Universidad Politécnica
de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Master's degree in Data Science

Master´s Final Project

Machine Learning applied to COVID-19

Author: Serena Alderisi
Tutor: Pedro Larrañaga, María Concepción Bielza

Madrid, October 2020

This Master's Final Project has been deposited in the ETSI Informáticos of the Polytechnic University of Madrid for its defense.

Master's final project
Master's degree in Data Science
Title: Machine Learning applied to COVID-19
October 2020

Author: Serena Alderisi
Tutor: Pedro Larrañaga, María Concepción Bielza
Artificial Intelligence Department
ETSI Informáticos
Universidad Politécnica de Madrid

Disclaimer

This work is carried out from a data science point of view.
No expert medical support has been consulted but scientific papers from medical literature published on PubMed, Google Scholar, ScienceDirect, SprigerLink, Elsevier and other trusted medical journal search engines.
The review of the papers presented in this work has academical purpose since it constitutes the Master's Final Project in Data Science MSc.

Abstract

This work is focused on the impact of machine learning, on the COVID-19 pandemic. Machine learning has proven to be invaluable in predicting risks in many spheres and since the spread of the virus started, its application is helping us against the viral pandemic. Like never before, people all around the world are collecting and sharing what they learn about the virus. Hundreds of research teams are combining their efforts to collect data and develop solutions every day.

Starting from this, the main goals of this work are: to shine a light on their work; going deep into how the application of machine learning techniques on different fields affected by the pandemic is helping us in the fight against the coronavirus; to identify strengths and weaknesses of machine learning techniques and the challenges for further progress in medical machine learning systems.
This final master thesis report addresses recent studies that apply machine learning on multiple angles: screening and diagnosis, contact tracing, drugs/vaccines development and prediction and forecasting for COVID-19.

Early and timely treatment computer-aided diagnosis, highlighted the importance of machine learning algorithms that proved to be urgently needed. Its application could largely reduce the efforts of clinicians and accelerate the screening and diagnosis process, as well as, reduce the human intervention in medical practice. Analyzing X-rays images with deep learning algorithms makes possible to achieve amazing results: in some cases models performance surpass the ability of a senior radiologists to early distinguish different types of pneumonia from COVID-19 pneumonia. In addition, deep neural networks are not only a valuable support to the activity of radiologists but they also represent a double check for false RT-PCR responses.

Using mobile apps for contact tracing helps to make people more aware about the consequences of their contacts and their movements. This consciousness will allow to contain the spread of the virus and individual protection. Having situational awareness by governments would benefit from increased access to previously unavailable population estimates and mobility information to enable different sectors to better understand COVID-19 trends and geographic distribution. This could be a big improvement, since the lack of available data is a big limitation for the implementation of machine learning solutions in the real world.

Machine learning application on laboratory research is allowing to speed up a vaccine development against SARS-CoV-2 virus and to discover which among the already existing drugs have potential anti-coronavirus activities.

Last but not least, knowing earlier who, among the cases, will develop severe symptoms, allows to manage better medical resources and treatments and mostly to avoid intensive care unit stress. Furthermore, knowing the possible development of the contagion dynamics lets governments plan more efficient and less-invasive containment strategies.

The outcome of this work can provide deep insights into this disease, re-evaluate existing medical diagnosis systems for this virus, and recommend solutions for developing reliable, faster and more efficient medical systems, with the aid of machine learning application.

Contents

Chapter 1

Introduction

On January 30, 2020, the International Health Regulations Emergency Committee of the World Health Organisation (WHO) declared the outbreak of the resulting disease from this new CoV called COVID-19, as a "public health emergency of international concern". Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has spread to every inhabited continent becoming an unprecedented public health crisis. In October 2020, *Coronavirus Resource Center at Johns Hopkins University of Medicine* [1] reported a total of more than 1 million deaths as worldwide COVID-19 infections surpassed 40 millions. Such quick spread is due to the fact that this virus is transmitted from person to person very easily through coughing, sneezing and respiratory droplets (*Coronavirus, WHO website, 2020*) [2]. It usually presents with symptoms of fever, cough, shortness of breath, and can have serious consequences such as pneumonia, multi-organ failure and death.

Today, a clear solution has not been developed on COVID-19. The vast majority of measures taken on a country basis and individually are to prevent the transmission of this virus to more people. Because of the uncertainty in the transmission dynamics of SARS-CoV-2 and high certainty in its virulence, it is understandable that early responses have relied on blunt interventions, such as movement bans and closures, to save lives.

Given the increasing caseload, there is an urgent need to augment medical and economical skills to face this critical illness. Hence, the scientific challenge now is to identify, through inference and simulation, measures that could provide as-good or better protection with less social cost (*Cobey, 2020*). The growing emphasis on machine learning techniques in medical fields can provide the right environment for change and improvement.

To address this global novel pandemic, WHO, scientists and clinicians in medical industries are searching for new technology to screen infected patients in various stages, find best clinical trials, control the spread of this virus, develop a vaccine for curing infected patients, and trace contacts. The role of the data science in this scenario consists in helping to speed up the process.
On March 16, 2020, the White House, collaborating with research institutes and tech companies, issued a call to action for global artificial intelligence researchers for developing novel text and data-mining techniques to assist COVID-19-related research.

The Allen Institute for AI in partnership with leading research groups issued an open source, weekly updated *COVID-19 Open Research Dataset* [3], which continuously documents COVID-19-related articles to accelerate novel research projects urgently requiring real-time data. Hundreds of research teams are combining their efforts to collect data and develop solutions every day.

Machine learning has proven to be invaluable in predicting risks in many spheres and since the spread of the virus started, its application is helping us fight the viral pandemic. Like never before, people all around the world are collecting and sharing what they learn about the virus.
Starting from this, the main goal of this work is to shine a light on their work, highlighting the importance of the role of machine learning to tackle SARS-CoV-2 (Figure 1.1).

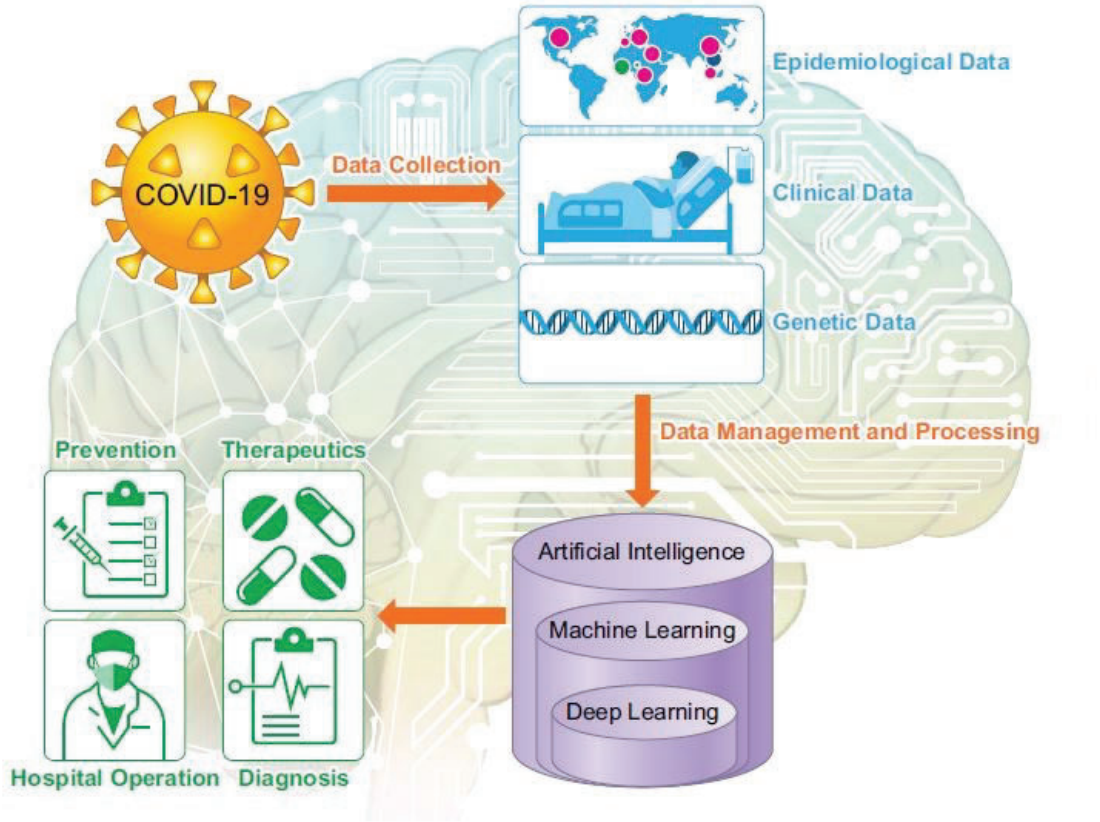

Figure 1.1: Role of AI in COVID-19 fight. *Source:* [8]

Worthy of mention is it also the effort of the European scenario, in particular of *ELLIS Society (European Lab for Learning and Intelligent Systems)* [4] that aims to boost economic growth in Europe by leveraging AI technologies applied in different areas. It involves the very best European academics while working together closely with basic researchers from industry. Some of the research works discussed in this document are part of collected COVID-19 related projects from the ELLIS network.
A further source for the papers involved in this work is the *Overview by Ad hoc Committee on Artificial Intelligence (CAHAI) - Council of Europe* [5].

In this context, the aims of this dissertation are:

- to explore the general panorama to outline recent breakthroughs in machine learning technologies and their applications on the fight against COVID-19;
- to identify achievements and the challenges for further progress in medical machine learning systems to tackle the virus;
- to go deep into how machine learning applications on different fields, affected the pandemic.

This work is a state of the art that provides an overview of some of the most relevant digital solutions achieved worldwide for screening and diagnosis, contact tracing, drugs/vaccines development and prediction and forecasting for COVID-19 disease, through the application of machine learning. Furthermore, for each scientific contribution distinct informations are provided, such as: application nature, machine learning techniques used and performance obtained from a data science point of view. It will be explained how the rapid development of automated diagnostic systems based on machine learning not only can contribute to increase diagnostic accuracy, save resources and to speed up progress, but they also represent an effective solution to protect healthcare workers by decreasing their contacts with COVID-19 patients.

To allow an easier reading, some main medical concepts related to these most recent research works have been briefly explained. Although no expert knowledge has revised this work, it can be considered as main resource the medical literature published on PubMed, Google Scholar, ScienceDirect, SpringerLink, Elsevier and other trusted medical journal search engines.

Chapter 2

State of the Art

Recent studies identified that machine learning (ML) is a promising technology employed by various healthcare providers as it results in better scalable, speed-up processing power, reliable and even outperforms humans in specific healthcare tasks (*Davenport et al., 2020*).
According to the American Physiological Society, the large-scale data of COVID-19 patients can be integrated and analyzed by advanced ML algorithms to better understand the pattern of viral spread, further improve diagnostic speed and accuracy, develop novel effective therapeutic approaches, and potentially identify the most susceptible people based on personalized genetic and physiological characteristics (*Alimadadi et al., 2020*).
With the recent progress in digitised data acquisition and ML and computing infrastructure, AI applications were expanding into areas that were previously believed to be only the area of human experts (*Yu et al., 2020*).

The introduction of ML techniques in radiology for instance, represents a qualitative and quantitative leap for clinical imaging interpretation and processing large volumes of data for imaging research. There is a certain reluctance from radiologists to use these new applications, even if ML and deep learning approaches have widely established their ability for image and lesion classification (*Noguerol et al., 2020*).
Indeed, a growing evidence of the advantages of ML in radiology lies in the ability of creating seamless imaging workflows for radiologists or even replacing radiologists. Deep learning algorithms provide performance as good as an expert radiologist (Section 2.2).
In the next pages it follows a comprehensive review of some of the most recent studies to tackle the novel COVID-19 pandemic, through the application of ML techniques on different areas: screening and diagnosis, contact tracing, drugs/vaccines development and prediction and forecasting for COVID-19.

2.1 ML technology in Coronavirus Family

From a specialised medical perspective, the features and classes of the CoV family are similar to one another. For that reason looking at the past researches could help to compensate the lack of knowledge about the new coronavirus.
Coronavirus is the common name for Coronaviridae, a large family of viruses.
In humans, the Coronaviruses cause respiratory infections, including the common cold till to rarer forms such as SARS (including the one causing COVID-19) and MERS.

- SARS-CoV: It was identified in 2003 as responsible for the epidemy of Severe Acute Respiratory Syndrome begun in China at the end of 2002.
- MERS-CoV: It was identified in 2012 as the cause of Middle East Respiratory Syndrome.
- SARS-CoV2: It's a novel coronavirus identified for the first time in Wuhan (China), at the end of 2019 as cause of COVID-19 Syndrome and it has been spread worldwide.

Albhari et al., 2020 is a systematic review based on the exploration of the CoV family by reviewing articles on data mining and ML algorithms, to understand how previous research has addressed prediction, regression, and classification methods before the spread of COVID-19.
The results of that study show in order, that the most used ML algorithms involved in the analysis of the previous coronaviruses, MERS-CoV and SARS-CoV, are decision tree algorithms, that were the most frequently used, naive bayes, support vector machine and k-nearest neighbours (from a research on 1305 articles from ScienceDirect, IEEE Xplore, Web of Science, PubMed and Scopus).
It is emerged that in MERS-CoV and SARS-CoV, age attribute is considered the most important and dangerous factor in the infected patients, since people over the age of 50 are more likely to be at risk and to have this type of virus than others. Shortness Of Breath (SOB) is also considered the most important concern with MERS-CoV and SARS-CoV illnesses.

In this context, the new epidemic of COVID-19 depends on the same features of MERS-CoV and SARS-CoV.
Several reliable reports and government news have mentioned that age is one of the most important features of patients with COVID-19 (some studies related to this topic are discussed in Section 2.5). Patients over the age of 50 are susceptible to contract the disease and be exposed to its risks and complications, also because generally higher age is correlated with a higher number of pathologies. The medical team has reported that SOB is the most important symptom attribute because it carries a high specificity for COVID-19, in patients that show symptoms.
COVID-19 seems not to be very different from SARS regarding its clinical features. However, it has a fatality rate of 2.3%, lower than that of SARS (9.5%) and much lower than that of MERS (34.4%).
A gastrointestinal route of transmission for SARS-CoV-2, which has been assumed for SARS-CoV and MERS-CoV for example, cannot be ruled out and needs further investigation (*Petrosillo et al., 2020*).

The lessons learned in the past from the SARS and MERS epidemics are among the best cultural weapons we have to face this new global threat.
Results obtained in previous analyses of MERS and SARS can serve as a guide for future research in the context of data mining algorithms and as consequence, detection and diagnosis can be remarkably enhanced. As we will see later, practical examples of what aforementioned are discussed in Section 2.4. Moreover, close cooperation among researchers in the biomedical engineering field and the medical community is necessary to stop the growing public health threat posed by the 2019 CoV.

In previous studies, such as *Sandhu al., 2016*, GPS was used to represent each MERS-CoV user on Google maps so that possibly infected users can be quarantined as early as possible. In this paper an effective cloud computing system is proposed which predicts MERS-CoV-infected patients using Bayesian networks and provides geographic-based risk assessment to control its outbreak.
MERS-CoV application user data were stored over shared cloud storage provided by Amazon. It could be accessed by doctors, users, healthcare departments, and the governmental agencies. Different classification algorithms such as k-nearest neighbors, linear regression, and neural networks were also implemented for one of the vital tasks of the proposed system: the classification of users into infected and possibly infected, so it required a high performance level.
As will be shown in Section 2.3, this general idea represents the starting point of the newest contact tracing app for COVID-19.

2.2 ML technology in COVID-19 Screening and Diagnosis

X-ray or computed tomography (CT) images are useful diagnostic tools for radiology doctors to detect COVID-19. Nevertheless, CT alone may have limited negative predictive value for ruling out SARS-CoV-2 infection, as some patients may have normal radiological findings at early stages of the disease.
X-rays and CT images are important complements also to reverse-transcription polymerase chain reaction (RT-PCR) tests due to the fact that they can be obtained and interpreted much more quickly than RT-PCR (*Ai et al., 2020*).
According to WHO, the most accurate diagnosis of COVID-19 infection is nucleic acid detection in secretional fluid collected from a throat swab using RT–PCR.
Even if it typically has high sensitivity and specificity, serial testing may be required to rule out the possibility of false negative results and coinfection with other viruses might influence the RT-PCR test accuracy. RT-PCR test takes up to 3 days to be completed and since in the past months there was a shortage of RT–PCR test kits, it emerged the urgent need for alternative methods that provide rapid and accurate diagnosis of patients with COVID-19.
In addition, the swab operation is categorically open to malfunctions by the expert mistakes and it should be also repeated (*Liu et al., 2020*).

In this regard, the study *Li et al., 2020* reported two false-negative results of RT-PCR for SARS-CoV-2 infection. The authors discussed the supplementary role of clinical data with RT-PCR, including laboratory examination results and CT features.
The first case is relative to a background chest or chest CT scans for 10-month-old patient. During the medical observation, two nucleic acid test presented weakly positive for influenza A and CT showed diffuse ground-glass opacities in both lungs. A deep learning-based computer-aided diagnostic system for pneumonia, which was trained with CT scans of patients with COVID-19, suggested this patient to have pneumonia based on relatively large proportion of abnormalities in lung, with the lesion volume accounting for 13.3% of the whole lungs. After two consecutive negative results, a third SARS-CoV-2 RT-PCR test confirmed the infection.
The second case regards a 36-year-old man presented with fever for 5 days (peak body temperature: 40°C). Respiratory symptoms at admission included dry throat and difficulty breathing; no cough, sputum, or stuffy/runny nose was observed. Chest CT showed emphysema in both upper lungs and diffuse ground-glass opacities in the right lower lobe, highly suggestive of viral pneumonia. In addition, the deep learning-based computer-aided diagnostic system also indicated a high risk of pneumonia with the infected area accounting for 8.9% of the whole lungs. Subsequently, throat swab specimens were promptly collected for SARS-CoV-2 RT-PCR. A negative result for SARS-CoV-2 was observed in the first RT-PCR test. A second consecutive SARS-CoV-2 RT-PCR test was conducted immediately thereafter, and a positive result was obtained. The patient was further confirmed with COVID-19 with additional positive RT-PCR tests. Also in this case ML alarmed suspected pneumonia based on relatively large proportion of abnormalities in lung.

