumonia severity on chest x-ray with deep learning. *Cureus* 12, 7 (2020).
- [38] Coronacases. 2020. CT images of confirmed COVID-19 cases. Mendeley Data. Retrieved from <https://coronacases.org>.
- [39] D. Su. 2020. Novel corona virus 2019 dataset. Retrieved from <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>.
- [40] Y. Decastro, F. Gamboa, D. Henrion, and R. Hess. 2019. Approximate optimal designs for multivariate polynomial regression. *Ann. Stat.* 47, 1 (2019), 127–155.
- [41] J. Degen, C. Wegscheid, A. Zaliani, and M. Rarey. 2008. On the art of compiling and using 'drug-like' chemical fragment spaces. *ChemMedChem: Chem. Enabl. Drug Discov.* 3, 10 (2008), 1503–1507.
- [42] M. Demirci and A. Adan. 2020. Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection. *PeerJ* 8 (2020), e9369.
- [43] E. Dong, H. Du, and L. Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Diseases* 20, 5 (2020), 533–534.

- [44] S. Fong, G. Li, N. Dey, and R. Crespo. 2020. Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. *Appl. Soft Comput.* 93 (2020), 106282.
- [45] S. Fong, G. Li, N. Dey, and R. Crespo. 2020. Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak. *Int. J. Interact. Multimedia Artific. Intell.* 6, 1 (2020), 132–140.
- [46] National Center for Biotechnology Information (NCBI). 2020. Genome sequencing data of SARS-CoV-2. Retrieved from <https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>.
- [47] Centers for Disease Control and Prevention. 2020. Weekly influenza confirmed cases. Retrieved from <https://www.cdc.gov/flu/weekly>.
- [48] European Centre for Disease Prevention and Control. 2020. Geographic distribution COVID-19 cases worldwide. Retrieved from <https://ecdc.europa.eu/en/publications-data>.
- [49] A. Gaulton, L. Bellis, A. Bento, J. Chambers, and M. Davies. 2012. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1 (2012), D1100–D1107.
- [50] R. Ge, Y. Qi, Y. Yan, M. Tian, and Q. Gu. 2020. The role of close contacts tracking management in COVID-19 prevention: A cluster investigation in Jiaying, China. *J. Infect.* 81, 1 (2020), e71–e74.
- [51] Global Health Drug Discovery Institute (GHDDI). 2020. Targeting COVID-19: GHDDI info sharing portal. Retrieved from <https://ghddi-ailab.github.io/Targeting2019-nCoV>.
- [52] I. Ghinai, T. McPherson, J. Hunter, and H. Kirking. 2020. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. *Lancet* 395, 10230 (2020), 1137–1144.
- [53] M. Gilman, C. Liu, A. Fung, and I. Behera. 2019. Structure of the respiratory syncytial virus polymerase complex. *Cell* 179, 1 (2019), 193–204.
- [54] GISAIID. 2020. GISAIID: global initiative on sharing all influenza data. Retrieved from <https://www.gisaid.org>.
- [55] D. Giuliani, M. Dickson, G. Espa, and F. Santi. 2020. Modelling and predicting the spatio-temporal spread of Coronavirus disease 2019 (COVID-19) in Italy. *BMC Infect. Diseases* 20, 1 (2020), 1–10.
- [56] R. Gómez, J. Wei, D. Duvenaud, and J. Hernández. 2018. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Sci.* 4, 2 (2018), 268–276.
- [57] GOV.UK. 2020. Coronavirus (COVID-19) cases in the UK. Retrieved from <https://coronavirus.data.gov.uk/>.
- [58] O. Gozes, M. Frid, and H. Greenspan. 2020. Rapid ai development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection and patient monitoring using deep learning CT image analysis. Retrieved from <https://arXiv:2003.05037>.
- [59] K. Greff, R. Srivastava, K. Jan, and B. Steunebrink. 2016. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* 28, 10 (2016), 2222–2232.
- [60] A. Hassaniien, L. Mahdy, K. Ezzat, and H. Elmousalami. 2020. Automatic X-ray COVID-19 lung image classification system based on multi-level thresholding and support vector machine. MedRxiv, 2020.
- [61] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR'16)*. 770–778.
- [62] X. He, E. Lau, P. Wu, X. Deng, and J. Wang. 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Med.* 26, 5 (2020), 672–675.
