medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

A Machine Learning Solution Framework for Combatting COVID-19 in Smart Cities from Multiple Dimensions

Ibrahim Abaker Targio Hashem[1], Absalom E. Ezugwu[2,*], Mohammed A. Al-Garadi[3], Idris N. Abdullahi[4], Olumuyiwa Otegbeye[2], Queeneth O. Ahman[5], Godwin C. E. Mbah[5], Amit K. Shukla[6], Haruna Chiroma[1,*]

[1]Future Technology Research Center, National Yunlin University of Science and Technology, Taiwan

[2]School of Computer Science, University of KwaZulu-Natal, Pietermaritzburg, KwaZulu-Natal South Africa

[3]Department of Biomedical Informatics, Emory University, Atlanta, USA

[4]Department of Medical Laboratory Science, College of Medical Sciences, Ahmadu Bello University, Zaria, Nigeria

[5]Department of Mathematics, Faculty of Physical Sciences, University of Nigeria, Nsukka, Nigeria

[6]IRISA Laboratory, ENSSAT, University of Rennes 1, France

[*]**Corresponding Authors**: Haruna Chiroma, Future Technology Research Center, National Yunlin University of Science and Technology, Yunlin, Taiwan; chiromaharun@fcetgombe.edu.ng. Absalom E. Ezugw, School of Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa; ezugwua@ukzn.ac.za.

Abstract. The spread of COVID-19 across the world continue as efforts are been made from multi-dimension to curtail it is spread and provide treatment. The COVID-19 triggered partial and full

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

lockdown across the globe in an effort to prevent it is spread. The COVID-19 causes serious fatalities with United States of America recording over 3,000 deaths within 24 hours, the highest in the world for a single day. In this paper, we propose a framework integrated with machine learning to curtail the spread of COVID-19 in smart cities. A novel mathematical model is created to show the spread of the COVID-19 in smart cities. The propose solution framework can generate, capture, store and analyze data using machine learning algorithms to detect, prevent the spread of COVID-19, forecast next epidemic, effective contact tracing, diagnose cases, monitor COVID-19 patient, COVID-19 vaccine development, track potential COVID-19 patients, COVID-19 drug discovery and provide better understand of the virus in smart cities. The study outlined case studies on the application of machine learning to help in the fight against COVID-19 in hospitals in smart cities across the world. The framework can provide guide for real world execution in smart cities. The propose framework has the potential for helping national healthcare systems in curtailing the COVID-19 pandemic in smart cities.

Keywords: COVID-19; Drugs discovery; Machine Learning Algorithms; Novel Coronavirus; Smart Cities.

1. Introduction

For over 4 months, the novel coronavirus that was later named Severe Acute Respiratory Syndrome Coronavirus 2 (SARSCoV-2) has caused unprecedented Coronavirus Disease 2019 (COVID-19) all over the world. The first human to human transmission of COVID-19 was first reported to World Health Organization (WHO) on 30th December, 2019. Thereafter, several retrospective study revealed that many COVID-19 patients started showing pneumonia symptoms

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

in early December. Even though, there are scientific controversies and theories over the date and origin of the SARS-CoV-2, it is widely proven that the novel coronavirus originated from the Wuhan live animal market (Hao, Zhong, Song, Fan, & Li, 2020).Genomic sequences of the early isolates of SARS-CoV-2 from infected patients in Wuhan showed over 88% nucleotide homology with two bad like SARS coronaviruses. Hence, strongly pointing towards zoonotic source and indeed bats serve as reservoir host of the SARS-CoV-2 (Hao et al., 2020). Currently, there are ongoing search for the possible intermediate host, which might have aided the transmission of the virus to human. The SARS-CoV-2 are droplet borne pathogen which gets in contact with humans when they are exposed to oral or nasal secretions of clinically symptomatically or asymptomatic infected persons (ECDC, 2020).

SARS-CoV-2 has tropism to cells and tissues that express the angiotensin converting enzyme 2. These receptors are mainly found in the respiratory tracts and limited extent the kidney, heart and gastrointestinal tract. The virus dock onto the receptor by means of receptor binding domain (RBD) of its spike glycoprotein. This represent the first step of viral replication and pathogenesis (Rothan & Byrareddy, 2020). As the virus replicates in the respiratory tract, it's provokes respiratory symptoms which mainly include dry cough, difficulty in breathing and sore throat. Then, it disseminates through the blood to other tissues and organs causing viremia and high fever. Hence, these symptoms together with others such as body weakness and pains represent the cardinal clinical symptoms of COVID-19 (Lin, Lu, Cao, & Li, 2020).