With an integrated approach of deep learning, CT features, and RT-PCR results, the screening and treatment of COVID-19 would be more effective.
That is why CT and chest X-rays are largely used and represent a key role in the diagnosis of coronavirus disease 2019 together with RT-PCR test and or clinical in-

formation. Another motivation that leads to the combination of these diagnostic tools is that symptoms of COVID-19 disease are similar to those of pneumonia, and a certain percentage of deaths due to the COVID-19 virus are on Pneumonia disease (*Chung et al., 2020*).

This section aims to show the potential of ML tools by suggesting new models that comes up with rapid and valid methods of SARS-CoV-2 diagnosis.
Among all the ML algorithms used in the following studies, convolutional neural networks play a relevant role in early and accurate COVID-19 screening.
Convolutional neural networks (CNN) is a deep neural network-based learning architecture which can take images as input for processing a massive amount of data. Deep convolutional neural networks (DCNNs) are one of the powerful deep learning architectures and have been widely applied in many practical applications.
It gives extremely good performance in computer vision and image analysis sector such as image recognition, object detection, semantic segmentation and nowadays, it is widely applied also for medical imaging analysis.
In previous studies, DCNNs have been exploited in X-ray image classification to successfully diagnose common chest diseases such as tuberculosis screening (*Pasa et al., 2019*) and mediastinal lymph nodes in CT images (*Miki et al., 2017*).
Like a regular artificial neural network (ANN), a CNN consists of sequence of hidden layers and they are basically denoted as convolutional and polling layer.
To learn more about CNNs structure and its application on magnetic resonance imaging (MRI), see *Lundervold et al., 2019*.

In this regard, several studies that cover deep learning-based solutions to detect the COVID-19 using chest X-rays will be reviewed.
Studies that diagnosed COVID-19 using chest X-rays have binary or multiple classifications, such as COVID19 infected and healthy (binary classification) or healthy, viral pneumonia affected and bacterical pneumonia infection (multi-class classification). Some studies use raw data from open dataset, Kaggle competitions or private ones from Hospitals around the world, while others have different feature extraction processes. The number of data used in studies also varies. Some studies combine different datasets to train the models, while others used small train sets.
Anyway, in all the following studies related to this section, the most preferred method is CNN.

In *Ahammed et al., 2020*, the authors investigated chest X-ray images of normal, viral pneumonia and COVID-19 affected patients. Several ML and deep learning classifiers (n=17) were used such as support vector machines (SVM), random forests, k-nearest neighbors (kNN), logistic regressions, Gaussian naïve Bayes, Bernoulli naïve Bayes, decision trees, extreme gradient boosting Xgboost (XGB), multilayer perceptrons (MLP), nearest centroids and perceptrons. Moreover, several deep learning classifiers were also employed such as CNN, deep neural network and several pre-trained CNNs such as residual neural network (ResNet50), visual geometry group network 16 (VGG16), and inception network V3 (inceptionV3) for transfer learning. Among all of these classifiers, the proposed CNN (Figure 2.1) shows the highest accuracy 94.03%, the receiver-operator characteristic (ROC) curve, Area Under the ROC Curve (AUC) of 95.53%, F-measure 94.03%, sensitivity 94.03%, specificity 97.01%. Besides, XGB shows the second maximum accuracy and minimum error rate along with other met-

rics. Then, logistic regression, SVM, MLP, random forest and Gaussian naive Bayes demonstrate better performances than other algorithms. As a result, this model might help to early detect COVID-19 patients and prevent community transmission compared to traditional methods.

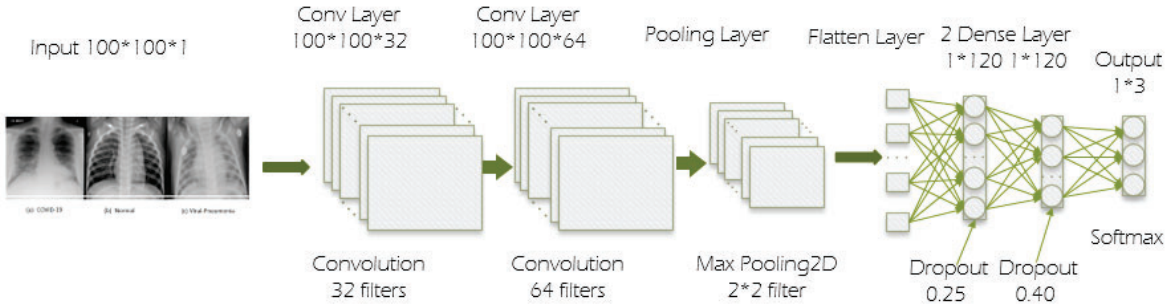

Figure 2.1: *Ahammed et al., 2020* CNN structure.

Another study that involves the application of deep neural network techniques coupled with radiological imaging for the identification of COVID-19 disease is *Shibly et al., 2020*. The aim of this work was to be supportive in overcoming the issue of a shortage of trained physicians in remote communities. In this article, the authors introduced a VGG-16 (also called OxfordNet) network-based faster regions with CNN (Faster R–CNN) framework to detect COVID-19 patients from chest X-Ray images using an available open-source dataset.
Figure 2.2 shows how the proposed architecture predicted the different samples of chest X-ray images and the confusion matrix. The model, however, made incorrect predictions mostly in poor images, and often it predicted the patient with Pneumonia as COVID-19 because they have similarities in image features. Figure 2.2 (G) depicts the generated confusion matrix based on 10-fold cross-validation method.
This approach provides a classification accuracy of 97.36%, 97.65% of sensitivity, and a precision of 99.28%. Hence, the proposed method might be of assistance for health professionals to validate their initial assessment towards COVID-19 patients.

Detecting coronavirus infected patient using X-ray images was the goal of *Sethy et al., 2020*. The idea was detecting Corona Virus by an SVM model based on deep features using X-ray images. Deep feature extraction is based on the extraction of features acquired from a pre-trained CNN (in this study the authors used 11 CNNs: AlexNet, DenseNet201, GoogleNet, Inceptionv3 ResNet18, ResNet50, ResNet101, VGG16, VGG19, XceptionNet and Inceptionresnetv2). The deep features obtained from each CNN network are used by SVM classifier. After that, the classification is performed, and the performance of all the classification models are measured. The best classification model is ResNet50 plus SVM that achieved 95.38% of accuracy, 97.29% of sensitivity and 93.47% of specificity.

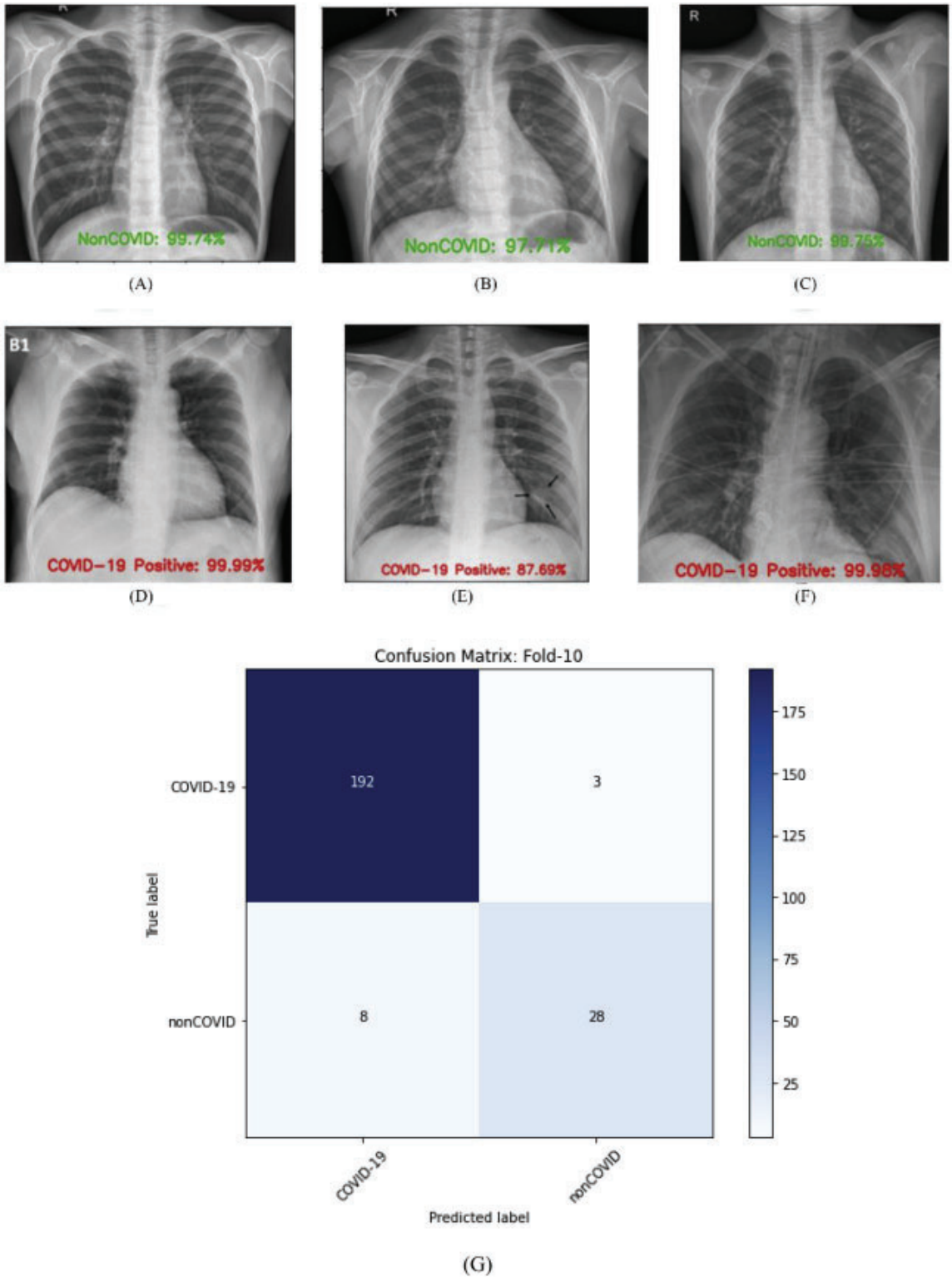

Figure 2.2: Predicted clinical outcomes CNN's *Shibly et al., 2020*

In *Narin et al., 2020* the authors proposed a deep transfer learning based approach using five pre-trained CNN based models (ResNet50, ResNet101, ResNet152, InceptionV3 and Inception-ResNetV2) for the detection of coronavirus pneumonia infected patients using chest X-ray radiographs. They implemented three different bi-

nary classifications with four classes (COVID-19, normal (healthy), viral pneumonia and bacterial pneumonia). Considering the performance results obtained, it has seen that the pre-trained ResNet50 model provides the highest classification performance: 96.1% accuracy for Dataset-1 (COVID-19 and normal), 99.5% accuracy for Dataset-2 (COVID-19 and viral Pneumonia) and 99.7% accuracy for Dataset-3 (COVID-19 and bacterial pneumonia) among other four used models.

Using *transfer learning*, the detection of various abnormalities in small medical image datasets is an achievable target, often yielding remarkable results. The results suggest that deep learning with X-ray imaging may extract significant biomarkers since CNN can learn to extract significant characteristics (features) of the image. The first strategy is called feature extraction via transfer learning and refers to the approach wherein the pre-trained model retains both its initial architecture, and all the learned weights. Hence, the pre-trained model is used only as a feature extractor, then the extracted features are inserted into a new network that performs the classification task (*Tan et al., 2018*).

Transfer learning procedure is adopted also in *Ioannis et al., 2020*. In this study, the performance of 5 nets were evaluated: VGG19, MobileNet v2, Inception, Xception, Inception ResNet v2. The classification was carried out for 2 datasets: Dataset_1 that contains instances of bacterial pneumonia, confirmed COVID- 19 disease and healty patients and Dataset_2 that contains COVID-19 cases, healthy instances and both bacterial and viral pneumonia. The authors performed both binary (COVID-19 and no COVID-19) and 3-class (normal, pneumonia and COVID-19) classification tasks. The best performance obtained are: for Dataset_1, 0.99 of accuracy for binary classification and 0.93 of accuracy for 3-class classification, 0.93 of sensitivity and 0.99 of specificity using VGG19 net; for Dataset_2, 0.97 of accuracy for binary classification and 0.95 of accuracy for 3-class classification, 0.99 of sensitivity and 0.96 of specificity using MobileNet v2.

In *Ying et al., 2020* a deep learning-based CT diagnosis system (DeepPneumonia) was developed to identify patients with COVID-19, bacteria pneumonia infected patients and healthy persons. Again, 2 classification tasks are considered: the first to discriminate the bacterial pneumonia and viral pneumonia (COVID-19) and the second to discriminate COVID-19 patients and 86 healthy persons. The first analysis achieved an AUC of 0.95, a sensitivity of 0.96 and an accuracy of 0.86, while in the second 0.99 of AUC, sensitivity of 0.93 and accuracy of 0.94. An additional skill of this model is the capability of localize the main lesion features, especially the ground-glass opacity (GGO) that is of great help to assist doctors in diagnosis. Moreover, the diagnosis for a patient could be finished in 30 seconds, and the implementation on Tianhe-2 supercomputer enables a parallel execution of thousands of tasks simultaneously.
On top of that, the authors developed an *online server for academic use only* [28] to predict COVID-19 uploading CT images file. This fully automated lung CT diagnosis system was developed by three main steps: First, the extraction of the main regions of lungs and filling the blank of lung segmentation with the lung itself to avoid noises caused by different lung contours. Then, the design of a details relation extraction neural network (DRE-Net constructed on the pretrained ResNet-5013 plus feature pyramid network (FPN)) to extract the top-details in the CT images and obtain

the image-level predictions. Finally, the image-level predictions were aggregated to achieve patient-level diagnoses.

In *Zheng et al., 2020*, a weakly-supervised deep learning-based software system was developed using 3D CT volumes to detect COVID-19. For each patient, the lung region was segmented using a pre-trained UNet; then the segmented 3D lung region was fed into a 3D deep neural network to predict the probability of COVID-19 infectious. The proposed 3D architecture is called DeCoVNet. It took a CT volume with its 3D lung mask as input and directly output the probabilities of COVID-19-positive and COVID-19-negative. To avoid the overfitting problem since the number of training CT volumes was limited, online data augmentation strategies were applied.
The deep learning algorithm obtained 0.959 AUC with 0.907 sensitivity and 0.911 specificity in the ROC curve. When using a probability threshold of 0.5 to classify COVID-19-positive and COVID-19-negative, the algorithm obtained an accuracy of 0.901, a positive predictive value of 0.840 and a very high negative predictive value of 0.982. Moreover, this algorithm took only 1.93 seconds to process a single patient's CT volume using a dedicated GPU.

Xu et al., 2020 aimed to establish an early screening model to distinguish COVID-19 pneumonia from Influenza-A viral pneumonia (IAVP) and healthy cases through pulmonary CT images using deep learning techniques. First, the CT images were preprocessed to extract the effective pulmonary regions. Second, a three-dimensional (3D) CNN segmentation model was used to "segment" multiple candidate image cubes. The center image together with its two neighbors of each cube was collected for further steps. Third, an image classification model was used to categorize all the image patches into three types: COVID-19, IAVP, and irrelevant to infection (ITI). Image patches from the same cube "voted" for the type and confidence score of this candidate as a whole. Finally, the overall analysis report for one CT sample was calculated using the noisy-or Bayesian function. The model achieved an overall accuracy rate of 86.7%, sensitivity of 86.7% and precision of 81.03%.

Abraham et al., 2020 investigated the effectiveness of a multi-CNN, a combination of several pre-trained CNNs, for the automated detection of COVID-19 from X-ray images. The method uses a combination of features extracted from multi-CNN with correlation based feature selection (CFS) technique and Bayesnet classifier for the prediction of COVID-19. It has to be underline that no existing state-of-the-art methods have employed multi-CNNs, CFS-based feature selection and Bayesnet classifier for the diagnosis of COVID-19. Moreover, this method used a relative large number of COVID19 cases, whereas the existing works have used a small dataset. The multi-CNN method extracted the features from the Xray images given in input. Then, the CFS algorithm in combination with subset size forward selection (SSFS) was utilized to determine the optimal feature subset and to reduce dimensionality.
Finally, the authors used a Bayesian network as a classifier. The Bayesian network was built with hill climbing search algorithm based on the K2 score and the algorithm for determining the conditional probability tables of the Bayes network was set to simple estimator algorithm in WEKA. Before selecting Bayesnet as best classifier, the authors performed other attempts using naive Bayes, SVM, logistic regression, AdaBoost and random forest. The best performing multi-CNN used in this study employs a combination of 5 pre-trained CNNs: Squeezenet, Darknet-53, MobilenetV2,

Xception and Shufflenet. The analysis was carried out using 2 datasets. The performance obtained is: 0.91 of accuracy, 0.98 sensitivity, 0.96 AUC for the first dataset, and 0.97 of accuracy, 0.99 sensitivity, 0.91 AUC for the second dataset.

Mei et al., 2020 proposed an ML model which combines CT imaging and clinical information showing equivalent accuracy of a senior chest radiologist to rapidly diagnose patients who are positive for COVID-19. The authors developed a DCNN (18-layer residual network: ResNet-18) to learn the imaging characteristics of patients with COVID-19 on the initial CT scan. Then, they used SVM, random forest and MLP classifiers to classify patients with COVID-19 only according to clinical information. MLP showed the best performance on the tuning set and finally to combine radiological data and clinical information to predict COVID-19 status a neural network model is used. Hence, these three ML models are used to generate the probability of a patient being COVID-19 (+): the first is based on a chest CT scan, the second on clinical information and the third on a combination of the chest CT scan and clinical information. Patient's age, presence of exposure to SARS-CoV-2, presence of fever, cough and cough with sputum, white blood cell counts, neutrophil counts, percentage neutrophils, lymphocyte counts and percentage lymphocytes were significant clinical features associated with SARS-CoV-2 status. The results show that this model, combining CT images with clinical history, achieved an AUC value of 0.92, and 0.84 of sensitivity and 0.83 of specificity.
A further step was to compare the obtained performance with the diagnosis made by two radiologists with 10 years of experience, providing them both CT imaging and clinical information. Diagnosis made by two senior radiologists achieved 0.84 of AUC, 0.74 of sensitivity and 0.94 in specificity.