- [63] Y. He, Z. Xiang, and H. Mobley. 2010. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *BioMed Res. Int.* 2010, Article 297505 (2010).
- [64] J. Hellewell, S. Abbott, A. Gimma, and N. Bosse. 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* 8, 4 (2020), e488–e496.
- [65] E. Hemdan, M. Shouman, and M. Karar. 2020. Covidx-net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images. Retrieved from <https://arXiv:2003.11055>.
- [66] C. Herst, S. Burkholz, J. Sidney, A. Sette, P. Harris, S. Massey, T. Brasel, E. Cunha, D. S. Rosa, and W. Chao. 2020. An Effective CTL Peptide Vaccine for Ebola Zaire Based on Survivors' CD8+ Targeting of a Particular Nucleocapsid Protein Epitope with Potential Implications for COVID-19 Vaccine Design. *Vaccine* 38, 28 (2020), 4464–4475.
- [67] G. Hinton, S. Sabour, and N. Frosst. 2018. Matrix capsules with EM routing. In *International Conference on Learning Representations*. 1–8.
- [68] Hanley J. Ho, Zoe Xiaozhu Zhang, Zhilian Huang, Aung Hein Aung, Wei-Yen Lim, and Angela Chow. 2020. Use of a real-time locating system for contact tracing of health care workers during the COVID-19 pandemic at an infectious disease center in Singapore: validation study. *J. Med. Internet Res.* 22, 5 (2020), e19437.
- [69] S. Hochreiter and J. Schmidhuber. 1997. Long short-term memory. *Neural Comput.* 9, 8 (1997), 1735–1780.
- [70] M. Hoffmann, H. Kleine, S. Schroeder, and N. Krüger. 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 2 (2020), 271–280.
- [71] M. Hofmarcher, A. Mayr, E. Rumetshofer, and P. Ruch. 2020. Large-scale ligand-based virtual screening for SARS-CoV-2 inhibitors using deep neural networks. *Soc. Sci. Res. Netw.* (2020). <https://ssrn.com/abstract=3561442>.

- [72] Jeremy Hsu. 2020. Can AI Make Bluetooth Contact Tracing Better? IEEE Spectrum. Retrieved from <https://spectrum.ieee.org/the-human-os/artificial-intelligence>.
- [73] F. Hu, J. Jiang, and P. Yin. 2020. Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model. Retrieved from <https://arXiv:2003.00728>.
- [74] Z. Hu, Q. Ge, L. Jin, and M. Xiong. 2020. Artificial intelligence forecasting of COVID-19 in China. Retrieved from <https://arXiv:2002.07112>.
- [75] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong. 2020. Evaluating the effect of public health intervention on the global-wide spread trajectory of COVID-19. Medrxiv, 2020.
- [76] C. Huang, Y. Chen, Y. Ma, and P. Kuo. 2020. Multiple-input deep convolutional neural network model for COVID-19 forecasting in China. *IEEE Access* 2019, 7 (2019), 74822–74834.
- [77] C. Huang, Y. Wang, X. Li, L. Ren, and J. Zhao. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 10223 (2020), 497–506.
- [78] G. Huang, Z. Liu, D. Van, and K. Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR'17)*. 4700–4708.
- [79] L. Huang, R. Han, T. Ai, P. Yu, and H. Kang. 2020. Serial quantitative chest CT assessment of COVID-19: Deep-learning approach. *Radiol.: Cardiothor. Imag.* 2, 2 (2020), e200075.
- [80] IEDB. 2020. Ellipro: An antibody epitope prediction tool. Retrieved from <http://tools.iedb.org/ellipro>.
- [81] GitHub. Inc. 2020. GitHub. Retrieved from <https://github.com>.
- [82] Clarivate Analytics Integrity. 2020. Clarivate analytics integrity. Retrieved from <https://integrity.clarivate.com/integrity/>.
- [83] M. Iqbal. 2020. Active surveillance for COVID-19 through artificial intelligence using concept of real-time speech-recognition mobile application to analyse cough sound. In *IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*. IEEE, 1–2.
- [84] S. Jaeger, S. Fulle, and S. Turk. 2018. Mol2vec: Unsupervised machine learning approach with chemical intuition. *J. Chem. Info. Model.* 58, 1 (2018), 27–35.
- [85] J. Jang. 1993. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybernet.* 23, 3 (1993), 665–685.
- [86] H. Jeffrey. 1990. Chaos game representation of gene structure. *Nucleic Acids Res.* 18, 8 (1990), 2163–2170.