Majority of the SARS-CoV-2 infected persons remain asymptomatic and infection being self-limiting. However, some 5% of infected persons suffer severe COVID-19 (Rothan & Byrareddy, 2020). Major determining factor for severe and fatal COVID-19 include old age (>60 years), underlying cardiovascular, immunological, metabolic and respiratory comorbidities. Based on

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

available scientific reports, the transmission of SARS-CoV-2 revolves around human, animals and the environment (Chakraborty & Maity, 2020). For now, preserving human life and health security is the major concern of most countries and territories. Hence, prompted the legislation, implementation and enforcement of adequate infection prevention and control measures for high priority pathogen (WHO, 2019). Therefore, combatting the COVID-19 need better understanding.

For better understanding of the COVID-19 pattern of spread, accurate and speed of diagnoses, development of new therapeutic methods, identification of the most susceptible people according their distinct genetic and physiological characteristics, machine learning algorithms are required to analyze the COVID-19 large scale datasets (Alimadadi et al., 2020).As a result of that, attempts have been made by scholars to apply machine learning algorithms in combating the COVID-19 in from different perspective. For example, drugs discovery targeting COVID-19is proposed in (Ge et al., 2020). Machine learning approach in CRISPR based COVID-19 surveillance using genome is proposed in (Metsky, Freije, Kosoko-Thoroddsen, Sabeti, & Myhrvold, 2020).The classification of novel pathogens for the COVID-19 is presented in (Randhawa et al., 2020). Automated deep learning COVID-19 patient detection and monitoring system is reported in (Gozes et al., 2020; Oyelade and Ezugwu, 2020). Machine learning approach is applied to develop system for creating awareness about wash to contain the COVID-19 (Pandey, Gautam, Bhagat, & Sethi, 2020).Generative network is used for the design of COVID-19 3C-like protease inhibitors (Zhavoronkov et al., 2020).The survival of severe COVID-19 patients was predicted based on machine learning approach by (Yan et al., 2020). Deep learning based system for quantifying the volume of lungs infection is reported in (Yan et al., 2020).

However, the major issue with the previous studies is that each of the study focuses on a particular aspect of combatting the COVID-19 pandemic leaving other critical aspect that is required to fight

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

the COVID-19 pandemic. (Spectrum, 2020) reported that hospital emergency rooms across the globe are experiencing floods of people infected with COVID-19 for urgent treatment. As a result of that, unprecedented influx of COVID-19 patient into the emergency rooms is experience, doctors are grabbling with the problem of patient's triage. Therefore, struggling to decide the COVID-19 patient that require intensive care. The condition of the patient lungs has to be assessed by doctors and nurses. However, volunteer doctors and nurses without pulmonary training can't assess the patient lungs. At the peak of the COVID-19 crises in Italy, doctors were face with a serious problem of taken decision on the patient that should be given resources and assistance. In view of these difficult COVID-19 cases facing doctors and nurses, a machine learning system can help to provide clinical decision support to play critical role in the COVID-19 crises, assist hospitals to be fully functional and keep COVID-19 patients alive.

A framework with multiple dimension to combat the COVID-19 in different front can provide better automated solutions to the fight against COVID-19. A COVID-19 pandemic involve a lot of measures to combat it via machine learning such as predictingCOVID-19 vaccine immunogenicity, detecting COVID-19 severity, predictingCOVID-19 mortality, COVID-19 resource allocation, COVID-19 drugs discovery, COVID-19 contact tracing, social distance, detecting wearing of mask, detectingCOVID-19 patients requiring ventilator, predictingCOVID-19 patient that is beyond medical intervention and triage COVID-19 patients. All these can be integrated into a single framework to work automatically within a city. Therefore, smart city can be repurposed to combat the COVID-19 by applying framework with multiple measures to fight the COVID-19. To provide smart applications in the smart cities, many technologies are connected together including sensor networks, broadband communications service, sensor devices, wireless sensor networks, internet, and cloud services. Sullivan (2008) pointed out that smart cities in 2020

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

will be embedded with smart structures such as the smart healthcare, smart security, smart mobility, smart building, smart governance, smart citizens, smart infrastructure, smart technology, smart energy and smart education.