In *Hurt et al., 2020*, the authors proposed a study to augment interpretation of chest radiographs with deep learning probability maps. They used a U-net CNN to predict pixel-wise probability maps for pneumonia on a dataset that contains publicly available frontal chest radiographs with bounding boxes representing pneumonia annotated by radiologists. So, the U-net17 was trained using the synthetic probability maps to predict the pixel-wise likelihood of pneumonia on each frontal chest radiograph. Predictions are represented by a pneumonia probability map. By establishing a probability threshold, these maps can be collapsed to a binary classification for the absence/presence of pneumonia. For performance evaluation of U-Net segmentation for detection of pneumonia, the authors considered the ROC curve. Including all radiographs, overall performance yielded an AUC of 0.854, which corresponds to an accuracy of 81.6%, sensitivity of 82.8%, specificity of 72.6%, positive predictive value of 47.9%, and negative predictive value of 93.3%. Excluding radiographs with other diagnoses (not pneumonia), AUC was 0.944. Excluding normal radiographs, AUC was 0.788. The pneumonia probability map produced by this approach may interface more naturally with radiologist interpretation than purely classification-based strategies.
Later, the authors in *Hurt et al., 2020*, generalized their algorithm on frontal chest X-ray images only. Their model predicted and correctly localized areas of pneumonia; it also assigns likelihoods that mirror the severity of the imaging findings, illustrating a surprising degree of generalizability and robustness.

An increasing number of recent studies is demonstrating the ability of ML models to separate COVID-19 from other pneumonias.
Another example is the study of *Li et al., 2020* that provides a fully automatic framework to detect COVID-19 and to differentiate it from community-acquired pneumonia and other lung conditions, using chest CT.
In this retrospective and multicenter study, a deep learning model, the COVID-19 detection neural network (COVNet), was developed to extract visual features from volumetric chest CT scans for the detection of COVID-19. CT scans of community-acquired pneumonia (CAP) and other non-pneumonia abnormalities were included to test the robustness of the model.
COVNet is a 3D deep learning framework consisting of RestNet50 as the backbone and is able to generate a probability score for each type of prediction: COVID-19, CAP, and non-pneumonia. The sensitivity and specificity for COVID-19 are 90% and 96% respectively. For the detection of CAP, COVNet model yielded a sensitivity of 87% and a specificity of 92%. Corresponding areas under the receiver operating characteristic curves for COVID-19 and CAP are 0.96 and 0.95, respectively.

As seen so far, it have been achieved impressive results with the application of computer vision technology, so much so that, nowadays there are ML models able to reach perfect classification rates for Corona detection.
That's the case of *Tuncer et al., 2020*.
Here, the authors proposed a method that achieved 100.0% classification accuracy for COVID-19 detection by using lung X-ray images. This method consists of preprocessing, feature extraction, and feature selection stages. The proposed feature generation method is called residual exemplar local binary pattern (ResExLBP). In the feature selection phase, a novel iterative reliefF (IRF) based feature selection is used. Decision tree, linear discriminant, SVM, kNN and subspace discriminant (SD) methods are chosen as classifiers in the classification phase. The classifier that reached best performance discriminating healthy and COVID-19 affected persons was SVM classifier, which achieved 99.55% of accuracy, 98.29% of sensitivity and 100% specificity, using 10-fold cross-validation.

In *Ozturk et al., 2020*, the authors designed an auxiliary tool to increase the accuracy of COVID-19 diagnosis. DarkCovidNet is a new model for automatic COVID-19 detection based on deep learning algorithms. The proposed model is developed to provide accurate diagnostics for binary classification (COVID-19 vs. no-findings) and multi-class classification (COVID-19 vs. No-findings vs. pneumonia).
DarkCovidNet model produced for binary classes a classification accuracy of 98.08%, sensitivity of 95.13%, specificity of 98.03% and for multi-class cases an accuracy of 87.02%, sensitivity of 85.35%, 92.18% specificity. The DarkNet, with 17 convolutional layers, model was used in this study as a classifier for the you only look once (YOLO) real time object detection system.

In *Wang et al., 2020* the authors built a transfer learning neural network, M-inception (based on the Inception network) from retrospectively collected CT images (including typical viral pneumonia and confirmed nucleic acid testing of SARS-COV-2 patients). The entire neural network can be roughly divided into two parts: the first part uses a pre-trained inception network to convert image data into one-dimensional feature vectors, and the second part uses a fully connected network with the main role of

classification prediction. The internal validation achieved a total accuracy of 82.9% with specificity of 80.5% and sensitivity of 84%. The external testing dataset showed a total accuracy of 73.1% with specificity of 67% and sensitivity of 74%.

Hemdan et al., 2020 introduced a new deep learning framework, namely COVIDX-Net to assist radiologists to automatically diagnose COVID-19 in X-ray images. The COVIDX-Net includes seven different architectures of DCNN models, such as modified VGG19, DenseNet121, InceptionV3, ResNetV2, Inception-ResNet-V2, Xception, MobileNetV2. Each deep neural network model is able to analyze the normalized intensities of the X-ray image, to classify the patient status either negative or positive COVID-19 cases. The results of COVIDX-Net verified that the best performance scores of deep learning classifiers are for the VGG19 and DenseNet201 models to automatically identify or confirm COVID-19 in 2D X-ray.

COVID-Net (*Wang et al., 2020*) is a deep CNN design tailored for the detection of COVID-19 cases from chest X-ray images that is open source and available to the general public. Developed by Linda Wang and Alexander Wong at the University of Waterloo and the AI firm DarwinAI in Canada, COVID-Net is one of the first open source network designs for COVID-19 detection from CXR images at the time of initial release (*available on Github* [41]). The model achieved an accuracy of 0.92 and a sensitivity of 0.91. The open access COVID-Net allows to accelerate the development of highly accurate yet practical deep learning solutions for detecting COVID-19 cases and accelerate treatment of those who need it the most.

Part of the existing studies use non-public datasets, others perform on complicated ML structures. In *Ucar et al., 2020* the authors used the same open dataset of COVID-Net (COVIDx) and introduced COVIDiagnosis-Net, a Bayes-SqueezeNet based model that comes forward with its light network design, tuned for the COVID-19 diagnosis with Bayesian optimization. The proposed system is composed of three main stages as offline augmentation of the raw dataset, training of the Bayesian optimization-based SqueezeNet model and decision-making of the network with the testing phase. The proposed method classifies the three-class X-ray images labeled as "normal" (no infection), "pneumonia" (bacterial or none-COVID viral infection) and "covid" (COVID-19 viral infection). Bayes SqueezeNet outperformed the existing COVID-Net test, achieving 0.98 of accuracy, 0.98 of sensitivity and 0.99 of specificity.

Caused by the nature of the Bayes' theorem, *Bayesian optimization* calculates the posteriori probability of a model with the aid of the learned data. Posteriori probability is proportional to the likelihood of observations and the multiplication of the prior probability. In brief, Bayesian optimization searches for the best model among many of them. The algorithm combines the prior distribution of the function with the samples of the prior knowledge to obtain the posteriors. New function evaluations are treated as observations used to update the posterior of the objective. At each iteration of the procedure, the acquisition function is cheaply optimized to determine the next point of function evaluation.

Also in *Nour et al., 2020*, the authors developed a diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. The proposed model is based on the CNN and can automatically reveal discriminative features on

chest X-ray images. Contrary to the generally used transfer learning approach, the proposed deep CNN model was trained from scratch. The extracted deep discriminative features were used to feed the ML algorithms, KNN, SVM, and decision tree to classify intances between COVID-19, normal and viral pneumonia patients. The hyperparameters of the ML models were optimized using the Bayesian optimization algorithm. The most efficient results were ensured by the SVM classifier with an accuracy of 98.97%, a sensitivity of 89.39% and a specificity of 99.75%.

In *Kang et al., 2020*, the authors proposed a study that used a different techinique from the others seen so far, developing a model to discriminate two types of CT images associated with COVID-19 and community-acquired pneumonia (CAP) with structured latent multi-view representation learning. Investigating different types of features extracted from CT images, it has been possible to learn latent representations which not only encode information of heterogeneous features but also reflect the class distribution. After obtaining the latent representation, the target was to train a latent-representation-based classifier which can diagnose the subjects between COVID-2019 and CAP. For simplicity, it has been employed a neural network with three fully-connected layers (FCNN) as the latent-representation-based classifier. The proposed multi-view representation learning technique, achieved a diagnosis performance of 95.5%, 96.6% and 93.2% in terms of accuracy, sensitivity and specificity, respectively. On top of that, rather stable performances are observed when varying the number of training data. In the future, the authors aim to consider diagnosis with more classes (i.e., normal, different COVID-19 severity, and CAP).

The main goal of *multi-view, or multi-modal, representation learning* is to learn a latent space which combines the information from different views available on the same data (*Li et al., 2019*). This technique allows a system to automatically discover the representations needed for feature detection or classification from raw data.

Ardakani et al., 2020 suggested a rapid and valid method for COVID-19 diagnosis. The authors used ten well-known pre-trained CNN to distinguish infection of COVID-19 from non-COVID-19 group. They trained and tested these 10 CNNs on the same dataset of 1020 CT images and compared the obtained results with the classification done by a radiologist. ResNet-101 could distinguish COVID-19 from non-COVID-19 cases with an AUC of 0.994 (sensitivity, 100%; specificity, 99.0%; accuracy, 99.51%). Xception achieved an AUC of 0.994 (sensitivity, 98.04%; specificity, 100%; accuracy, 99.02%). These results have been compared to the performance of a radiologist that achieved an AUC of 0.873 (sensitivity, 89.21%; specificity, 83.33%; accuracy, 86.27%). ResNet-101 can be considered as a high sensitivity model to characterize and diagnose COVID-19 infections, and can be used as an adjuvant tool in radiology departments.

During the review process, radiologists mistakenly diagnosed COVID-19 pneumonia or non-COVID-19 pneumonia from CT images. As seen so far, this led to assess even more the performance of radiologists, beside the performance of the algorithms. Indeed, studies like *Harrison et al., 2020*, focused their attention on the performance of radiologists rather than ML tools. The radiologists involved in this study were from China and United States and they were retrospectively capable of distinguishing COVID-19 from viral pneumonia at chest CT images with high specificity but

moderate sensitivity.

Part of the discussed studies in this section represent a starting point for further improvements in terms of generalization and dataset augmentation.
Nevertheless, there are some ML solutions already available on the market to aid radiologists in the screening and diagnosis phases.
Hospitals like Beijing Haidian Hospital in China, have been already facilitated with a deep learning-based computer-aided diagnostic system for pneumonia, called *InferRead CT Pneumonia, InferVision (2020)*.
InferVision's deep learning medical imaging platform significantly helped screen patients for the coronavirus in China. It acts as a second pair of eyes to identify multiple diseases from one set of chest scans due to ML that can provide a complete view of the clinical conditions of the patient (Figure 2.3). The implementation of this system, greatly improved the detection efficiency for patients highly suspected with COVID-19 by alarming the technician within 2 minutes when any suspected cases were found after CT examination.
InferVision Coronavirus ML solution has been in use also at the center of the epidemic outbreak at Tongji Hospital in Wuhan, along with sites in other cities, such as the Third People's Hospital of Shenzhen - in Shenzhen City, to accelerate pneumonia diagnosis and epidemic monitoring efforts.

In early February, Alibaba Group's research and innovation institute *DAMO Academy* [49] developed an ML-enabled system that could diagnose COVID-19 in 20 seconds with 96% accuracy via CT scans (Technology.org [50]). The detection and recognition process are extremely fast, where the CT-image processing task and diagnosis are done within 20 seconds, even when typical patient scans may include over 300 images. Usually a similar process lasts 5-15 minutes when carried out by a qualified doctor. With a training data set consisting of more than 5000 confirmed SARS-CoV-2 coronavirus computed tomography scans of the chest, the system classifies the results into confirmed coronavirus, the common flu, or other respiratory diseases. This ML system does two things: one is to track treatment responses in confirmed cases, and the other is to provide diagnoses for suspected cases. The resulting algorithm works impressively effectively, and boasts not only an exceptionally high detection rate, but also the ability to differentiate between 'regular' viral pneumonia and the new pandemic coronavirus. Alibaba already stated that the new tool would be adopted in more than 100 hospitals in several provinces of China [51].

All the proposed deep learning based models to detect and classify COVID-19 cases from X-ray images indicate that CNNs achieved encouraging results with high accuracy in COVID-19 detection. These systems have the aim to aid radiologist's activity, but they could be used also in remote places in countries affected by COVID-19 to overcome a shortage of radiologists and/or provide support in the hospitals.
From another point of view, even if ML classifiers and deep learning led to amazing achievements, a significant part of the studies discussed in this section require further data collection to test the generalization of the ML models on other patient populations before being able to provide a concrete implementation in real time diagnosis.

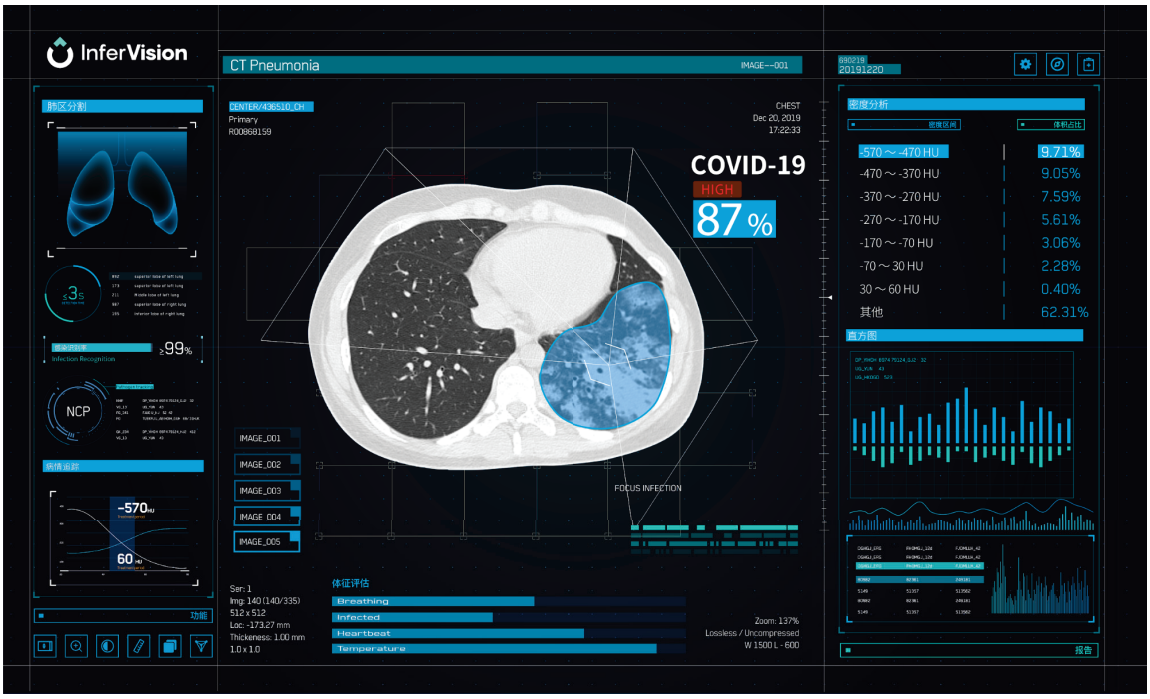

Figure 2.3: InferRead CT Pneumonia, InferVision. *Source:* [32]

These results support the idea that deep learning algorithms will become increasingly valuable as they become further integrated into the clinical diagnostic workflow and that the rapid recognition of pneumonia in patients may allow for early isolation precautions and administration of supportive therapies.
As next step, it would be important not only to predict the presence of COVID-19, but also the severity degree from X-ray images to further help monitor and treat patients.

2.3 ML technology in COVID-19 Contact Tracing

This section describes how mobile phone data can guide government and public health authorities in determining the best course of action to control the COVID-19 pandemic and in assessing the effectiveness of control measures.
If a person is diagnosed and is confirmed with COVID-19, the next important step is contact tracing prevention of the wider spread of the disease. To take control on the spread of SARS-Cov-2, contact tracing is an essential public health tool used to break the chain of virus transmission (*Contact Tracing in the context of COVID-19 - WHO 2020*).
The process of contact tracing is to identify and manage people who are recently exposed to an infected COVID-19 patient to avoid further spread. Generally, it provides a follow-up for 14 days since the exposure for each infected person.
Various infected countries come up with a digital contact tracing process with the mobile application, utilizing different technologies like Bluetooth, Global Positioning System (GPS), social graph, contact details, network based API, mobile tracing data and card transaction data (*Lalmuanawma et al., 2016*).
The digital contact tracing process can perform virtually real-time and much faster compared to the non digital system. All these digital apps are designed to collect individual personal data, which will be analyzed by ML tools to trace a person who is vulnerable to the novel virus due to his/her recent contact chain. However, there are still limitations in addressing the scenario, privacy, control over the data, and even data security breach.

According to *Nuria et al. 2020*, the use of mobile phone data into analytical efforts to control the COVID-19 pandemic can offer a critical contribution to four broad areas of investigations:

1. *Situational awareness* would benefit from increased access to previously unavailable population estimates and mobility information to enable stakeholders across sectors better understand COVID-19 trends and geographic distribution.

2. *Cause-and-effect* use cases can help stakeholders identify the key drivers and consequences of implementing different measures to contain the spread of COVID-19. They aim to establish which variables make a difference for a problem and whether further issues might be caused.

3. *Prediction* tasks would leverage real-time population counts and mobility data to enable new predictive capabilities and allow stakeholders to assess future risks, needs, and opportunities.