- [87] J. Jumper, K. Tunyasuvunakool, P. Kohli, and D. Hassabis. 2020. Computational predictions of protein structures associated with COVID-19. Retrieved from <https://deepmind.com>.
- [88] V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, and M. Nielsen. 2017. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* 199, 9 (2017), 3360–3368.
- [89] O. Kadioglu, M. Saeed, H. Johannes, and T. Efferth. 2021. Identification of novel compounds against three targets of SARS CoV-2 coronavirus by combined virtual screening and supervised machine learning. *Computers in Biology and Medicine* 133 (2021), 104359.
- [90] N. Kandel, S. Chungong, A. Omaar, and J. Xing. 2020. Health security capacities in the context of COVID-19 outbreak: An analysis of international health regulations annual report data from 182 countries. *Lancet* 395, 10229 (2020), 1047–1053.
- [91] M. Keeling, T. Hollingsworth, and J. Read. 2020. The efficacy of contact tracing for the containment of the 2019 novel coronavirus(COVID-19). *J Epidemiol. Commun. Health* 74, 10 (2020), 861–866.
- [92] P. Keeling. 2008. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 362.
- [93] D. Kermany, M. Goldbaum, and W. Cai. 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172, 5 (2018), 1122–1131.
- [94] J. Kim, M. Kang, E. Park, and D. Chung. 2019. A simple and multiplex loop-mediated isothermal amplification assay for rapid detection of SARS-CoV. *BioChip J.* 13, 4 (2019), 341–351.
- [95] S. Kim, P. Thiessen, E. Bolton, and J. Chen. 2016. PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1 (2016), D1202–D1213.
- [96] S. Kissler, C. Tedijanto, M. Lipsitch, and Y. Grad. 2020. Social distancing strategies for curbing the COVID-19 epidemic. MedRxiv, 2020.
- [97] Robert A. Kleinman and Colin Merkel. 2020. Digital contact tracing for COVID-19. *Can. Med. Assoc. J.* 192, 24 (2020), E653–E656.
- [98] W. Kong, Y. Li, M. Peng, and D. Kong. 2020. SARS-CoV-2 detection in patients with influenza-like illness. *Nature Microbiol.* 5, 5 (2020), 675–678.
- [99] J. Koo, A. Cook, M. Park, Y. Sun, and H. Sun. 2020. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: A modelling study. *Lancet Infect. Diseases* 20, 6 (2020), 678–688.

- [100] Genia Kostka and Sabrina Habich-Sobiegalla. 2020. In Times of Crisis: Public Perceptions Towards COVID-19 Contact Tracing Apps in China, Germany, and the U.S. *Soc. Sci. Res. Netw.* (2020). <https://ssrn.com/abstract=3693783>.
- [101] A. Kozomara, M. Birgaoanu, and S. Griffiths. 2019. miRBase: From microRNA sequences to function. *Nucleic Acids Res.* 47, D1 (2019), D155–D162.
- [102] A. Krizhevsky, I. Sutskever, and G. Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Adv. Neural Info. Process. Syst.* 25 (2012), 1097–1105.
- [103] S. Lai, I. Bogoch, and N. Ruktanonchai. 2020. Assessing spread risk of Wuhan novel coronavirus within and beyond China, January–April 2020: A travel network-based modelling study. *MedRxiv*, 2020.
- [104] V. Lamos, S. Moura, E. Yom, and I. Cox. 2021. Tracking COVID-19 using online search. *NPJ Dig. Med.* 4, 1 (2021), 1–11.
- [105] Larxel. 2020. COVID-19 X-rays. Retrieved from <https://www.kaggle.com/andrewmvd/convid19-X-rays>.
- [106] G. Li and E. De. 2020. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nature Rev. Drug Discov.* 19, 3 (2020), 149–150.
- [107] L. Li, L. Qin, Z. Xu, and Y. Yin. 2020. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology* (2020), 200905.
- [108] L. Li, Q. Zhang, X. Wang, J. Zhang, and T. Wang. 2020. Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on weibo. *IEEE Trans. Comput. Soc. Syst.* 7, 2 (2020), 556–562.
- [109] S. Li, K. Song, B. Yang, Y. Gao, and X. Gao. 2020. Preliminary assessment of the COVID-19 outbreak using 3-staged model e-ISHR. *J. Shanghai Jiaotong Univ. (Sci.)* 25 (2020), 157–164.