To the best of our knowledge, no dedicated framework integrated with machine learning to fight COVID-19 from multiple dimension automatically in smart cities reported in the literature. The framework can address the unique challenges in fighting COVID-19 to ease the work of healthcare workers in saving life's and provide guide for real world execution in smart cities. (Yu, Wang, Liu, & Zomaya, 2016) pointed out that exploring theory is critical for providing guide for effective applications.

In this paper, we propose a solution framework integrated with machine learning to combat COVID-19 in Smart Cities from multiple dimensions.

The other section of the paper are organized as follows: Section 2 presents the background information about COVID-19, Section 3 presents the basic concept of smart city including case studies and machine learning, Section 4 presents mathematical model of COVID-19 spread in smart city, Section 5 presents the applications of machine learning in fighting COVID-19 pandemic in smart cities, Section 6 outlined case studies on combatting COVID-19 via machine learning in smart city, Section 7 presents the propose solution framework for combatting COVID-19 in smart city from multiple front, before the concluding remarks in Section 8.

The summaries of the contributions of the study are provided as follows:

- A detail solution framework integrated with machine learning to fight COVID-19 in smart cities with guide for effective real world execution is proposed.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

- Mathematical model of the COVID-19 spread in smart cities is presented for easy understanding.
- Use cases of machine learning in combatting COVID-19 in smart cities in presented in the study.
- As a novel virus, the study outlined cardinal clinical and laboratory features of the COVID-19 for different stages to properly guide the machine learning community in developing healthcare decision support related to COVID-19 to be embedded in smart healthcare.

2. The Novel Coronavirus Diseases: background and cardinal clinical features

Out of the seven Coronaviruses, SARS-CoV-2 is the third highly pathogenic coronaviruses that has afflicted human race. As the incidence and fatality rates of SARS-CoV-2 infection, the etiological agent of COVID-19, continues to rise across more than 210 countries and territories, several preventive and control measures have been adopted to halt the spread of the SARS-CoV-2 and minimize COVID-19 associated death. As at 7:30AM GMT+1, 27th April 2020, there were over 3 million global confirmed cases of SARS-CoV-2 infection with case fatality rate (CFR) of over 7.0% (Worldometer, 2020). European and American countries appeared to have the worst CFRs associated with COVID-19 and least in Africa. Although, there is no categorical explanation for this variation. Several observers have attributed the low incidence rate of COVID-19 in Sub-Saharan Africa to under-diagnosis probably due to inadequate molecular diagnostic capacity. But variation in the genetics, strains, viral proteins mutations and host immune response could have contributed to SARS-CoV-2 virulence and pathogenesis (van Dorp et al., 2020).

Although, there have been controversies on the origin of SARS-CoV-2. However, several studies have been able to trace the source of this infection to zoonotic origin where COVID-19 patients

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

were exposed to an animal market in Wuhan where live animals were sold. Subsequently, efforts have been made to search for a reservoir host and/or intermediate hosts of SARS-CoV-2 from which the infection might have spread to humans. Initially, two snake species were identified as the possible reservoir hosts of SARS-CoV-2. However, the only consistently identified SARS-CoV-2 reservoirs are mammals and bats (Bassetti et al., 2020). Particularly, genomic sequencing of some SARS-CoV-2 isolates show 88% nucleotide homology with two bat-derived-SARS-like CoVs (Lu et al., 2020). Thus, indicating bats as the most likely reservoir hosts for SARS-CoV-2 (Lu et al. 2020).

Virologically, SARS-CoV-2 is a single stranded RNA virus with positive polarity and variable open reading frames (ORFs) (Cui, Li, & Shi, 2019). It has been shown that two-third of SARS-CoV-2 genome are located within the 1st ORF which translates the pp1a and pp1ab polyproteins. These polyproteins encode 16 non-structural proteins (Cui et al., 2019). Conversely, the remaining ORFs code viral structural and accessory proteins of SARS-CoV-2. The remaining one third genome codes the nucleocapsid (N) protein, spike (S) glycoprotein, matrix (M) protein, and small envelope (E) protein of SARS-CoV-2. Out of these 4 proteins, the S glycoprotein is key because of it is role in host cells attachment and pathogenesis of COVID-19. This protein alongside the viral RNA dependent RNA Polymerase (RdRP) have largely been utilized in the synthesis of primers and antigens for molecular and serological tests of SARS-CoV-2 infection, respectively (WHO, 2020).