4. *Impact* assessment aims to determine which, whether, and how various interventions affect the spread of COVID-19 and requires data to identify the obstacles hampering the achievement of certain objectives or the success of particular interventions.

Passively generated mobile phone data has emerged as a potentially valuable data source to infer human mobility and social interactions and they are extremely useful to provide value throughout the whole epidemiological cycle intervals: investigation,

recognition, initiation, acceleration, deceleration and preparation (Figure 2.4 of *Pandemic Intervals Framework (PIF)*):

I. In the early *recognition and initiation phase* of the pandemic, the focus is on situational analysis and the fast detection of infected cases and their contacts. Research has shown that quarantine measures of infected individuals and their family members, combined with surveillance and standard testing procedures, are as effective as control measures in the early stages of the pandemic (*Koo et al. 2020*). Individual mobility and contact data offer information about infected individuals, their locations and social network. Contact data can be collected through mobile apps.

II. During the *acceleration phase*, when community transmission reaches exponential levels, the focus is on interventions for containment, which typically involves social contact and mobility restrictions. Aggregated mobile phone data is here crucial to assess the efficacy of implemented policies through the monitoring of mobility between and within affected municipalities. Mobility information also contributes to the building of more accurate epidemiological models that can explain and anticipate the spread of the disease. These models, in turn, can inform the mobilization of resources (e.g. respirators, intensive care units).

III. Finally, *during the deceleration and preparation phases*, as the peak of infections is reached, restrictions will likely be lifted. Near real-time data on mobility and hotspots will be important to understand how lifting and re-establishing various measures translate into behavior, especially to find the optimal combination of measures at the right time (e.g. general mobility restrictions, school closures, banning of large gatherings), and to balance these restrictions with aspects of economic vitality.

There are two different types of tracing technology:

- Local tracking: it uses GPS data to determine individuals who were in the same place at the same time (it's not accurate to indicate close physical contact)
- Proximity tracking: it uses bluetooth low energy (BLE) to determine whether two smartphones are close enough for their users to transmit the virus (it's better than GPS or cell site location).

Infectious disease surveillance web-based tools like *Flu Near You* have been rapidly adapted for COVID-19–specific collection (*Covid Near You*). Alternatively, web portals have been developed for researchers to report patient-level information on behalf of participants already enrolled in clinical registries, such as *The COVID-19 & Cancer Consortium*.

Along this line, researchers at the Massachusetts Institute of Technology and other collaborators are working on Private Kit: Safe Paths (*Barbar et al. 2020*), a free open-source and privacy-first contact-tracing technology that provides individuals with information on their interaction with diagnosed COVID-19 people while also empowering governments' efforts to contain an epidemic outbreak. The solution is a 'pull'

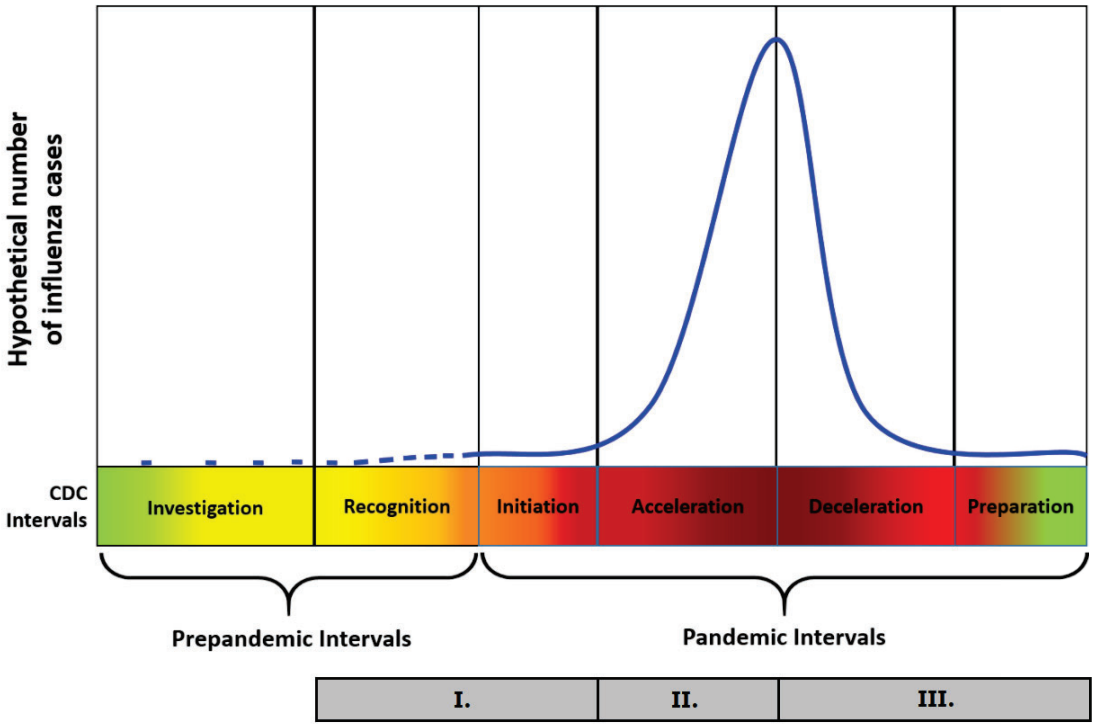

Figure 2.4: Pandemic intervals as defined by Center for Disease Control. *Source:* [55]

model where users can download encrypted location information about carriers so the users can self-determine their likely exposure to COVID-19 and coordinate their response with their doctor using their symptoms and personal health history.

Several contact tracing apps have been developed worldwide, such as *Immuni App in Italy*, *Radar COVID in Spain*, *Corona-Warn-App in Germany*, *StopCovid in France* , *Covidsafe in Australia*, *TraceTogether App in Singapore* , *CovidTracker in Ireland*, *Cocoa App in Japan*.
All these apps have the common aim of minimizing the spread of COVID-19 disease through community-driven contact tracing. This kind of technology helps people re-starting a careful social life without too much stress and further minimizing the spread of a disease. It works identifying people exposed to coronavirus, alerting the users who have had a risky exposure, even if they are asymptomatic and mainly to prevent the spread of the coronavirus.

Nevertheless, a smartphone application does not represent a random sampling of the population, an inherent limitation of any epidemiologic study that relies on voluntary participation (*Drew et al. 2020*).
In this context, digital solutions proposed for contact tracing app to reduce the spread of COVID-19 can be a powerful tool but there are still some limitations (*European Centre for Disease Prevention and Control*):

- only 79% of the population own a smartphone, with less than 40% in the over-65 age group;
- not everyone will have a smartphone, in particular the elderly, and not everyone

will have downloaded the app;
- some populations may be particularly wary of downloading the app;
- people with the app may not carry their phone with them at all times or may have the phone switched off;
- how to effectively deal with people who might have two or more devices;
- some apps may not work on older smartphones or operating systems;
- mobile apps have limitations in terms of their utility in investigating outbreaks in healthcare settings or long-term care facilities;
- how to ensure a limited number of people unnecessarily notified.

In addition, the use of mobile phone data raises legitimate public concerns about privacy, data protection and civil liberties. An intervention of this kind raises ethical questions regarding access, transparency, the protection and use of personal data, and the sharing of knowledge with other countries.
Anyway, as published by the European Parliament, the guidelines and toolboxes for developing any COVID-19-related apps, prepared by the Commission in cooperation with member states, *European Data Protection Supervisor*, and *European Data Protection Board*, aim at guaranteeing sufficient protection of data and limiting intrusiveness. *Guidance on Data Protection* is an essential part of the Commission guidelines, stressing that the apps must fully comply with EU data protection rules, most notably the *General Data Protection Regulation (GDPR)* and the *Privacy Directive*.
Recently, *Apple and Google* [76] have released a joint announcement describing their system to support Bluetooth based privacy-preserving proximity tracing across iOS and Android smartphones, where privacy, transparency, and consent are of utmost importance in this collaboration. Apple and Google will be launching a comprehensive solution that includes application programming interfaces (APIs) and operating system-level technology to assist in enabling contact tracing with user privacy and security central to the design. [77].

Mongan Institute at Massachusetts General Hospital launched a mobile *COVID Symptom Study* app that was co-developed by King's College and Zoe Global Ltd. It was deployed in the UK on March 24, 2020 (*Mongan Institute Website, 2020*). COVID Symptom Study, before known as COVID-19 Symptom Tracker, it's a COVID-19 epidemiological research mobile app that runs on Android and iOS. It collects data from both asymptomatic and symptomatic individuals and tracks in real time how the disease progresses by recording self-reported health information on a daily basis, including symptoms, hospitalization, RT-PCR test outcomes, demographic information and pre-existing medical conditions.
With broader implementation, data generated from the COVID Symptom Study app are increasingly being linked to the public health response within the National Health Service (NHS) in the United Kingdom. The COVID Symptom Study app is now a global public science project supported by the UK government and crowd-funding, with more than 4.2 million participants providing vital health data to help researchers and the NHS understand and beat COVID-19. Data collected by this app, allowed to generate insight and make prediction (see Section 2.5).

COVID Symptom Study app follows an approach based on a real-time epidemiology (*Ferretti et al. 2020*). The authors developed a mathematical model for infectiousness (the mean rate at which individuals infect others at a certain time after they

themselves were infected) to estimate the basic reproductive number and to quantify the contribution of different transmission routes. According to them, practical and logistical factors will determine whether a contact tracing app is sufficient to control viral spread on its own, or whether additional measures (e.g., physical distancing) are required. The results of this mathematical model of infectiousness and interventions are presented in this *Shiny Web Application* [81], through which users can test the effects of alternative infection parameters.
Furthermore, the author explained how an instant contact tracing app works. Proximity events between two phones running the app are recorded. Upon an individual's COVID-19 diagnosis, contacts are instantly, automatically, and anonymously notified of their risk and asked to self-isolate. The core functionality is to replace a week's work of manual contact tracing with instantaneous signals transmitted to and from a central server. Figure 2.5 shows how epidemic control can be achieved if enough people use a contact-tracing app that builds a memory of proximity contacts and immediately notifies contacts of positive cases. The contacts of individual A (and all individuals using the app) are traced using BLE connections with other app users. Individual A requests a SARS-CoV-2 test (using the app) and that person's positive test result triggers an instant notification to individuals who have been in close contact. The app advises isolation for the case (individual A) and quarantine of the individual's contacts. Coronavirus diagnoses are communicated to the server, enabling recommendation of risk-stratified quarantine and physical distancing measures in those now known to be possible contacts, while preserving the anonymity of the infected individual.

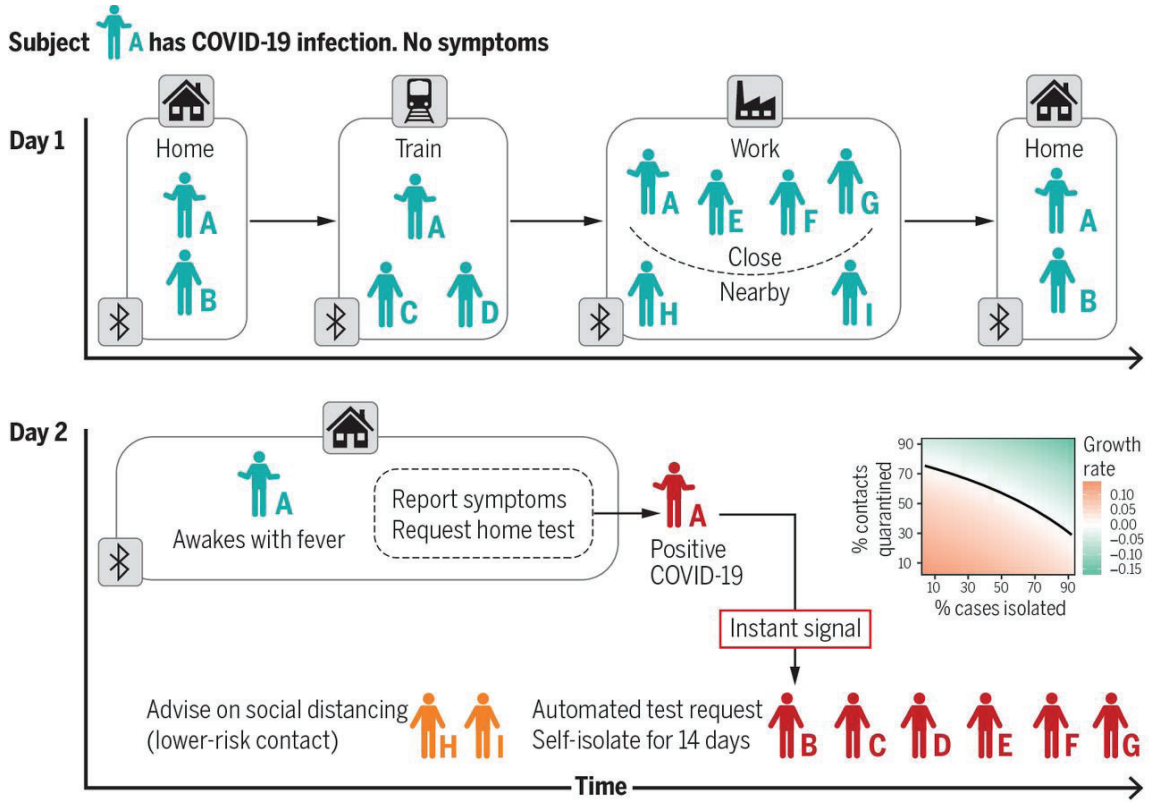

Figure 2.5: A schematic of app-based COVID-19 contact tracing. *Source:* [80]

Anyway, in the context of a mobile phone app, Figure 2.5 paints an optimistic scenario, since there is no delay between case confirmation and notification of contacts. Anyway, digital contact tracing smartphone app promises to be an effective method for achieving COVID-19 epidemic control.

On September 24th the *NHS COVID-19 app* was launched in England and Wales, that requires to be used with COVID Symptom Study app. The difference between them is that COVID Symptom Study app is a large-scale scientific project to understand and map COVID-19, providing estimated national and local COVID-19 cases based on algorithmic prediction. It asks users to log daily health updates and records a wide range of symptoms and triggers an invitation to book a test if people report symptoms that might be caused by COVID-19. It does not use your phone's Bluetooth, GPS, location or contacts and does not track you as you move around.
The NHS COVID-19 app uses Apple/ Google API exposure notification system which allows an app to measure the distance and duration between two devices, alerting users if someone they have been in close contact has had later positive tests for the virus using random unique IDs. If any of those users later tests positive for coronavirus, other app users they may have been in contact with will then receive an anonymised exposure alert with advice on what to do next. A scientific calculation, using an algorithm, has been developed by scientists to work out which app users are 'close contacts'.
The NHS COVID-19 app supports the government's efforts to control the spread of the virus through testing and contact tracing (NHS COVID-19 app is only available in England and Wales, there are different contact tracing apps in Scotland and Northern Irelands called Test and Protect app and StopCOVID NI app, respectively).
Oxford University team proposed the development of a digital contact tracing system to the NHS in order to detect close proximity contacts and alert people of their risk of coronavirus infection (*Coronavirus Fraser Group* [83]). Their research showed that COVID-19 spreads before we develop symptoms, and so we need a fast and effective test and trace system. Their epidemic model provided the scientific basis for configuration of the NHS COVID-19 app and enabled policy makers to consider how best to scale-up and adjust the contact tracing app alongside other vital infection control measures. Research analyses of app predicted that with as little as 15% of the population using the NHS COVID-19 contact tracing app, it started to have a meaningful impact to reduce the number of new coronavirus infections, hospitalisations and deaths.

The Coronavirus Fraser Group involves scientists based at Oxford University's Nuffield Department of Medicine, involves experts in infectious disease epidemiology, medicine, virology, immunology, mathematical modelling, phylogenetics, behavioural economics, ethics and communications.
They later developed an individual-based epidemic simulation *Hinch et al. 2020* which enables epidemiologists, app designers and policy makers to compare a variety of algorithm configurations for digital contact tracing under a range of assumptions about the epidemic, the technology, a country's demographics, and user engagement.
Contact tracing is difficult to model accurately in a simple mathematical model, because a history of previous contact events must be recalled. Therefore an individual-based model offers the most parsimonious method for accurately capturing the effects of this intervention. Other interventions can be modelled simultaneously in the same

framework. The individual-based model code is open source and can be accessed on Github.

Individual-based model (IBM) has been used to simulate and to design control strategies for dynamic systems that are subject to stochasticity and heterogeneity, such as infectious diseases. In the IBM, an individual is represented by a set of specific characteristics that may change dynamically over time. This feature allows a more realistic analysis of the spread of an epidemic. An IBM allows to emulate simulation experiences in a computational environment taking into account the characteristics and interaction of each individual. In this way, aspects that are generally ignored by other models can be considered making the system more realistic.

A retrospective big data infoveillance study is *Mackey et al., 2020*, where the authors analyzed Twitter for the purposes of characterizing conversations regarding self-reporting of COVID-19-related symptoms, access to testing, and experiences with purported recovery for the purposes of digital contact tracing. Tweets were collected from the Twitter public streaming application programming interface in March 3-20, 2020, filtered for general COVID-19-related keywords and then further filtered for terms that could be related to COVID-19 symptoms as self-reported by users. Tweets were analyzed using an unsupervised ML approach called the *biterm topic model (BTM)*, where groups of tweets containing the same word-related themes were separated into topic clusters related to COVID-19 symptoms, testing, and recovery. BTM was used to identify relevant topic clusters. Then, tweets in these clusters were extracted and manually annotated for content analysis and analyzed for statistical and geographical characteristics (63% of the analyzed tweets came from USA).
Many users reported symptoms they thought were related to COVID-19, but they were not able to get tested to confirm their concerns. In the absence of testing availability and confirmation, accurate case estimations for that period of the outbreak may never be known.