- [110] Y. Li, Z. Zhu, Y. Zhou, Y. Xia, and W. Shen. 2019. Volumetric medical image segmentation: A 3D deep coarse-to-fine framework and its adversarial examples. In *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*. Springer, 69–91.
- [111] J. Liu, R. Cao, M. Xu, X. Wang, and H. Zhang. 2020. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection *in vitro*. *Cell Discov.* 6, 1 (2020), 1–4.
- [112] P. Liu, P. Beeler, and R. Chakrabarty. 2020. COVID-19 progression timeline and effectiveness of response-to-spread interventions across the United States. *MedRxiv*, 2020.
- [113] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. Gilson. 2007. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 35, 1 (2007), D198–D201.
- [114] R. Lu, X. Wu, Z. Wan, Y. Li, and L. Zuo. 2020. Development of a novel reverse transcription loop-mediated isothermal amplification method for rapid detection of SARS-CoV-2. *Viol. Sinica* 35, 3 (2020), 344–347.
- [115] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, and H. Wu. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* 395, 10224 (2020), 565–574.
- [116] M. Paul. 2020. Chest X-ray images (pneumonia). Retrieved from <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia/metadata>.
- [117] L. Maaten and G. Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008), 2579–2605.
- [118] Airlines Magazine. 2020. International air travel association (IATA). Retrieved from <https://www.iata.org/>.
- [119] H. Maghdid, A. Asaad, K. Ghafoor, A. Sadiq, and M. Khan. 2020. Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. Retrieved from <https://arXiv:2004.00038>.
- [120] H. Maghdid, K. Ghafoor, A. Sadiq, K. Curran, and K. Rabie. 2020. A novel AI-enabled framework to diagnose coronavirus COVID 19 using smartphone embedded sensors: Design study. In *IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE, 180–187.
- [121] M. Marini, C. Brunner, N. Chokani, and R. Abhari. 2020. Enhancing response preparedness to influenza epidemics: Agent-based study of 2050 influenza season in Switzerland. *Simul. Model. Pract. Theory* 103 (2020), 102091.
- [122] M. Marini, N. Chokani, and R. Abhari. 2020. COVID-19 epidemic in Switzerland: Growth prediction and containment strategy using artificial intelligence and big data. *MedRxiv*, 2020.
- [123] MediaCloud. 2020. MediaCloud. Retrieved from <https://www.mediacloud.org/>.
- [124] Metabiota. 2020. How AI is battling the coronavirus outbreak. Retrieved from <https://www.metabiota.com>.
- [125] H. Metsky, C. Freije, and T. Kosoko. 2020. CRISPR-based surveillance for COVID-19 using kosoko-thoroddsen, tinnasolveig f-comprehensive machine learning design. *BioRxiv*, 2020.
- [126] H. Mi, A. Muruganujan, and P. Thomas. 2012. PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41, D1 (2012), D377–D386.
- [127] T. Mihara, Y. Nishimura, Y. Shimizu, H. Nishiyama, G. Yoshikawa, H. Uehara, P. Hingamp, S. Goto, and H. Ogata. 2016. Linking virus genomes with host taxonomy. *Viruses* 8, 3 (2016), 66.
- [128] Shervin Minaee, Rahele Kafieh, Milan Sonka, Shakib Yazdani, and Ghazaleh Jamalipour Soufi. 2020. Deep-COVID: Predicting COVID-19 from chest x-ray images using deep transfer learning. *Med. Image Anal.* 65 (2020), 101794.

- [129] M. Moskal, W. Beker, R. Roszak, and E. Gajewska. 2020. Suggestions for second-pass anti-COVID-19 drugs based on the artificial intelligence measures of molecular similarity, shape and pharmacophore distribution. *ChemRxiv*. 2020.
- [130] A. Narin, C. Kaya, and Z. Pamuk. 2020. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Retrieved from <https://arXiv:2003.10849>.
- [131] Radiological Society of North America. 2020. RSNA pneumonia detection challenge. Retrieved from <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>.
- [132] Natioanl Health Commission of the People's Republic of China. 2020. Real-time COVID-19 report. Retrieved from <http://www.nhc.gov.cn/>.
- [133] British Society of Thoracic Imaging. 2020. COVID-19 BSTI imaging database. Retrieved from <https://www.bsti.org.uk/training-and-education/COVID-19-bsti-imaging-database>.