Indeed, RNA viruses including SAR-CoV-2 have high mutation rates which are significantly correlated with enhanced virulence and evolvability (Duffy, 2018). At proteomic level, amino acid substitutions have been reported in NSP2, NSP3 and S protein (Wu et al., 2020). Interestingly, another study suggested that NSP2 and NSP3 mutations play significant role in virulence and

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

differentiation mechanism of SARS-CoV-2 (Angeletti et al., 2020). Of interest, is the mutation in S-protein. This has made scientists explore the possible differences of the host tropism and transmission rate of SARS-CoV-2. Worthy to note is the NSP2 and NSP3 mutations in SARS-CoV-2 isolated from many COVID-19 patients in China (Angeletti et al., 2020). These have caused scientists to embark on genomic surveillance of SARS-CoV-2 in order to determine the correlation of these mutations to virulence diversity and their implications on reinfection, immunity and vaccines development (Liu et al., 2020).

Based on available genetic analysis, SARS-CoV-2 is related to SARS-CoV-1 and an early mathematical modelling report revealed an R_0 of 2 - 3 for SARS-CoV-2 (Yan et al., 2020). This could possibly explain why SARS-CoV-2 is more contagious than SARS-CoV-1 and MERS-CoV which are endemic in certain countries. Infection rate can be referred to as the R_0, a measure of the transmissibility of SARS-CoV-2. R_0 predicts how many people an infected person could transmit SARS-CoV-2 in a population with no prior immunity to the pathogen. Generally, the higher the R_0, the more contagious the pathogen. An R_0 of < 1 means that the outbreak dies out while $R_0 > 1$ means the infection will continue to spread (Prompetchara, Ketloy, & Palaga, 2020).

First, it is worthy to note that a single SARS-CoV-2 infected individual has the ability to infect approximately three uninfected persons (Kampf, Todt, Pfaender, & Steinmann, 2020). This occurs through the nano-particles of respiratory droplets which can spread, contaminate surfaces and hands where they remain stable for hours (Kampf et al., 2020). The hands now become a mechanical vector. Thus, a potential site to terminate and prevent it from invading the body. However, if the virus is not eliminated at this stage, it can move towards it is predilection site (i.e. cells of the lungs) where it uses it is spikes to dock and attach angiotensin-converting enzyme-2 (ACE-2) as receptors to gain access into epithelial cells of the respiratory tract. At this stage,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

SARS-CoV-2 compromises innate lung immunity (Kampf et al., 2020). It then takes advantage of these cells as replication site. The virus regenerates and sheds by disassembling itself and utilizing the machinery of the alveoli cells, precisely Golgi apparatus, to reproduce and repackage itself (Kampf et al., 2020).

The SARS-CoV-2 exists in such a way that as it replicates and disrupts the protective function of the ACE-2 receptor which induces the process of fibrosis (scarring). It has been shown that patients with fatalities associated with SARS-CoV-2 present with a characteristic ground glass effect in their lungs and this sequela impedes efficient oxygenation. As the body tries to compensate for this deficiency, it gradually results to Severe Acute Respiratory Syndrome (SARS), where it becomes impossible for the respiratory system to make available oxygen to the rest of the body (hypoxia) (Kampf et al., 2020). This ultimately results in multiple organ failure. Based on available clinical data, those susceptible to contracting severe form of SARSCoV-2 infection include the elderly (>65 years) population, persons with underlying disease conditions (cardiovascular, metabolic, respiratory and immunological disorders) (Lippi and Plebiani, 2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