In *McLachlan et al. 2020*, the authors built on some of the digital solutions already under development, with the addition of a Bayesian network model that predicts likelihood for infection supplemented by traditional symptoms and contact tracing. This solution focuses on enabling users to diagnose the possible presence of COVID-19 themselves, through a causal probabilistic model: a Bayesian network.
The app provides the user with information about how likely it is s(he) has or has not mild or severe COVID-19. When this probabilistic information is combined with data about the GPS-location of the smartphone, together with information about the age group of the person it is possible to provide information about the distribution of mild and severe COVID-19.
The main difference between this solution and the others presented before is that in the latter the contact tracing apps act retrospectively. The app advises the user they were previously in close contact with an infected, and in the case of COVID-19, this advice often comes only after they have already begun asymptomatically shedding the disease. The solution that integrates the retrospective contact tracing app with symptom tracking and a Bayesian network, instead, provides the users with a prospective view of the probability that s(he) may have developed COVID-19.

An efficient sampling algorithm can be used to predict the spread of infectious diseases such as COVID-19 under different testing and tracing strategies, social distancing measures, and business restrictions, given location or contact histories of individuals. That's the idea adopted by *Lorch et al., 2020*. The authors introduced a novel modeling framework for studying epidemics that is specifically designed to make use of fine-grained spatiotemporal data. Experiments using measured COVID-19 data and mobility patterns from Tübingen, a town in the southwest of Germany, demonstrate that this model can be used to quantify the effects of tracing, testing, and containment strategies at an unprecedented spatio-temporal resolution.
This model uses marked temporal point processes to represent individual mobility patterns and the course of the disease for each individual in a population. Building on this algorithm, they used Bayesian optimization to estimate the risk of exposure of each individual at the sites they visit, the percentage of symptomatic individuals, and the difference in transmission rate between asymptomatic and symptomatic individuals from historical longitudinal testing data. Simulator for the spatio-temporal model for COVID-19 is available on *Github* [89] to facilitate research and informed policy-making.

To sum up, mobile apps should be one tool among many general preventative population measures such as physical distancing, enhanced hand and respiratory hygiene, and regular decontamination.
In addition, using mobile apps for contact tracing helps to make people more aware about the consequences of their contacts and their movements, allowing to contain the spread of the virus and individual protection.
However, a legitimate concern is around the ethics, potential loss of privacy and long-term impact on civil liberties resulting from the use of individual mobile data to monitor COVID-19. That is just one of the reasons why the proportion of people that are using these apps is still very little. Privacy-aware and ethically acceptable solutions to use mobile phone data should be prepared and vetted in advance and then, read on national and international levels, to rapidly act when the crisis hits. After the pandemic has subsided, mobile data will be helpful for post-hoc analysis of the impact of different interventions on the progression of the disease, and cost-benefit analysis of mobility restrictions.
Furthermore, in the future, it will be important also to share knowledge and collaborate between different countries. On 13 May 2020, the Commission listed the use of contact-tracing apps among the guidelines for resuming travel in Europe and noted they have to be interoperable so that people can use them to be alerted wherever in Europe they are (*European Parliament Website, 2020* [90]).
Other studies continue to explore the utility of social media and other forms of electronic data to estimate COVID-19 disease severity, symptoms, trends and case counts (see Section 2.5).

2.4 ML technology in COVID-19 Drugs/Vaccines Development

Since the coronavirus epidemic fury started, researchers and healthcare experts around the globe ubiquitously urged to develop a possible choice to tackle the SARS-CoV2 pandemic with the development of drugs and vaccines.
ML technology constitutes to be an enthralling road to speed up this process.
The studies discussed in this section may result hard to read by a non-medical expert audience. Therefore, with the support of the medical litterature, some scientific concepts involved in the following research works are explained in brief.
The following concepts are shown in Figure 2.6 and Figure 2.7.
Proteomic methods are useful to study virus-virus and virus-host interactions, providing valuable insight into the protein interactions that allow viruses to infect and replicate within the host cell (*Maxwell et al. 2007*).

Genome analysis is a powerful tool for understanding viral disease outbreaks, for example providing informations about how a disease began and how it is transmitted (*Wohl et al. 2016*).
The main difference between genomics and proteomics is that genomics is the study of the entire set of genes in the genome of a cell whereas proteomics is the study of the entire set of proteins produced by the cell (*Tyers et al. 2003*).
Both of them have important implications for COVID-19 clinical trials. Studying the genome evolution, as well, of the new coronavirus allows scientists to obtain important information for developing new drugs and vaccines research and to help those searching for synthetic antibodies to treat infection (*Medical Express, 2020* [96]).

The transcriptome analysis is an integral part of almost all genomic studies of disease and biological processes. Transcriptome of a cell or a tissue is the collection of RNAs transcribed in it. Since transcriptome is dynamic, it's a good representative of the cellular state (*Anuj et al., 2019*).

Spike (S) protein represented a key target for developing therapeutics to block viral entry and inhibit membrane fusion in MERS (*Duet al. 2016*). In the SARS vaccine development, it has been frequently used as the vaccine antigen due to its ability to induce neutralizing antibodies that prevent host cell entry and infection (*Du et al., 2009*). The S protein of SARS-CoV-2 is widely considered as a promising antigen (*Zhang et al., 2020*). The term antigen is derived from antibody generation, referring to any substance that is capable of eliciting an immune response.

An epitope or antigenic determinant is a group of amino acids or other chemical groups exposed on the surface of a molecule, frequently a protein, which can generate an antigenic response and bind antibody. In brief it means that is capable of stimulating an immune response (*Carolyn et al., 1996*).

As aforementionated in Section 2.1, the study *Ong et al., 2020* is a practical example of how previous studies related to coronavirus family, are helpful to contrast the novel COVID-19. In this paper the authors used Vaxign reverse vaccinology tool and the newly developed Vaxign-ML tool to predict COVID-19 vaccine candidates.
Reverse vaccinology (RV) is a widely used approach to identify potential vaccine can-

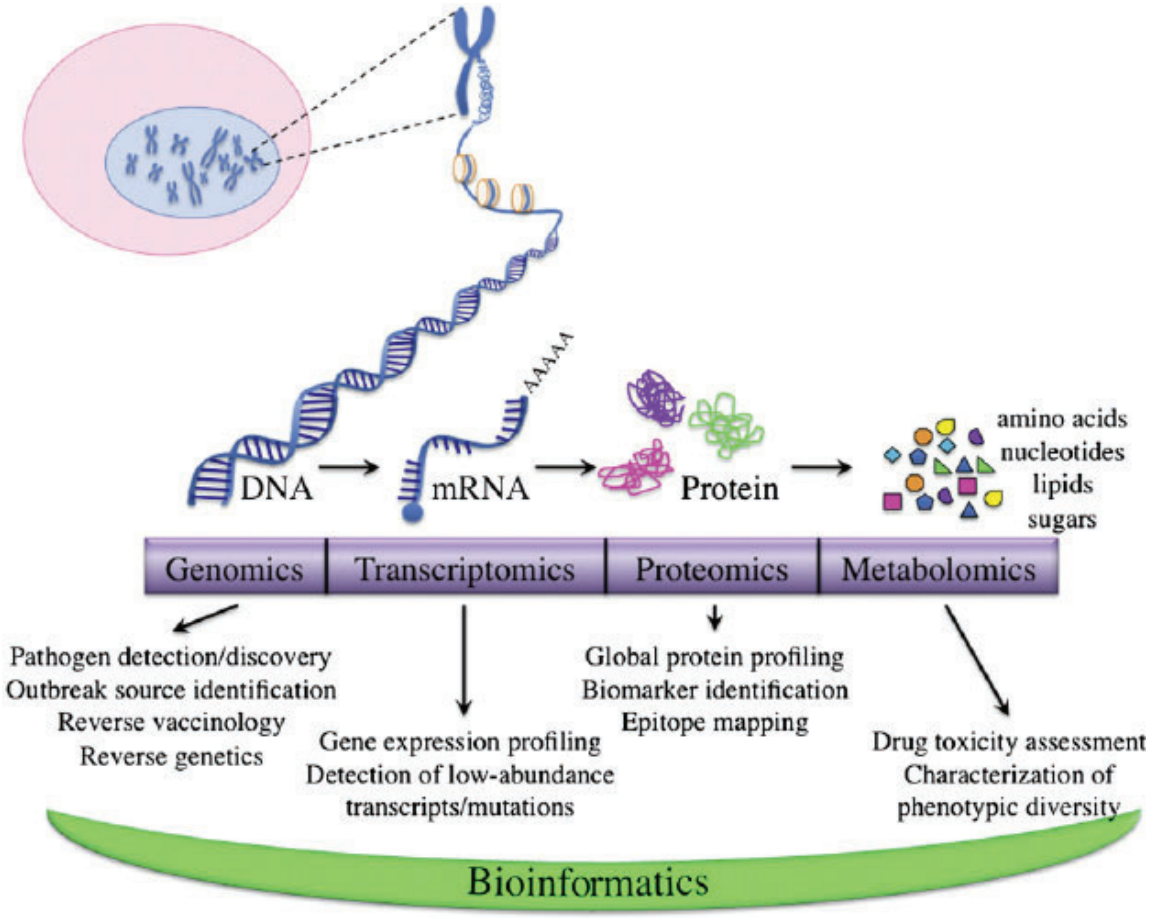

Figure 2.6: A top-down explanation of "omics." *Source: Fontana et al., 2012*

didates by screening the proteome of a pathogen through computational analyses. Starting from the studies about vaccine development against SARS and MERS, the authors investigated the entire proteome of SARS-CoV-2 and six proteins, including the S protein and five non-structural proteins.
The S, nsp3, and nsp8 proteins were predicted by both tools to induce high protective antigenicity. Besides the commonly used S protein, the nsp3 protein has not been tested in any coronavirus vaccine studies and was selected for further investigation. The nsp3 was found to be more conserved among SARS-CoV-2, SARS-CoV, and MERS-CoV than among 15 coronaviruses infecting human and other animals.
Five supervised ML classification algorithms, including logistic regression, SVM, kNN, random forest, and XGB were trained on a proteins dataset. The best performing XGB model was selected to predict the protegenicity score of proteins: a protein with protegenicity score over 0.9 is considered as strong vaccine candidate (weighted F1-score > 0.94 in (N5CV) nested five-fold cross-validation).

The retrospective study using the past SARS-CoV and MERS-CoV data demonstrated that the ML based method developed in *Ge et al., 2020* can successfully predict effective drug candidates against a specific coronavirus.
The authors developed a data-driven drug repositioning framework, which applies both ML and statistical analysis approaches to systematically integrate and mine large-scale knowledge graph, literature and transcriptome data to discover the po-

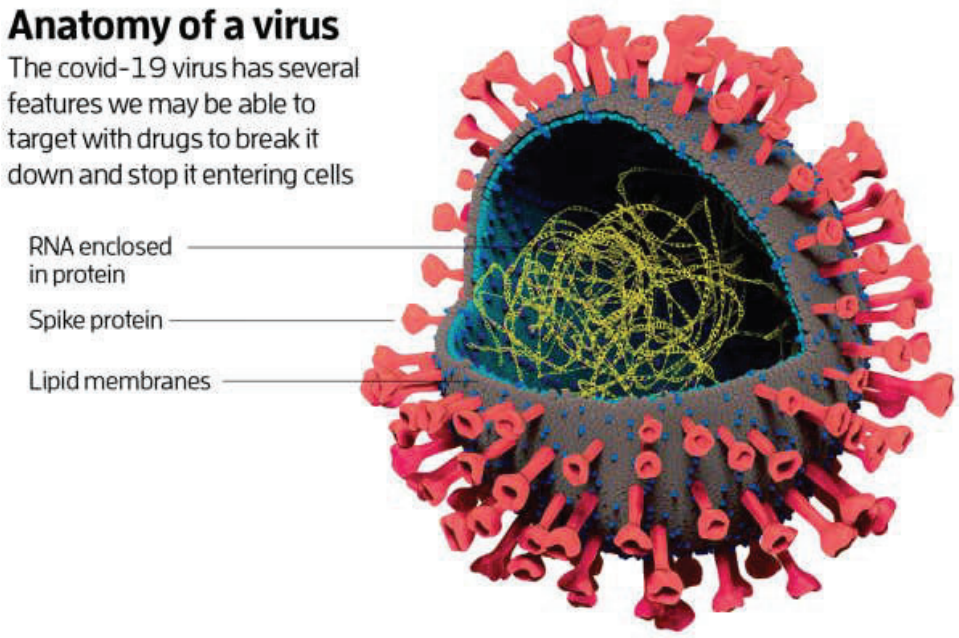

Figure 2.7: SARS-CoV-2 structure. *Source: NewScientist, 2020* [92]

tential drug candidates against SARS-CoV-2.
It has been discovered that the PARP1 inhibitor CVL218 can serve as a potential therapeutic agent for the treatment of COVID-19. CVL218 is able to suppress the CpG-induced production of IL-6, which has been reported previously to be of high relevance to the viral pathogenesis of COVID-19, especially for those intensive care unit (ICU) patients infected by SARS-CoV-2. After constructed a virus related knowledge graph consisting of drug-target interactions, protein-protein interactions and similarity networks from publicly available databases, a network-based knowledge mining algorithm is applied to predict an initial list of drug candidates that can be potentially used to treat SARS-CoV-2 infection. Next, they further narrowed down the list of drug candidates with the previously reported evidences of antiviral activities based on the text mining results from the large-scale literature texts, which were derived through a deep learning based relation extraction method named BERE.

EVQLV [104] is a startup creating algorithms capable of computationally generating, screening, and optimizing hundreds of millions of therapeutic antibodies to discover treatments most likely to help those infected by the virus responsible for COVID-19.
Studies show that it takes an average of five years and a half billion dollars to discover and optimize antibodies in a lab. ML algorithms can significantly reduce that time and cost, speeding up the first stage of the process.
ML helps with antibody discovery, rapidly screening for therapeutic antibodies with a high probability of success (*Columbia University, 2020* [105]). EVQLV chief technology officer states that their algorithms reduce the likelihood of drug-discovery failure in the lab: failing in the computer as much as possible help to reduce the possibility of downstream failure in the laboratory and saving a significant amount of time from laborious and time-consuming work.
Moreover, EVQLV collaborates with *Immunoprecise Antibodies (IPA)* (a company focused on the discovery of therapeutic antibodies to develop therapeutic candidates

against COVID-19) in such a way that the collaboration will accelerate the effort to develop therapeutic candidates against COVID-19. EVQLV will identify and screen hundreds of millions of potential antibody treatments in only a few days far beyond the capacity of any laboratory. IPA will produce and test the most promising antibody candidates.

In *Ke et al., 2020*, the authors used therapeutic medicine that has prior use experiences in patients in order to resolve the current pandemic situation before it could become worsening. ML technology is hereby applied to identify the marketed drugs with anti-coronavirus activities by using two different learning databases (one using the 3C-like protease constraint and other data-holding records of infected SARS-CoV, SARS-Cov-2, influenza, and human immunodeficiency virus (HIV)). Using a deep neural network on the eighty old drugs with potential for COVID-19 treatment, the study suggested that eight drugs, i.e., vismodegib, gemcitabine, clofazimine, celecoxib, brequinar, conivaptan, bedaquiline and tolcapone are found virtually effective against feline infectious peritonitis coronavirus. Furthermore, other five drugs like homoharringtonine, salinomycin, boceprevir, tilorone and chloroquine are also found operational during ML experimental environment. Three types of molecular descriptors, extended connectivity fingerprint (ECFP), functional-class fingerprints (FCFPs), and octanol–water partition coefficient (ALogP_count), were found by ML.

In *Beck et al., 2020* the authors used a pre-trained deep learning-based drug-target interaction model called Molecule Transformer-Drug Target Interaction (MT-DTI) to identify commercially available drugs that could act on viral proteins of SARS-CoV2. The proposed model employed a deep learning algorithm on 3C-like proteinase of COVID-19 and the Food and Drug Administration (FDA) approved 3,410 existing drugs available in the market. The result revealed that a popular antiretroviral drug used to treat HIV called Antazanavir is the best drug for COVID-19 medication. It was also found that various antiviral compounds like Kaletra might be utilized for the medicine of COVID-19 human patients.

In *Randhawa et al., 2020* the authors identified an intrinsic COVID-19 virus genomic signature and used it together with an ML-based alignment-free approach for an ultra-fast, scalable, and highly accurate classification of whole COVID-19 virus genomes. The proposed method combines supervised ML with digital signal processing (MLDSP) for genome analyses, augmented by a decision tree approach to the ML component, and a Spearman's rank correlation coefficient analysis for result validation. Each genomic sequence is mapped into its respective genomic signal (a discrete numeric sequence) using a numerical representation. Then, classification of COVID-19's novel pathogens is performed using six supervised-learning based classification models: linear discriminant, linear SVM, quadratic SVM, fine KNN, subspace discriminant, and subspace KNN.
This method achieves 100% accurate classification of the COVID-19 virus sequences, and discovers the most relevant relationships among over 5000 viral genomes within a few minutes, using raw DNA sequence data alone, and without any specialized biological knowledge, training, gene or genome annotations. It suggests that, for novel viral and pathogen genome sequences, this alignment-free whole-genome ML approach can provide a reliable real-time option for taxonomic classification of the COVID-19 virus as Sarbecovirus, within Betacoronavirus, as well as quantitative ev-

idence supporting a bat origin hypothesis. These results are obtained through a comprehensive analysis of over 5000 unique viral sequences, using different dataset to validate the model. This study suggests that such alignment-free approaches to comparative genomics can be used to complement alignment-based approaches when timely taxonomic classification is of the essence, such as at critical periods during novel viral outbreaks.

Ongoing vaccine development efforts are focused on raising an immune response against the spike protein. In *Prachar et al., 2020* the authors investigated potential SARS-CoV-2 epitopes. To assess whether current peptide-HLA prediction tools could be suitable for identification of epitopes relevant in a vaccine against SARS-CoV-2, they evaluate predicted binders using fifteen prediction tools tested on a relevant dataset of peptides from the SARS-CoV-2 genome. The analysis revealed that ANN achieved the highest score (AUC = 97.47) to identify 174 peptides with epitope potential in a future COVID-19 vaccine. This creates a challenge for vaccine development efforts, especially for the design of epitope vaccines, where only a limited number of epitopes may be included.