- [134] S. Oh, W. Pedrycz, and B. Park. 2003. Polynomial neural networks architecture: Analysis and design. *Comput. Electric. Eng.* 29, 6 (2003), 703–725.
- [135] E. Ong, H. Wang, M. Wong, and M. Seetharaman. 2020. Vaxign-ML: Supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics* 36, 10 (2020), 3185–3191.
- [136] E. Ong, M. Wong, A. Huffman, and Y. He. 2020. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Front. Immunol.* 11 (2020), 1581.
- [137] Il Sole 24 ore. 2020. The coronavirus datasets in Italy. Retrieved from <https://lab24.ilssole24ore.com/coronavirus>.
- [138] Lara Orlandic, Tomas Teijeiro, and David Atienza. 2020. The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms. Retrieved from <https://arXiv:2009.11644>.
- [139] J. Ortega, M. Serrano, F. Pujol, and H. Rangel. 2020. Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: NN *in silico* analysis. *EXCLI J.* 19 (2020), 410.
- [140] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, and L. Guo. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Commun.* 11, 1 (2020), 1–12.
- [141] X. Pan, X. Da, W. Zhou, and L. Wang. 2020. Identification of a potential mechanism of acute kidney injury during the COVID-19 outbreak: A study based on single-cell transcriptome analysis. *Intens. Care Med.* 46, 6 (2020), 1114–1116.
- [142] The Paper. 2020. The paper news network. Retrieved from <https://www.thepaper.cn>.
- [143] J. Phua, L. Wong, L. Ling, M. Egi, and C. Lim. 2020. Intensive care management of coronavirus disease 2019 (COVID-19): Challenges and recommendations. *Lancet Respir. Med.* 8, 5 (2020), 506–517.
- [144] B. Pierce, K. Wiehe, H. Hwang, and B. Kim. 2014. ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 30, 12 (2014), 1771–1773.
- [145] M. Pourhomayoun and M. Shakibi. 2020. Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making. *MedRxiv*, 2020.
- [146] M. Prachar, S. Justesen, D. Steen, S. Thorgrimsen, E. Jurgons, O. Winther, and F. Bagger. 2020. COVID-19 Vaccine Candidates: Prediction and Validation of 174 SARS-CoV-2 Epitopes. *BioRxiv*, 2020.
- [147] K. Preuer, G. Klambauer, F. Rippmann, and S. Hochreiter. 2019. Interpretable deep learning in drug discovery. In *Explainable AI*. Springer, 331–345.
- [148] N. Punn, S. Sonbhadra, and S. Agarwal. 2020. COVID-19 epidemic analysis using machine learning and deep learning algorithms. *MedRxiv*, 2020.
- [149] X. Qi, Z. Jiang, Q. Yu, C. Shao, and H. Zhang. 2020. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: A multicenter study. *MedRxiv*, 2020.
- [150] X. Qian, R. Ren, Y. Wang, Y. Guo, and J. Fang. 2020. Fighting against the common enemy of COVID-19: A practice of building a community with a shared future for mankind. *Infect. Diseases Poverty* 9, 1 (2020), 1–6.
- [151] R. Qiao, N. Tran, B. Shan, A. Ghodsi, and M. Li. 2020. Personalized workflow to identify optimal T-cell epitopes for peptide-based vaccines against COVID-19. Retrieved from arXiv:2003.10650.
- [152] R. Adrian. 2020. Detecting COVID-19 in X-ray images with keras, tensorflow, and deep learning. Retrieved from <https://www.pyimagesearch.com/category/medical>.
- [153] R. Tawsifur. 2020. Novel corona virus 2019 dataset. Retrieved from <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>.
- [154] Radiopaedia. 2020. Images of COVID-19 cases. Mendeley Data. Retrieved from <https://radiopaedia.org>.
- [155] M. Rahman, M. Hoque, M. Islam, S. Akter, A. Rubayet, M. Siddique, O. Saha, M. Rahaman, M. Sultana, and M. Hossain. 2020. Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2 etiologic agent of global pandemic COVID-19: An *in silico* approach. *PeerJ* 8 (2020), e9572.
- [156] S. Rahmatizadeh, S. Valizadeh, and A. Dabbagh. 2020. The role of artificial intelligence in management of critical COVID-19 patients. *J. Cell. Mol. Anesthesia* 5, 1 (2020), 16–22.