The selected review papers adopted various methodologies and technologies addressing the classical method of classification based on statistics to advanced modern ML algorithms. The use of computational tools, was found to be more active in predicting the reusability of an existing old drug on COVID-19 medication and dramatically minimize the level of a risk factor in the development of more cost-effective process.
During this urgency, the use of ML can augment the drug development process by lessening the time slot on discovering a supplementary treatment and medication for the carrier by drawing a vast probability over security, manageability, and clinical information on the existing drug compound.
Issues and challenges found in this area states out how big is the impact of ML tools to identify the drug candidates against SARS-CoV-2. Furthermore it's remarked how important is a close cooperation between ML and medical resources.

2.5 ML technology in COVID-19 Prediction and Forecasting

Despite that models for COVID-19 diagnosis and prognosis were developed, the lack of prediction models makes the early detection difficult. Given the increasing caseload, there is an urgent need to augment clinical skills with the support of ML models. For instance, prediction of severe/critical cases before symptom occurs, can effectively save medical resources, identifying from among the many mild cases the few that will progress to critical illness.
In addition, forecasting and monitor the ongoing growth of COVID-19 cases, can assist the managers in the decision-making support systems, developing strategic planning in the public health system to avoid deaths.
The research works discussed in this section involve the application of ML and statistical models with the aim of providing support to healthcare in early identifying critical and severe symptoms, but also to government resources engaged, in contrasting the spread of the COVID-19 and their economical consequences.

In *Sun et al., 2020*, the authors developed a model for predicting the COVID-19 patients who will progress into severe/critical cases. The dataset contains diagnosed patients with PCR in Shanghai where temperature, heart rate, blood pressor and other clinical and laboratory features were collected. Among all the clinical and laboratory features analyzed, 36 of them were found to be statistically significantly associated with the clinical outcome (severe/critical symptom) of these patients infected of COVID-19. Indicators associated with severe/critical symptoms are mainly thyroxine, immune related cells and products. To select the features a combination of less than five features was considered. So, exhaustive attack method (numeration method, which means list all combinations of the features) by combining 2, 3, and 4 features was used. Then, SVM was applied to develop model in training set by the selected features, and to predict the outcome in testing set. The combination of age, GSH, CD3 percentage reached the best performance: 0.9757 of AUC in testing and 100% of recall. The limitations of this study stay again is the relatively small sample siz: a total of 336 patients diagnosed as COVID-19 with PCR Kit were enrolled in this study (310 non-severe/critical cases and 26 severe/critical cases).

In *Jiang et al., 2020*, the authors presented a first step towards building an ML framework, in order to predict patients at risk for more severe illness on initial presentation and to identify the combinations of clinical characteristics of COVID-19 that predict outcomes. They used filter and wrapper methods to select the best subset of clinical characteristics that most contribute to the accuracy of predicting who will develop acute respiratory distress syndrome (ARDS), a severe outcome in COVID-19. The most predictive clinical features on presentation, are: a mildly elevated alanine aminotransferase (ALT) (a liver enzyme), the presence of myalgias (body aches), and an elevated hemoglobin (red blood cells), gender, temperature, Na+, K+, lymphocyte count, creatinine, age and white blood count, in this order. SVM was the classifier with the best performance with 0.80 of accuracy. The other algorithms taken into account were, in order of performance obtained: kNN, decision tree, random forest and logistic regression.

To support decision making and logistical planning in healthcare systems, the study of *Yan et al., 2020* suggests a simple and operable decision rule to quickly predict pa-

tients at the highest risk, allowing them to be prioritized and potentially reducing the mortality rate. The authors leveraged a database of blood samples from 485 infected patients in the region of Wuhan, China, to identify crucial predictive biomarkers of disease mortality.
This study uses a supervised XGBoost classifier as the predictor model. XGBoost is a high-performance ML algorithm that benefits from great interpretability potential due to its recursive tree-based decision system. The importance of each individual feature in XGBoost is determined by its accumulated use in each decision step in trees. This computes a metric characterizing the relative importance of each feature, which is particularly valuable to estimate features that are the most discriminative of model outcomes, especially when they are related to meaningful clinical parameters. To evaluate the markers of imminent mortality risk, the authors assessed the contribution of each patient parameter to decisions of the algorithm. Features were ranked by multi-tree XGBoost according to their importance. ML tools selected three biomarkers that predict the mortality of individual patients more than 10 days in advance with more than 90% accuracy: lactic dehydrogenase (LDH), lymphocyte and high-sensitivity C-reactive protein (hs-CRP). In particular, relatively high levels of LDH alone seem to play a crucial role in distinguishing the vast majority of cases that require immediate medical attention.
The results show that the model is able to accurately identify the outcome of patients, regardless of their original diagnosis upon hospital admission. Notably, the performance of the external test set is similar to that of the training and validation sets, which suggests that the model captures the key biomarkers of patient mortality.

Wu et al., 2020 presented a model to rapidly identify COVID-19 infection through random forest algorithm based on clinical available blood test results. Eleven key blood indices were extracted through random forest algorithm to build the final assistant discrimination tool from 49 clinical available blood test data which were derived by commercial blood test equipments. Eleven laboratory parameters could achieve high-level performance to accurately identify COVID-19. Once the virus invaded the body, the composition of the blood would change. The abnormalities sometimes cannot be observed directly from these routine laboratory blood tests, but the parameters could play a material impact on identifying the disease after processed by ML. It also indicated that the parameters were highly correlated to the infection of SARS-CoV-2 and the powerful ability of the tool to effectively distinguish patients with COVID-19 and patients with pneumonia. The statistical variations of 11 parameters were analyzed from different groups including group of COVID-19, group of common pneumonia, group of tuberculosis, and group of lung cancer. Except for glucose (GLU) and magnesium (Mg), almost all the parameters manifested the significant differences between COVID-19 and general pneumonia.
The method presented robust outcome to accurately identify COVID-19 from a variety of suspected patients with similar CT information or similar symptoms, combining ML algorithms and laboratory parameters.
After multiple verification, the reliability and repeatability of the tool has been fully evaluated, and it has the potential to develop into an emerging technology to identify COVID-19 and lower the burden of global public health. The tool also demonstrated its outstanding performance on an external validation set that was completely independent of the modeling process, with sensitivity, specificity, and overall accuracy of 0.9512, 0.9697, and 0.9595, respectively and AUC of 0.9926, a sensitivity of 1.0000

and a specificity of 0.9444 in the test set.

Another study that proposed a model to predict the risk for critical COVID-19, is *Assaf et al., 2020*. The primary outcome was risk for critical disease, defined as mechanical ventilation, multi-organ failure, admission to the ICU, and/or death. In order to predict deterioration, three different machine-learning algorithms were used: ANN, random forest, and classification and regression tree (CRT). Patients with severe COVID-19 at admission, based on low oxygen saturation, low partial arterial oxygen pressure, were excluded. Patients included had confirmed COVID-19 infection based on RT-PCR for the SARS-CoV-2 ribonucleic acid (RNA). The most contributory variables to the models were APACHE II score, white blood cell count, time from symptoms to admission, oxygen saturation and blood lymphocytes count. The APACHE II score estimates ICU mortality based on a number of laboratory values and patient signs taking both acute and chronic disease into account. It is applied within 24 hours of admission of a patient to an intensive care unit (ICU): an integer score from 0 to 71 is computed based on several measurements; higher scores correspond to more severe disease and a higher risk of death. Machine-learning models demonstrated high efficacy in predicting critical COVID-19 compared to the most efficacious tools available. ANN demonstrated accuracy improvement of 11.0% from the APACHE II score with sensitivity, specificity and accuracy of 59.0%, 96.3% and 90.5%, respectively, reaching AUC of 0.92. Random forest classification achieved an accuracy improvement of 12.0% from the APACHE II score with sensitivity, specificity and accuracy of 75.0%, 95.8% and 92.9%, respectively, with a AUC of 0.93. Finally, the best performance was obtained by a CRT model that reached sensitivity, specificity and accuracy of 88.0%, 92.7% and 92.0%, respectively, with AUC of 0.90.

In *Ribeiro et al., 2020*, the authors presented a forecast model for cumulative confirmed cases of COVID-19 in Brazil, to assist governors in decision-making with the aim of containing the pandemic and planning strategies concerning the health system. In this paper, six ML approaches named moving average (ARIMA), cubist regression (CUBIST), random forest (RF), ridge regression (RIDGE), support vector regression (SVR), and stacking-ensemble learning were employed in the task of forecasting one, three, and six-days-ahead the COVID-19 cumulative confirmed cases in ten Brazilian states with a high daily incidence.
Mean absolute error, and symmetric mean absolute percentage error criteria were adopted to evaluate the performance of the compared approaches based on the results of predictions over ten datasets and three forecasting horizons adopted. SVR and stacking-ensemble learning model were the best suitable tools to forecast COVID-19 cases for most of the adopted states, thanks to the fact that these approaches were able to learn the nonlinearities inherent to the evaluated epidemiological time series. In the stacking-ensemble learning modelling, as base-learners six ML approaches named CUBIST, RF, RIDGE, SVR are trained and their forecasting are used as inputs for a meta-learner Gaussian process (GP) giving error in range of 0.87%-3.51% in one-day prediction, 1.02%–5.63% in three-days prediction and 0.95% -6.90% six days ahead.

Stacked generalization or stacking-ensemble learning is an ensemble-based approach which combines through a meta-learner the predictions of a set of weak models (base-learners) to obtain a stronger learner. This approach usually operates into two levels,

where in the first level the base-learners are trained and their predictions are obtained. In the next stage, a meta-learner uses, as inputs, the predictions of the previous level in the training phase. The stacking predictions are obtained from the meta-learner. The main advantage of the stacking-ensemble learning is that, this approach can improve the accuracy and additionally reduce error variance.

In *Chimmula et al., 2020* to help frontline health workers and government policy makers, the authors proposed a time series model to forecast COVID-19 transmission in Canada using long short-term memory networks (LSTM). LSTM network is a deep learning approach, used in this paper, to forecast the future COVID-19 cases, evaluate the key features to predict the trends and possible stopping time of the current COVID-19 outbreak in Canada and around the world. Based on the results obtained, the predicted possible ending point of this outbreak was around June 2020 (this paper is based on the available data until March 31, 2020). The model generated predictions with a root mean square error of about 45.70 with an accuracy of 92.67% for long term predictions.

An *LSTM network* has the capability of addressing the limitations of traditional time series forecasting techniques by adapting nonlinearities of the given COVID-19 dataset and can result in state of the art results on temporal data. Each block of LSTM operates at different time steps and passes its output to the next block until the final LSTM block generates the sequential output.

Real-time forecasting and risk assessment of COVID-19 is proposed in *Chakraborty et al., 2020*. This study carried two analyses: generating short term (real-time) forecasts of the future COVID-19 cases for multiple countries and provide risk assessment (in terms of case fatality rate) of the novel COVID-19 for some profoundly affected countries by finding various important demographic characteristics of the countries along with some disease characteristics. To solve the first problem, the authors presented a hybrid approach based on autoregressive integrated moving average model and wavelet-based forecasting model that can generate short-term (ten days ahead) forecasts of the number of daily confirmed cases for Canada, France, India, South Korea, and the UK. In the second problem, an optimal regression tree algorithm has been applied to find essential causal variables that significantly affect the case fatality rates for different countries. The estimates of the performance metrics for the fitted tree are as follows: roat mean square error = 0.013, R-square = 0.896, Rsquare-adjusted = 0.769.

Regression tree has built-in feature selection mechanism, easy interpretability, and it consists in three stages. The first stage involves growing the tree using a recursive partitioning technique to select essential variables from a set of possible causal variables and split points using a splitting criterion. The standard splitting criterion for regression tree is the mean squared error. After a large tree is identified, the second stage of regression tree methodology uses a pruning procedure that gives a nested subset of trees starting from the largest tree grown and continuing the process until only one node of the tree remains. The cross-validation technique is popularly used to provide estimates of future prediction errors for each subtree. The last stage of the regression tree methodology selects the optimal tree that corresponds to a tree yielding the lowest cross-validated or testing set error rate. To avoid instability of

trees in this stage, trees with smaller sizes, but comparable in terms of accuracy, are chosen as an alternative. This process can be tuned to obtain trees of varying sizes and complexity. A measure of variable importance can be achieved by observing the drop in the error rate when another variable is used instead of the primary split. In general, the more frequent a variable appears as a primary split, the higher the importance score assigned.

Also public social media data can be usefully harnessed to predict infection cases and inform timely responses. In *Shen et al., 2020* the aim was to collect and analyze posts related to COVID-19 on Weibo, a popular Twitter-like social media site in China, to predict COVID-19 case counts. The authors used an ML classifier (decision trees, kNN, MLP, SVM and random forest) to identify sick posts, in which users report their own or other people's symptoms and diagnoses related to COVID-19. Random forest achieved the best performance with F1 score of 0.880, accuracy of 0.888 and sensitivity of 0.888.
Using the officially reported daily case counts as outcome, the authors found that post reporting symptoms and diagnosis of COVID-19 significantly predicted daily case counts up to 14 days ahead of official statistics. For the subset of geotagged posts, they found that the predictive pattern held true for both Hubei province and the rest of mainland China regardless of the unequal distribution of health care resources and the outbreak timeline.
Ordinary least squares regression with robust standard errors was used to estimate the final model.
Researchers and disease control agencies should pay close attention to the social media infosphere regarding COVID-19. In addition to monitoring overall search and posting activities, leveraging ML approaches and theoretical understanding of information sharing behaviors is a promising approach to identify true disease signals and improve the effectiveness of infoveillance.

A model combining symptoms to predict probable COVID-19 infection was applied to the data from all app users who reported symptoms to predict participants that are likely to have COVID-19. In *Menni et al., 2020* the authors investigated whether loss of smell and taste is specific to COVID-19 in 2,618,862 individuals who reported symptoms through the use of Symptom Tracker smartphone app afore mentioned [78]. Using a logistic regression-based model they combined symptoms that included loss of smell (anosmia) and taste, fatigue, persistent cough and loss of appetite to obtain a symptoms prediction model for COVID-19. All ten symptoms queried (fever, persistent cough, fatigue, shortness of breath, diarrhea, delirium, skipped meals, abdominal pain, chest pain and hoarse voice) were associated with testing positive for COVID-19 in the UK cohort, after adjusting for multiple testing; while in the US cohort, only loss of smell and taste, fatigue and skipped meals were associated with a positive test result.
In the UK test set the model achieved on average: AUC=0.76, sensitivity=0.65, specificity= 0.78, while in the US validation set the model achieved on average: AUC=0.76, sensitivity=0.66, specificity= 0.83. Among all the users who reported their symptoms on the app, 17.42% are likely to be infected.

In *Agosto et al., 2020* the authors presented a statistical model which can be employed to understand the contagion dynamics of the COVID-19 and can heavily im-

pact health, economics and finance. The model is a Poisson autoregression of the daily new observed cases, and can reveal whether contagion has a trend, and where is each country on that trend. In this way preventive measures (such as mobility restrictions) can be applied and/or relaxed. Model results are presented from the observed series of China, Iran, Italy and South Korea. All the estimated autoregressive coefficients of the model are statistically significant, confirming the presence of both a short-term dependence and a long-term trend.

Among patients with COVID-19, the ability to identify patients at risk for deterioration during their hospital stay is essential for effective patient allocation and management. To predict patient risk for critical COVID-19 based on status at admission it is possible using ML models.
Further refinement of these models with more data, would strengthen their predictive power and allow them to be a useful tool to identify early from the many with COVID-19, who will develop more serious disease and require closer clinical attention and resources.
Anyway, ML tools need to be developed iteratively and include clinicians in their development to be clinically applicable.

Chapter 3

Discussion

3.1 Strengths and Weaknesses of ML Tools

Every process or technique has some sort of pros and cons. Even in the case of ML, there are some factors that lead to advantages or disadvantages.
However, the fundamental theory behind ML-driven tools is that they require sufficient training data (of all possible cases).
Often, traditional ML requires a clean set of annotated data so that classifiers can possibly be well trained, which falls under scope of supervised learning.
Collecting large amounts of data is not trivial, and one has to wait for a long time. AI experts state the fact that limited data may skew results away from the severity of coronavirus outbreak not providing optimal performance.

The Wall Street Journal [122] reported that coronavirus reveals limits of ML health tools: some diagnostic-app makers are holding off updating their tools, highlighting the shortage of data on the new coronavirus and the limitations of health services billed as ML when faced with novel, fast-spreading illnesses.

According to *Santosh, 2020*, to detect COVID-19, ML-driven tools are expected to have active learning-based cross-population train/test models that employ multimodal data. This means that instead of having a conventional set of train, validation, and test set, there is the need of ML-driven tools that can learn over time without having complete knowledge about the data as in active learning (AL).
Beside the use of AL in ML, the author suggests that cross-population train/test models are the must in such scenarios, since we do not have enough data from the particular regions, as well the use of multimodal data can help support decision-making process with higher confidence.
ML models provide the advantage of learning non-linear relations between the input and the output, allowing inclusion of heterogeneous variable types in one model, which is not feasible in univariate analysis or risk-score prediction models.
In this context, publishers, journals and researchers are still urged to research different domains and stop the spread of this deadly virus.

3.2 SWOT Analysis

As aforesaid, the current ML methods have some internal and external disadvantages and limitations that are impeding their ultimate implementation in the clinical arena. As such, ML can be considered a portion of a business trying to be introduced in the health care market (*Agarwal et al., 2020*).
For this reason, this section analyzes strengths, weaknesses, opportunities, and threats (SWOT analysis) emerged from the papers discussed (Figure 3.1).

(a) Screening and Diagnosis

(b) Contact Tracing

(c) Drugs/Vaccines Development

(d) Prediction and Forecasting

Figure 3.1: SWOT Analysis

3.3 Performance Metrics

The performance of each individual ML algorithm applied in the research works aformentioned was assessed by different evaluation metrics. A brief description of them is given in Figure 3.2 for classification metrics and in Figure 3.3 for regression metrics.