- [157] G. Randhawa, M. Soltysiak, H. El, and C. de Souza. 2020. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PLoS One* 15, 4 (2020), e0232391.

- [158] A. Rao and J. Vazquez. 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. *Infect. Control Hosp. Epidemiol.* 41, 7 (2020), 826–830.
- [159] N. Rapin, O. Lund, M. Bernaschi, and F. Castiglione. 2010. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* 5, 4 (2010).
- [160] R. Rizk and A. Hassanien. 2020. COVID-19 forecasting based on an improved interior search algorithm and multi-layer feed forward neural network. Retrieved from <https://arXiv:2004.05960>.
- [161] R. Sameni. 2020. Mathematical modeling of epidemic diseases: A case study of the COVID-19 coronavirus. Retrieved from <https://arXiv:2003.11371>.
- [162] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR'18)*. 4510–4520.
- [163] K. Santosh. 2020. AI-driven tools for coronavirus outbreak: Need of active learning and cross-population train/test models on multitudinal/multimodal data. *J. Med. Syst.* 44, 5 (2020), 1–5.
- [164] B. Sarkar, M. Ullah, F. Johora, and M. Taniya. 2020. The essential facts of Wuhan novel coronavirus outbreak in china and epitope-based vaccine designing against 2019-nCoV. *BioRxiv*, 2020.
- [165] J. Sarkar and P. Chakrabarti. 2020. A machine learning model reveals older age and delayed hospitalization as predictors of mortality in patients with COVID-19. *MedRxiv*, 2020.
- [166] B. Schuller, D. Schuller, K. Qian, J. Liu, H. Zheng, and X. Li. 2020. COVID-19 and computer audition: An overview on what speech and sound analysis could contribute in the SARS-CoV-2 Corona crisis. Retrieved from <https://arXiv:2003.11117>.
- [167] A. Senior, R. Evans, J. Jumper, and J. Kirkpatrick. 2020. Improved protein structure prediction using potentials from deep learning. *Nature* 577, 7792 (2020), 706–710.
- [168] A. Senior, R. Evans, J. Jumper, J. Kirkpatrick, and L. Sifre. 2019. Protein structure prediction using multiple deep neural networks CASP13. *Proteins: Struct. Funct. Bioinform.* 87, 12 (2019), 1141–1148.
- [169] P. Sethy and S. Behera. 2020. Detection of coronavirus Disease (COVID-19) based on deep features. Preprints, 2020.
- [170] F. Shan, Y. Gao, J. Wang, and W. Shi. 2021. Abnormal Lung Quantification in Chest CT Images of COVID-19 Patients with Deep Learning and its Application to Severity Prediction. *Medical Physics* 48, 4 (2021), 1633–1645.
- [171] Feng Shi, Jun Wang, Jun Shi, Ziyang Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, and Dinggang Shen. 2020. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* 14 (2020), 4–15.
- [172] F. Shi, L. Xia, F. Shan, and D. Wu. 2021. Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. *arXiv*, 2021.
- [173] W. Shi, X. Peng, T. Liu, and Z. Cheng. 2020. Deep learning-based quantitative computed tomography model in predicting the severity of COVID-19: A retrospective study in 196 patients. *Ann Transl Med.* 9, 3 (2020), 216.
- [174] B. Shin, S. Park, K. Kang, and J. Ho. 2019. Self-attention based molecule representation for predicting drug-target interaction. In *Proceedings of the Machine Learning for Healthcare Conference*. PMLR, 230–248.
- [175] Gitanjali R. Shinde, Asmita B. Kalamkar, Parikshit N. Mahalle, Nilanjan Dey, Jyotismita Chaki, and Aboul Ella Hassanien. 2020. Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput. Sci.* 1, 4 (2020), 1–15.
- [176] K. Simonyan and A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. Retrieved from <https://arXiv:1409.1556>.
- [177] SIRM. 2020. COVID-19 database. Retrieved from <https://sirm.org/category/senza-categoria/COVID-19>.
- [178] M. Siwiak, P. Szczesny, and M. Siwiak. 2020. From a single host to global spread. the global mobility based modelling of the COVID-19 pandemic implies higher infection and lower detection rates than current estimates. *Soc. Sci. Res. Netw.* (2020) <https://ssrn.com/abstract=3562477>.
- [179] M. Skalic, J. Jiménez, D. Sabbadin, and G. De. 2019. Shape-based generative modeling for de novo drug design. *J. Chem. Info. Model.* 59, 3 (2019), 1205–1214.