Metrics	Definition	Formula
Accuracy	Accuracy is the summation of TP and TN divided by the total instance values of the confusion matrix.	$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$
AUC	It measures the capability of the model to distinguish between classes.	$\text{AUC} = \int_{x=0}^{1} TPR\left(FPR^{-1}(x)\right) dx$
F-Measure	The harmonic mean of precision and recall. It is also called F1-score .	$\text{F-Measure} = \frac{2TP}{2TP+FP+FN}$
Sensitivity	The ratio of true positive that are correctly identified. It is also called Recall.	$\text{Sensitivity} = \frac{TP}{TP+FN}$
Specificity	The ratio of true negative that are correctly identified.	$\text{Specificity} = \frac{TN}{TN+FP}$

Figure 3.2: Classification metrics

Metrics Definition	
$MSE = \frac{\sum_{i=0}^{N-1}(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{N}$	$\mathbf{y}_i$ is the actual expected output $\hat{\mathbf{y}}_i$ is the model's prediction.
$RMSE = \sqrt{MSE}$	
$MAE = \sum_{i=0}^{N-1} \lvert\mathbf{y}_i - \hat{\mathbf{y}}_i\rvert$	
$R^2 = 1 - \frac{MSE}{\text{VAR}(\mathbf{y})\cdot(N-1)} = 1 - \frac{\sum_{i=0}^{N-1}(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_{i=0}^{N-1}(\mathbf{y}_i - \bar{\mathbf{y}})^2}$	

Figure 3.3: Regression metrics

MAE (mean absolute error) represents the difference between the original and predicted values extracted by averaging the absolute difference over the data set.
MSE (mean squared error) represents the squared differences between the original and predicted values extracted by averaging these differences over the data set.
RMSE (root mean squared error) is the error rate, the square root of MSE.
R-squared (coefficient of determination) represents how well the values fit compared to the original values. The value ranges from 0 to 1 interpreted as percentages. Hence, the higher the value is, the better the model is.

3.4 Comparison

Details of the literature evaluation discussed in the previous chapter are summarized to compare techniques and obtained results in screening and diagnosis, contact tracing, drugs/vaccines development and prediction and forecasting COVID-19.
See respectively Tab 3.1, Tab 3.2, Tab 3.3, Tab 3.4 and Tab 3.5.

Table 3.1: State-of-the-art ML technology in COVID-19 Screening and Diagnosis - part 1

Ref	*Application nature*	*ML algorithm*	*Performance*
[21]	Detection of COVID-19 cases from chest X-Ray images	CNNs, LR, SVM, MLP, NB, DT, GNB, NC, BNB (Classification)	Accuracy: 0.94 Sensitivity: 0.94 Specificity: 0.97
[22]	Diagnose novel Coronavirus from X-ray images	Faster R–CNN (Classification)	Accuracy: 0.97 Sensitivity: 0.98 Specificity: 0.95
[23]	COVID-19 detection by SVM based on deep feature using X-ray images	ResNet50+SVM (Classification)	Accuracy: 0.95 Sensitivity: 0.97 Specificity: 0.93
[24]	Binary classifications with four classes: COVID-19, normal, viral and bacterial pneumonia	Deep CNN ResNet-50 (Classification)	(d1) Accuracy: 0.96 (d2) Accuracy: 0.99 (d3) Accuracy: 1.00
[26]	COVID-19 detections from X-ray binary and 3-class classification using 2 datasets	VGG-19 and MobileNet v2	Accuracy: 0.97 Sensitivity: 0.99 Specificity: 0.97
[27]	Pneumonia classification, COVID-19 diagnosis and localization of the main lesions from X-ray	DRE-Net (ResnNet50 + FPN) (Classification)	(task1) Accuracy: 0.86 Sensitivity: 0.96 (task2) Accuracy: 0.94 Sensitivity: 0.93
[29]	Weakly-supervised software system to detect COVID-19 from 3D CT volumes	UNet+DeCoVNet (3D DNN) (Classification)	Accuracy: 0.91 Sensitivity: 0.91 Specificity: 0.91
[30]	Distinguish COVID-19 pneumonia from Influenza-A viral pneumonia (IAVP)	ResNet-18 (Classification)	Accuracy: 0.87 Sensitivity: 0.87 Specificity: 0.81
[31]	Detect COVID-19 from chest X-ray images	Multi-CNN (Feature extraction) CFS (feature selection) BayesianNet (classification)	(d1) Accuracy: 0.91 Sensitivity: 0.98, AUC:0.96 (d2) Accuracy: 0.97 Sensitivity: 0.99, AUC:0.91
[32]	Rapid diagnosis for COVID-19 combined with clinical info	ResNet18 + MLP (Feauture extraction) + (Classification)	AUC:0.92, Sensitivity: 0.84 Specificity: 0.83
[33]	Probability maps to augment COVID-19 diagnosis from CT images	U-Net17 (Classification)	Accuracy: 0.81 Sensitivity: 0.83 Specificity: 0.73

Table 3.2: State-of-the-art ML technology in COVID-19 Screening and Diagnosis - part 2

Ref	*Application nature*	*ML algorithm*	*Performance*
[35]	Identify COVID-19, CAP non-pneumonia from chest CT	COVNet (Classification)	(COVID-19) Sensitivity: 0.90 AUC:0.96, Specificity:0.96 (CAP) Sensitivity: 0.87 AUC: 0.95, Specificity: 0.92
[36]	COVID-19 detection by using lung X-ray images	ResExLBP (Feature extraction) IRF (Feature selection) SVM (Classification)	Accuracy: 1.00 Sensitivity: 0.98 Specificity: 1.00
[37]	Early detection of COVID-19 cases using X-ray images	DarkCovidNet YOLO (Object detection)	(Binary) Accuracy: 0.98 Sensitivity: 0.95 Specificity: 0.92 (Multi-class) Accuracy: 0.87 Sensitivity: 0.85 Specificity: 0.92
[38]	Prediction of COVID-19 on CT images	M-Inception (Classification)	Accuracy :0.83 Sensitivity: 0.814 Specificity: 0.8
[39]	Automatically diagnose COVID-19 in X-ray images	COVIDX-Net (Classification)	Accuracy: 0.92 Sensitivity: 100 Specificity: 0.99
[40]	Detection of COVID-19 cases from chest X-ray images	COVID-Net (Classification)	Accuracy: 0.92 Sensitivity: 0.91
[42]	COVID-19 diagnosis	*"COVIDiagnosis-Net"* Deep Bayes-SqueezeNet	Accuracy: 0.98 Sensitivity: 0.98 Specificity: 0.99
[43]	Detect COVID-19 from chest X-ray images	CNN from scratch + SVM (Deep Feature Extraction) + (Classification)	Accuracy: 0.99 Sensitivity: 0.89 Specificity: 1.00
[44]	Classify COVID-19 CAP	FCNN Structured latent multi-view representation learning	Accuracy: 0.95 Sensitivity: 0.97 Specificity: 0.93
[46]	Rapid COVID-19 diagnosis: distinguish infection of COVID-19 from non COVID-19 groups	ResNet-101 (Classification)	Accuracy: 0.99 Sensitivity: 1.00 Specificity: 0.99
	Other ML medical systems		
[48]	InferVision, InferRead	Detect COVID-19 pneumonia	
[51]	Alibaba Damo Academy algorithm	Diagnose COVID-19 in 20"	

Table 3.3: State-of-the-art ML technology in COVID-19 Contact Tracing

Ref	*Application nature*	*ML algorithm*
[86]	Characterize tweets regarding self-reporting symptoms	Biterm topic model
[87]	App based auto-diagnosis of mild or severe COVID-19	Bayesian network
[88]	Quantify the effects of social strategies	Sampling algorithm + Bayesian optimization
[84]	Individual epidemic simulation using contact tracking app	Individual based model
	Other ML Solutions	
[78]	Symptom Study app UK	Real time disease tracking by self reporting
[82]	NHS COVID-19 app UK	Alert people of their risk of COVID-19 infection
[57]	Flu Near You	Disease surveillance web tool for flu
[58]	Covid Near You	Disease surveillance web tool for COVID-19 infection
[59]	Cancer Consortium	Report patient-level information for Cancer and COVID-19 infection web portal for researchers
[60]	Private Kit: Safe Paths	Free open-source and privacy-first contact-tracing technology
[61]	Immuni	COVID-19 Contat Tracing app in Italy
[62]	Radar COVID	COVID-19 Contat Tracing app in Spain
[63]	Corona-Warn	COVID-19 Contat Tracing app in Germany
[64]	StopCovid	COVID-19 Contat Tracing app in France
[65]	CovidSafe	COVID-19 Contat Tracing app in Australia
[66]	Trace Together	COVID-19 Contat Tracing app in Singapore
[67]	Covid Tracker	COVID-19 Contat Tracing app in Ireland
[68]	Cocoa	COVID-19 Contat Tracing app in Japan

Table 3.4: State-of-the-art ML technology in COVID-19 Drugs/Vaccines Development

Ref	*Application nature*	*ML algorithm*	*Performance*
[102]	Predict COVID-19 vaccine candidates.	XGB	F1: 0.95
[103]	Predict effective drug candidates against a specific coronavirus	Knowledge graph Text mining	-
[107]	Identify potential old drugs with anti-coronavirus activities	Deep neural network	-
[108]	Prediction of commercially available antiviral drugs for SARS-CoV-2	MT-DTI deep learning model	-
[109]	Classification of COVID-19 novel pathogens	ML allignement-free (Genome analysis) Linear discriminant, SVM, kNN (Classification)	Accuracy: 1.00
[110]	Identify potential SARS-CoV-2 epitopes	NetMHC (peptide prediction) + Fully connected Feed-forward NN	AUC: 0.97
	Other ML medical systems		
[104]	EVQLV	Rapid antibodies discovery	

CNN, as well as classification algorithms, including decision trees, SVM, MLP, kNN, logistic regressions, random forests and also Bayesian networks are the most used and versatile ML algorithms encountered in these papers.

Table 3.5: State-of-the-art ML technology in COVID-19 Prediction and Forecasting

Ref	*Application nature*	*ML algorithm*	*Performance*
[111]	Predict patient development into severe/critical COVID-19 symptoms	SVM (Classification)	AUC:0.98 Sensitivity: 1.00
[112]	Predict clinical severity (ARDS) and relative triggers	SVM (Classification)	Accuracy: 0.80
[113]	Mortality prediction model for COVID-19 patients	Multi-tree XGBoost	Accuracy: 0.90
[114]	Real-time forecasting model for COVID-19 transmission	LSTM network	RMSE: 45.70 Accuracy:0.93
[115]	Real-time forecasting and risk assessment for COVID-19	Hybrid ARIMA-WBF Regression trees	(RT) R-squared:0.90 (RT) R-adj:0.77
[116]	Predict COVID-19 infection from clinical available blood test results	Random forest (Classification)	AUC: 0.99, Sensitivity: 1.00 Specificity: 0.94
[117]	Predict the risk for critical COVID-19	Random forest (Classification)	Accuracy: 0.92 Sensitivity: 0.88 Specificity: 0.93
[118]	Short-term forecasting of cumulative cases of COVID-19	Stacking-ensemble learning	(1 day) Err: 0.87-3.51 (3 days) Err: 1.02-5.63 (6 days) Err: 0.95- 6.90
[119]	Predict COVID-19 case counts in China from Weibo posts	Random forest Regression trees	Accuracy: 0.89 Sensitivity: 0.89 F1:0.89
[120]	Investigate COVID-19 new symptoms from smartphone-based app reporting	Logistic regression (Classification)	AUC:0.76 Sensitivity: 0.66 Specificity: 0.83
[121]	COVID-19 contagion trend	Poisson autoregression model	-

Chapter 4

Conclusions

This state of the art provided a quite complete overview of the research works based on ML and deep learning algorithms applied to fight the COVID-19 disease, including distinct information, such as application nature, ML techniques used and performance obtained for each study.
Some of the exposed models are not deployed enough to show their real world operation. They require to be tested with more data, that's why it is so important to share the knowledge and results, to develop open-source solutions that allow more people to work on and improve the new proposals quickly.
The shortage of studies in the literature that are really applicable is a great concern and may have serious implications for detecting and minimising the spread of this virus.
Numerous challenges and research limitations have been indicated in the academic literature and need to be addressed in the future.
Nevertheless, the ongoing development in ML has already significantly improved screening and diagnosis, contact tracing, drugs/vaccines development process and prediction and forecasting for the COVID-19 pandemic reducing the human intervention in medical practice:

(2.2) *Screening and Diagnosis*
Deep learning based technology proved to be crucial to provide correct and rapid diagnosis using chest scan, such that recent digital solutions reached the same or even more accuracy than a senior radiologist into early distinguish different types of pneumonia from COVID-19 pneumonia.
In this context, ML algorithms used in diagnostic operations primarily rely on CNN and computer vision techniques able to manipulate and extract insight from image analysis. Based on the results, it is demonstrated that deep learning with CNNs may have significant effects on the automatic detection and automatic extraction of essential features from X-ray images, related to the diagnosis of the COVID-19.
Analyzing X-rays image and CT scans with deep learning algorithms together with computer vision technology, made possible achieve amazing results.
Since by now, all diagnostic tests show failure rates such as to raise concerns, the probability of incorporating X-rays based DL models into the diagnosis of the disease could be assessed by the medical community, thanks to the findings achieved. In addition, DNNs are not only a valuable support in the activity of radiologists but they also represent a double check for false

RT-CPR responses.

(2.3) *Contact tracing*
Mobile apps are still not diffused enough but are helping us to contain the spread of the virus. Public social media data can be usefully harnessed to predict infection cases and inform timely responses.
Furthermore, knowing the possible development of the contagion dynamics lets governments plan more efficient and less-invasive containment strategies. Having situational awareness by governments would benefit from increased access to previously unavailable population estimates and mobility information to enable different sectors to better understand COVID-19 trends and geographic distribution.
Prediction tasks would leverage real-time population counts and mobility data, enabling new predictive capabilities to assess future risks, needs, opportunities as well the identification of the key drivers and consequences of implementing different measures to contain the spread of COVID-19.
This could be a big improvement, since the lack of available data is a big limitation for ML solutions implementation in real world.
In addition, using mobile apps for contact tracing helps to make people more aware about the consequences of their contacts and their movements, allowing to contain the spread of the virus and individual protection.
Nevertheless, a key concern about mobile data is that the pandemic is used to create and legitimize surveillance tools used by government and technology companies that are likely to persist beyond the emergency, threatening individual privacy.

(2.4) *Drugs/Vaccines development*
ML is helping to discover which treatments can be used to cure infected patients meanwhile a vaccine is found and in parallel, their application is helping us also saving time significantly with automated research of the antibodies to develop the vaccine itself.
Studies show that it takes an average of five years and a half billion dollars to discover and optimize antibodies in a laboratory: the contribution of ML models in this contest is huge. Hence, ML helps in developing vaccines and treatments at much of faster rate than usual and is also helpful for clinical trials during the development of the vaccine. ML algorithms are speeding up the process of discovering potential vaccine candidates and identifying commercially available drugs with anti SARS-CoV-2 activities for the treatment of COVID-19 patients.

(2.5) *Prediction and Forecasting*
In its peak, the COVID-19 emergency department inquiries went beyond the ICU capacity, forcing researchers to seek for the best tool identifying the high-risk patients. ML models provide new tools to utilize during emergency department triage to anticipate disease progression for the best placement on the one hand, and perhaps a more aggressive treatment regimen on the other hand.
Hence, artificial intelligence may be applied for accurate risk prediction of patients with COVID-19, to optimize patients triage and in-hospital alloca-

tion, better prioritization of medical resources and improved overall management of the COVID-19 pandemic.
Knowing in advance who, among cases, will develop severe symptoms allows to manage better the medical resources and treatments, as well as to avoid ICU stress.
Furthermore, knowing the possible development of the contagion dynamics and case counts lets governments plan more efficient and less-invasive containment strategies.

On top of that, exploring the previous solutions adopted to fight the other coronaviruses lets us know better this invisible enemy.
In conclusion, it is evident that ML can significantly improve screening and diagnosis, contact tracing, drugs/vaccines development and prediction and forecasting for the COVID-19 pandemic and reduce human intervention in medical practice.