- [180] Y. Song, S. Zheng, L. Li, X. Zhang, and X. Zhang. 2020. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 99, 1 (2020), 1–14.
- [181] M. Stepniewska, P. Zielenkiewicz, and P. Siedlecki. 2018. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. *Bioinformatics* 34, 21 (2018), 3666–3674.
- [182] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.
- [183] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, and Z. Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR'16)*. 2818–2826.

- [184] B. Tang, F. He, D. Liu, M. Fang, Z. Wu, and D. Xu. 2020. AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *BioRxiv*, 2020.
- [185] Z. Tang, W. Zhao, X. Xie, and Z. Zhong. 2021. Severity assessment of COVID-19 using CT image features and laboratory indices. *Phys. Med. Biol.* 66, 3 (2021), 035015.
- [186] Z. Tanoli, Z. Alam, and V. Markus. 2018. Drug target commons 2.0: A community platform for systematic analysis of drug-target interaction profiles. *Database* (2018), 083.
- [187] P. Teles. 2020. Predicting the evolution Of SARS-Covid-2 in Portugal using an adapted SIR model previously used in South Korea for the MERS outbreak. Retrieved from <https://arXiv:2003.10047>.
- [188] I. Thevarajan, T. Nguyen, M. Koutsakos, J. Druce, L. Caly, C. Sandt, X. Jia, S. Nicholson, M. Catton, and B. Cowie. 2020. Breadth of concomitant immune responses prior to patient recovery: A case report of non-severe COVID-19. *Nature Med.* 26, 4 (2020), 453–455.
- [189] H. Tian, Y. Liu, Y. Li, and C. Wu. 2020. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 368, 6491 (2020), 638–642.
- [190] K. To, O. Tsang, W. Leung, and A. Tam. 2020. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: An observational cohort study. *Lancet Infect. Diseases* 20, 5 (2020), 565–574.
- [191] A. Toda. 2020. Susceptible-infected-recovered (SIR) dynamics of COVID-19 and economic impact. Retrieved from <https://arXiv:2003.11221>.
- [192] F. Touret and X. Lamballerie. 2020. Of chloroquine and COVID-19. *Antiviral Research* 177 (2020), 104762.
- [193] Google Trends. 2020. Coronavirus search trends. Retrieved from https://trends.google.com/trends/story/US_cu_4Rjdh3ABAABMHM_en.
- [194] O. Trott and A. Olson. 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 2 (2010), 455–461.
- [195] R. Viner, S. Russell, H. Croker, J. Packer, J. Ward, C. Stansfield, O. Mytton, and C. Bonell. 2020. School closure and management practices during coronavirus outbreaks including COVID-19: A rapid systematic review. *Lancet Child Adolesc. Health* 4, 5 (2020), 397–404.
- [196] R. Vita, L. Zarebski, J. Greenbaum, and H. Emami. 2010. The immune epitope database 2.0. *Nucleic Acids Res.* 38, suppl_1 (2010), D854–D862.
- [197] I. Wallach, M. Dzamba, and A. Heifets. 2015. AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. Retrieved from <https://arXiv:1510.02855>.
- [198] A. Walls, X. Xiong, and Y. Park. 2019. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* 176, 5 (2019), 1026–1039.
- [199] A. C. Walls, Y. Park, M. Tortorici, and A. Wall. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 2 (2020), 281–292.
- [200] H. Wang, Y. Zhang, S. Lu, and S. Wang. 2020. Tracking and forecasting milestone moments of the epidemic in the early-outbreak: Framework and applications to the COVID-19. *F1000Res.* 9 (2020).
- [201] L. Wang and A. Wong. 2020. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. *Sci. Rep.* 10, 1 (2020), 1–12.
- [202] M. Wang, R. Cao, L. Zhang, and X. Yang. 2020. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. *Cell Res.* 30, 3 (2020), 269–271.
- [203] S. Wang, B. Kang, J. Ma, and X. Zeng. 2021. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *European Radiology* 31 (2021), 6096–6104.
- [204] X. Wang, Y. Peng, L. Lu, and R. Summers. 2017. Hospital-scale Chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR'17)*.
- [205] Y. Wang, M. Hu, Q. Li, and X. Zhang. 2020. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. Retrieved from <https://arXiv:2002.05534>.