List of Figures

List of Tables

Bibliography

[1] Johns Hopkins University (JHU). *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE), (October 2020).*

[2] Coronavirus. *WHO: World Health Organization (2020).*

[3] Semantic Scholar team - COVID-19 Open research dataset (CORD-19). *Allen Institute for AI (2020).*

[4] *ELLIS Society (European Lab for Learning and Intelligent Systems)*

[5] AI and control of Covid-19 coronavirus - Overview by ad hoc committee on artificial intelligence (CAHAI) secretariat. *Council of Europe Portal (2020)*

[6] Cobey S. - Modeling infectious disease dynamics. *Science, 15 May 2020:Vol. 368, Issue 6492, pp. 713-714 (2020).*

[7] Davenport, T., & Kalakota, R. - The potential for artificial intelligence in healthcare. *Future Healthcare Journal, 6(2), 94–98. (2019).*

[8] Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. - Artificial intelligence and machine learning to fight COVID-19. *Physiological Genomics, 52(4), 200–202, (2020).*

[9] Yu, K., Beam, A.L. & Kohane, I.S. - Artificial intelligence in healthcare. *Nature Biomedical Engineering 2, 719–731 (2018).*

[10] Noguerol T., Paulano-Godino, Félix Martín-Valdivia M., Me- nias, C., Luna, A. - Strengths, weaknesses, opportunities, and threats analysis of artificial intelligence and machine learning. *Applications in Radiology. Journal of the American College of Radiology. 16. 1239-1247 (2019).*

[11] Albahri A.S., Hamid R.A., Alwan J.k. et al. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel Coronavirus (COVID-19): a systematic review. *Journal of Medical Systems 44, 122 (2020).*

[12] Petrosillo N., Viceconte G., Ergonul O., Ippolito G., Petersen E. - COVID-19, SARS and MERS: are they closely related? *Clinical Microbiology and Infection, volume 26, Issue 6, Pages 729-734 (2020).*

[13] Sandhu R., Sood, S.K., Kaur, G. An intelligent system for predicting and preventing MERS-CoV infection outbreak. *The Journal of Supercomputing volume 72, pages 3033–3056 (2016).*

Screening and Diagnosis section relative papers

[14] Ai T., Yang Z., Hou H., Zhan C., Chen C., Lv W., Tao Q., Sun Z., Xia L. - Correlation of chest CT and RT-PCR testing for Coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology, 296(2), E32–E40 (2020).*

[15] Liu R., Han H., Liu F., Lv Z., Wu K., Liu Y., et al. Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clinica Chimica Acta Volume 505, Pages 172-175 (2020)*

[16] Li D., Wang D., Dong J., Wang N., Huan H., Xu H., Xia C. - False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome Coronavirus 2: role of deep-learning-based CT diagnosis and insights from two cases. *Korean Journal of Radiology, 21(4), 505–508 (2020).*

[17] Chung M. et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology 295, 202–207 (2020).*

[18] Pasa F., Golkov V. , Pfeiffer F., Cremers D., and Pfeiffer D., - Efficient deep network architectures for fast chest X-Ray tuberculosis screening and visualization. *Scientific Reports, vol. 9, no. 1, p. 6268 (2019).*

[19] Miki Y. et al. - Classification of teeth in cone-beam CT using deep convolutional neural network. *Computers in Biology and Medicine, vol. 80, pp. 24-29 (2017).*

[20] Selvikvåg Lundervold A., Lundervold A. - An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik, volume 29, Issue 2, Pages 102-127, ISSN 0939-3889 (2019).*

[21] Ahammed K., Md. S. Satu, M. Z. Abedin, Md. A. Rahaman, S. M. S. Islam - Early detection of Coronavirus cases using chest X-ray images employing machine learning and deep learning approaches. *medRxiv 2020.06.07.20124594 (2020).*

[22] Shibly K.H., Dey S.K., Islam M.A., Rahman M. (2020). COVID faster R–CNN: A novel framework to diagnose novel Coronavirus disease (COVID-19) in X-ray images. *Informatics in Medicine Unlocked, 20, 100405 - 100405 (2020).*

[23] Sethy, P.K., Behera, S.K. - Detection of Coronavirus disease (COVID-19) based on deep features. *Preprints 2020, (2020).*

[24] Narin A., Kaya C., Pamuk Z. (2020). Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and deep convolutional neural networks. *ArXiv, 2003.10849 (2020).*

[25] Tan C. et al. - A survey on deep transfer learning. *ArXiv, 1808.01974 (2018)*

[26] Ioannis D. A., Tzani B. - COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine 43:635–640 (2020)*

[27] Song Y., Zheng S., Li L., Zhang X., Zhang X., Huang Z., Chen J., Zhao H., Jie Y., Wang R., Chong Y., Shen J., Zha Y., Yang Y.. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *medRxiv 2020.02.23.20026930 (2020).*

[28] Song Y., et al. *Discriminating COVID-19 Pneumonia from CT Images Server, 2019*

[29] Zheng C., Deng X., Fu Q., Zhou Q., Feng J., Ma H., Liu W., Wang X. - Deep learning-based detection for COVID-19 from chest CT using weak label. *medRxiv 2020.03.12.20027185 (2020).*

[30] Xu X., Jiang X., Ma C., Du P., Li X., Lv S., Yu L., Ni Q., Chen Y., Su Y., Lang G., Li Y., Zhao H., Liu J., Xu K., Ruan L., Sheng J., Qiu Y., Wu W., Liang T., Li L. - A Deep learning system to screen novel Coronavirus disease 2019 pneumonia. *Engineering, ISSN 2095-8099 (2020).*

[31] Abraham B., Nair M. S. - Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier. *Biocybernetics and Biomedical Engineering, vol. 40(4), pages 1436–1445 (2020).*

[32] Mei X., Lee H., Diao K. et al. - Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. *Nature Medicine 26, 1224–1228 (2020).*

[33] Hurt B., Yen A., Kligerman S., Hsiao A. - Augmenting Interpretation of chest radiographs with deep learning probability maps. *Journal of Thoracic Imaging, 35(5), 285–293 (2020).*

[34] Hurt B., Kligerman S., Hsiao A. - Deep learning localization of pneumonia: 2019 Coronavirus (COVID-19) outbreak. *Journal of Thoracic Imaging, 35(3), W87–W89 (2020).*

[35] Li L., et al. - Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology, vol. 296, No. 2 (2020).*

[36] Tuncer T., Dogan S., Ozyurt F. - An automated residual exemplar local binary pattern and iterative relief based COVID-19 detection method using chest X-ray image. *CChemometrics and Intelligent Laboratory Systems, vol. 203, 104054 (2020).*

[37] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra A. U. Automated detection of COVID-19 cases using deep neural networks with X- ray images. *Computers in Biology and Medicine, volume 121 (2020).*

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

[39] Ezz E.D.H., Shouman M. A., Karar M. E. - COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images. *arXiv:2003.11055v1 (2020).*

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

[41] Wang L. - COVID-Net open source initiative open access. *Github (2020)*

[42] Ucar F., Korkmaz D. (2020). COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Medical Hypotheses, 140, 109761 (2020).*

[43] Nour M., Cömert Z., Polat K. - A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. *Applied Soft Computing,106580 (2020).*

[44] Kang H. et al. - Diagnosis of Coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning. *IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2606-2614 (2020).*

[45] Li Y., Yang M. and Zhang Z. - A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 10, pp. 1863-1883 (2019)*

[46] Ardakani A., Kanafi R., Acharya U.,Khadem N., Mohammadi A. - Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. *Computers in Biology and Medicine, 121, 103795 (2020).*

[47] Harrison X. Bai, Hsieh B., Xiong Z., Halsey K., Choi L.W., Thi My Linh Tran, Pan I., Shi L.B., Wang D.C., Mei J., Jiang X.L., Zeng Q.A., Egglin T., Ping-Feng Hu, Agarwal S., Xie F.F., Li S., Healey T., Atalay M.K., and Wei-Hua Liao - Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. *Radiology, 296:2, E46-E54 (2020).*

[48] Beijing Infervision Technology Co., Ltd. global high-tech enterprise in medical artificial intelligence. *InferVision (2020).*

[49] Alibaba DAMO Academy. *DAMO Website (2020)*

[50] AI algorithm detects coronavirus infections in patients from CT scans with 96% accuracy. *Technology.org (2020)*

[51] So D. - Alibaba news roundup: tech takes on the outbreak *Alizila News from Alibaba Group (february 21, 2020).*

Contact Tracing section relative papers

[52] Contact tracing in the context of COVID-19. *WHO Headquarters (HQ) - ref: WHO/2019-nCoV/Contact_Tracing/2020.1 (May 10, 2020).*

[53] Lalmuanawma S., Hussain J., Chhakchhuak L. - Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos, Solitons & Fractals, vol. 139, 110059 (2020).*

[54] Oliver N. and Letouzé E. and Sterly H. and Delataille S. and De Nadai M. and Lepri B.and Lambiotte R. and Benjamins R. and Cattuto C.and Colizza V. and de Cordes N. and Fraiberger S.P. and Koebe T. and Lehmann S. and Murillo J. and Pentland A. and Phuong N Pham and Pivetta F. and Ali SalahA. and Saramäki J. and Scarpino S.V. and Tizzoni M. and Verhulst S. and Vinck P. - Mobile phone data and COVID-19: Missing an opportunity? *arXiv:2003.12347 (2020).*

[55] Pandemic Intervals Framework (PIF). *Centers for Disease Control and Prevention (2020)*

[56] Koo J.R., Cook A.R., Park M., Sun Y., Sun H., Tao Lim J., Tam C. and Dickens B. - Interventions to mitigate early spread of SARS-CoV-2 in Singapore. *Lacet Infectous Disease (2020).*

[57] *Flu Near You web based tool*

[58] *Covid Near You web based tool*

[59] *The Covid-19 & Cancer Consortium web portal*

[60] Barbar R., Beaudry R., Benedetti F.M., Clough A., Das R., Gupta R., Jain K., Kanaparti R., Kanaparti S., Kapa S., Keegan C., Louisy K., Nadeau G., Nuzzo A., Penrod S., Rajaee Y., Raskar R., Schunemann I., Singh A., Storm G., Vepakomma P., Vilcans K., and Werner J. - Apps gone rogue: maintaining personal privacy in an epidemic. *Massachusetts Institute of Technology (2020).*

[61] *Immuni App - Italy, 2020.*

[62] *Radar COVID App- Spain, 2020.*

[63] *Corona Warn App - Germany, 2020.*

[64] *StopCovid App - France, 2020.*

[65] *Covidsafe App - Australia, 2020.*

[66] *TraceTogether App - Singapore, 2020.*

[67] *Covid Tracker App - Ireland, 2020.*

[68] *Cocoa App - Japan, 2020.*

[69] Drew D. A., Nguyen L. H., Steves C. J., Menni C., Freydin M., Varsavsky T., Sudre C. H., Cardoso M. J., Ourselin S., Wolf J., Spector T. D., Chan A. T., COPE Consortium - Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science, vol. 368 (6497), 1362–1367 (2020).*

[70] Mobile applications in support of contact tracing for COVID-19 - A guidance for EU EEA Member States. *European Centre for Disease Prevention and Control. Mobile applications in support of contact tracing for COVID-19 (2020).*

[71] *European Data Protection Supervisor.*

[72] *European Data Protection Board.*

[73] Coronavirus: guidance to ensure full data protection standards of apps fighting the pandemic. *European Commision Website (2020).*

[74] EU data protection rules. *European Commission Website (2020).*

[75] Privacy-preserving contact tracing *Apple Website (2020)*

[76] Apple and Google partner on COVID-19 contact tracing technology. *Apple Website (2020)*

[77] *Apple and Google partner on COVID-19 contact tracing technology (2020).*

[78] *COVID Symptom Study app*

[79] About the COVID Symptom Study App - COronavirus Pandemic Epidemiology (COPE) Consortium. *Mongan Istitute Website (2020)*

[80] Ferretti L., Wymant C., Kendall M., Zhao L., Nurtay A., Abeler-Dörner L., Parker M., Bonsall D., Fraser C. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science, vol. 368 (6491), eabb6936 (2020)*

[81] Digital contact tracing for SARS-COV-2 (2020) *Shiny Web Application*

[82] *NHS COVID-19 app*

[83] Oxford University Research Team Support launch of the NHS COVID-19 contact tracing app. *https://Coronavirus-fraser-group.org (2020).*

[84] Hinch R., Probert W., Nurtay A., Kendall M., Wymant C., Hall M. Fraser C. - Effective configurations of a digital contact tracing app: a report to NHSX. *Github (2020).*

[85] Nepomuceno E.G., Resende D.F., Lacerda M. A Survey of the individual- based model applied in biomedical and epidemiology. *Journal of Biomedical Research and Reviews, vol. 1, no. 1, pp. 11-24 (2018)*

[86] Mackey T., Purushothaman V., Li J., Shah N., Nali M., Bardier C., Liang B., Cai M., Cuomo R. - Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: retrospective big data infoveillance study. *JMIR Public Health and Surveillance, vol. 6 (2), (2020).*

[87] McLachlan S., Lucas P., Dube K., McLachlan G. S., Hitman G. A., Osman M., Kyrimi E, Neil M, Fenton N.- The fundamental limitations of COVID- 19 contact tracing methods and how to resolve them with a Bayesian network approach *ResearchGate (2020)*

[88] Lorch L. and Kremer H. and Trouleau W. and Tsirtsis S. and Szanto A. and Schölkopf B. and Rodriguez M.G. - Quantifying the effects of contact tracing, testing, and containment. *arXiv:2004.07641 (2020)*

[89] Lorch L. Simulator for the spatiotemporal model for Covid-19 *Github (2020)*

[90] COVID-19 tracing apps: ensuring privacy and data protection. *European Parliament Website (2020).*

Drugs/Vaccines Development section relative papers

[91] Fontana J. M., Alexander E., Salvatore M. - Translational research in infectious disease: current paradigms and challenges ahead. *Translational Research vol. 159(6), pages 430–453 (2012).*

[92] Marshall M. - We're beginning to understand the biology oof the covid-19 virus. *NewScientist (2020)*

[93] Maxwell K. L., Frappier L. Viral proteomics. *Microbiology and Molecular Biology Reviews 71(2), 398–411 (2007).*

[94] Wohl S., Schaffner S. F., Sabeti P. C. - Genomic analysis of viral outbreaks. *Annual Review of Virology, vol. 3(1), 173–195 (2016).*

[95] Tyers M., Mann M. - From genomics to proteomics. *Nature 422, 193–197 (2003).*

[96] University of Bristol - Genome analysis has important implications for COVID-19 clinical trials. *Medical Express (2020).*

[97] Anuj S., Radha K.M. K. - Transcriptome analysis. *Encyclopedia of Bioinformatics and Computational Biology, volume 3, Pages 792-8052019 (2019).*

[98] Du L., Yang Y., Zhou Y., Lu L., Li F., Jiang S.. MERS-CoV spike protein: a key target for antivirals. *Expert Opinion on Therapeutic Targets, vol. 21(2), 131-143 (2016).*

[99] Du L., He Y., Zhou Y. et al. - The spike protein of SARS-CoV - a target for vaccine and therapeutic development. *Nature Reviews Microbiology 7, 226–236 (2009).*

[100] Zhang B., Hu Y., Chen L. et al. - Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell Research 30, 702–704 (2020).*

[101] Feldkamp C.S. , Carey. J. L. - 2 - Immune function and antibody structure. *Immunoassay pages 5-24 (1996).*

[102] Ong E., Wong M. U., Huffman A., He Y. - COVID-19 coron- avirus vaccine design using reverse vaccinology and machine learning. *bioRxiv 2020.03.20.000141 (2020).*

[103] Ge Y., Tian T., Huang S., Wan F., Li J., Li S., Yang H., Hong L., Wu N., Yuan E.,Cheng L., Lei Y., Shu H., Feng X., Jiang Z., Chi Y., Guo X., Cui L., Xiao L., Li Z., Yang C., Miao Z., Tang H., Chen L., Zeng H., Zhao D., Zhu F., Shen X., Zeng J. - A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *bioRxiv 2020.03.11.986836 (2020)*

[104] *EVQLV AI-powered technology company.*

[105] Data science institute alumni use machine learning to discover Coronavirus treatments. *Columbia University (2020*

[106] *Immunoprecise Antibodies Ltd. (IPA).*

[107] Ke Y-Y, Peng T-T, Yeh T-K, Huang W-Z, Chang S-E, Wu S-H, Hung H-C, Hsu T-A, Lee S-J, Song J-S, Lin W-H, Chiang T-J, Lin J-H, Sytwu H-K, Chen C-T. Artificial intelligence approach fighting COVID-19 with repurposing drugs. *Biomed Journal, volume 43, issue 4, August 2020, Pages 355-362 (2020).*

[108] Beck B.R., Shin B., Choi Y., Park S., Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computational and Structural Biotechnology Journal vol.18 pages 784–90 (2020).*

[109] Randhawa G.S., Soltysiak M.P.M., El Roz H., de Souza C.P.E., Hill K.A., Kari L. (2020) Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PLoS ONE 15(4): e0232391 (2020).*

[110] Prachar M., Justesen S., Steen-Jensen D., Thorgrimsen S., Jurgons E., Winther O., Bagger F.O.- COVID-19 vaccine candidates: prediction and validation of 174 SARS-CoV-2 epitopes. *bioRxiv 2020.03.20.000794 (2020).*

Prediction and Forecasting section relative papers

[111] Sun L., Liu G., Song F., Shi N., Liu F., Li S., Li P., Zhang W., Jiang X., Zhang Y., Sun L., Chen X., Shi Y. Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. *Journal of Clinical Virology, volume 128, 104431 (2020).*

[112] Jiang X., Coffee M., Bari A., Wang J., Jiang X. et al. - Towards an artificial intelligence framework for data-driven prediction of Coronavirus clinical severity. *CMC-Computers, Materials & Continua, vol.63(1), 537–551 (2020).*

[113] Yan L., Zhang H., Goncalves, J. et al. - An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence 2, 283–288 (2020).*

[114] Chimmula V.K.R., Zhang L. - Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals vol.135, Article 109864 (2020).*

[115] Chakraborty T., Ghosh I. - Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. *Chaos, Solitons & Fractals vol.135, article 109850 (2020).*

[116] Wu J., Zhang P., Zhang L., Meng W., Li J., Tong C., Li Y., Cai J., Yang Z., JZhu J., Zhao M., Huang H., Xie X. and Li S. - Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. *medRxiv 2020.04.02.20051136 (2020).*

[117] Assaf D., Gutman Y., Neuman Y., Segal G., Amit S., Gefen-Halevi S., Shilo N., Epstein A., Mor-Cohen R., Biber A., Rahav G., Levy I., Tirosh, A. - Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Internal and Emergency Medicine, 1–9 (2020).*

[118] Dal Molin Ribeiro M. H., da Silva R. G., Cocco Mariani V., dos Santos Coelho L. - Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil. *Chaos, Solitons & Fractals, volume 135, 109853, ISSN 0960-0779 (2020).*

[119] Shen C., Chen A., Luo C., Zhang J., Feng B., Liao W. (2020). Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in Mainland China: observational infoveillance study. *Journal of Medical Internet Research, vol. 22(5), e19421.*

[120] Menni C., Valdes A.M., Freidin M.B. et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine 26, 1037–1040 (2020).*

[121] Agosto A., Giudici P. - A Poisson autoregressive model to understand COVID-19 contagion dynamics. *DEM Working Papers Series 185, University of Pavia, Department of Economics and Management (2020).*

[122] Olson P. - Coronavirus reveals limits of AI health tools. *The Wall Street Journal (2020).*

[123] Santosh K.C. AI-Driven tools for Coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. *Journal of Medical Systems 44, 93 (2020).*

[124] Agarwal S., Punn N. S., Sonbhadra S. K., Nagabhushan P., Pandian K. K., Saxena P. (2020). Unleashing the power of disruptive and emerging technologies amid COVID 2019: a detailed review. *arXiv preprint arXiv:2005.11507 (2020).*