- [206] D. Ward, M. Higgins, J. Phelan, M. Hibberd, S. Campino, and T. Clark. 2021. An integrated *in silico* immuno-genetic analytical platform provides insights into COVID-19 serological and vaccine targets. *Genome Med.* 13, 1 (2021), 1–12.
- [207] WHO. 2021. Novel coronavirus 2019 (COVID-19). Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- [208] R. Wölfel, V. Corman, W. Guggemos, and M. Seilmaier. 2020. Virological assessment of hospitalized patients with COVID-2019. *Nature* 581, 7809 (2020), 465–469.
- [209] Worldpop. 2020. The statistics datasets on holidays and air travel. Retrieved from <https://www.worldpop.org>.
- [210] F. Wu, S. Zhao, B. Yu, Y. Chen, and W. Wang. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579, 7798 (2020), 265–269.

- [211] J. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, and P. Salazar. 2020. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Med.* 26, 4 (2020), 506–510.
- [212] J. Wu, K. Leung, and G. Leung. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* 395, 10225 (2020), 689–697.
- [213] J. Wu, P. Zhang, L. Zhang, W. Meng, and J. Li. 2020. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. MedRxiv, 2020.
- [214] B. Xu, B. Gutierrez, S. Mekaru, and K. Sewalk. 2020. Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci. Data* 7, 1 (2020), 1–6.
- [215] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, and S. Lv. 2020. A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* 6, 10 (2020), 1122–1129.
- [216] Y. Xu, X. Li, B. Zhu, H. Liang, and C. Fang. 2020. Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nature Med.* 26, 4 (2020), 502–505.
- [217] L. Yan, H. Zhang, J. Goncalves, and Y. Xiao. 2020. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. MedRxiv, 2020.
- [218] L. Yan, H. Zhang, Y. Xiao, and M. Wang. 2020. Prediction of criticality in patients with severe COVID-19 infection using three clinical features: A machine learning-based prognostic model with clinical data in Wuhan. MedRxiv, 2020.
- [219] H. Yang, W. Xie, X. Xue, and K. Yang. 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* 3, 10 (2005).
- [220] W. Yang, Q. Cao, L. Qin, X. Wang, and Z. Cheng. 2020. Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): A multi-center study in Wenzhou city, Zhejiang, China. *J. Infect.* 80, 4 (2020), 388–393.
- [221] X. Yang. 2012. Flower pollination algorithm for global optimization. In *Proceedings of the International Conference on Unconventional Computing and Natural Computation*. Springer, 240–249.
- [222] Z. Yang, Z. Zeng, and K. Wang. 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 under public health interventions. *J. Thoracic Dis.* 12, 3 (2020), 165.
- [223] B. Zareie, A. Roshani, M. Mansournia, and M. Rasouli. 2020. A model for COVID-19 prediction in Iran based on China parameters. *Arch. Iran. Med.* 23, 4 (2020), 244–248.
- [224] C. Zavaleta. 2020. COVID-19: Review Indigenous peoples' data. *Nature* 580, 7802 (2020), 185.
- [225] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia. 2020. COVID-19 screening on chest X-ray images using deep learning based anomaly detection. Retrieved from <https://arXiv:2003.12338>.
- [226] J. Zhang, H. Zeng, J. Gu, H. Li, L. Zheng, and Q. Zou. 2020. Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines* 8, 2 (2020), 153.
- [227] J. Zhao, Y. Zhang, X. He, and P. Xie. 2020. Covid-CT-dataset: A CT scan dataset about COVID-19. Retrieved from <https://github.com/UCSD-AI4H/COVID-CT>.
- [228] A. Zhavoronkov, V. Aladinskiy, and A. Zhebrak. 2020. Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *Insilico Med.* 307 (2020), E1.
- [229] C. Zheng, X. Deng, Q. Fu, Q. Zhou, and J. Feng. 2020. Deep learning-based detection for COVID-19 from chest CT using weak label. MedRxiv, 2020.
- [230] P. Zhou, X. Yang, X. Wang, and B. Hu. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 7798 (2020), 270–273.
- [231] Y. Zhou, Y. Hou, J. Shen, and Y. Huang. 2020. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 6, 1 (2020), 1–18.
- [232] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang. 2018. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 3–11.
- [233] ZINC. 2020. ZINC database. Retrieved from <https://zinc.docking.org>.

Received May 2020; revised March 2021; accepted May 2021