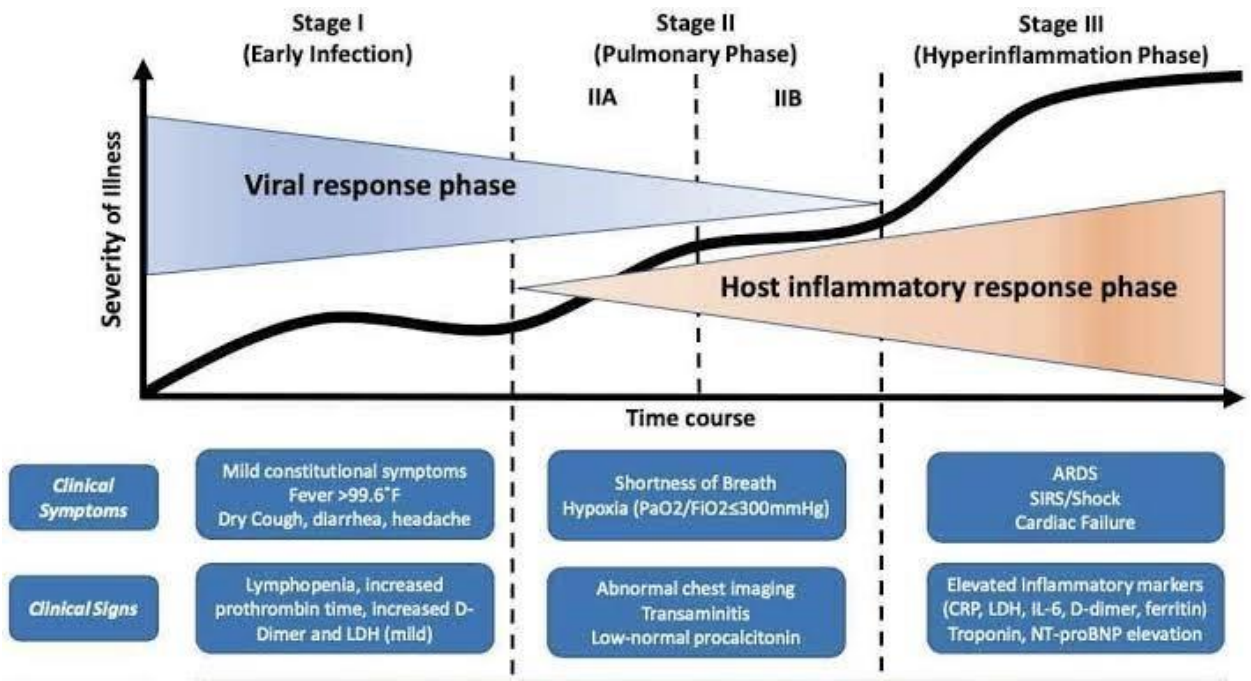

Figure 1: Cardinal Clinical, Laboratory Features and Phases of COVID-19

When susceptible individuals get infected by SARS-CoV-2, its either the infected person remain asymptomatic (no apparent illness) or symptomatic. If symptomatic, the disease passes through three stages of severity. At the early infection stage (Stage I), patients present with mild clinical symptoms which includes dry cough, diarrhea, fever and headache. This could last for 3-5 days. These are usually accompanied by lymphopenia (low white blood cell counts), elevated prothrombin time, D- dimer and mild increase in Lactose Dehydrogenase (LDH). Majority (98%) of SARS-CoV-2 infected patients remains at this stage and eventually get cured. However, for those with underlying medical disorders, they may proceed to stage II (Pulmonary Phase), which is predominantly characterized by Shortness of breath and hypoxia (inadequate oxygen supply to the body). This could last from 5 days to 3 weeks). At this stage, patients display abnormal chest radiograph, Transaminitis and declined procalcitonin level. Very few patients (2%) proceed to the severe stage of COVID-19. The stage II (hyper-inflammation phase) is largely characterized by

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

acute respiratory distress syndrome (ARDS), severe inflammatory response syndrome (SIRS), shock and cardiac failure. Majority of patients at this stage eventually die. At this stage COVID-19 patients experience significantly high blood inflammatory markers such as elevated C-reactive protein (CRP), Interleukin-6 (IL-6), D-dimer and ferritin. In addition, affected patients present with increased blood level of Cardiac markers especially troponin and N-terminal (NT)-pro hormone B-type natriuretic peptide (NT-proBNP) (Gao et al., 2020), the summary is presented in Figure 1.

Diagnostically, the use of viral culture for establishing acute COVID-19 diagnosis is not practicable due to it is long turnaround time (3 days) for SARS-CoV-2 to cause obvious cytopathic effects (CPE) on Vero E6 cells. In addition, isolation of SARS-CoV-2 is laborious and requires biosafety level-3 (BSL-3) facilities which are unavailable in most healthcare centers, especially in developing countries. So far, all available serum antigen (such as the S-glycoprotein) and antibody (IgA, IgM and IgG) detection tests have not been validated by the WHO. However, it has been suggested that serological assays could assist in the analysis of an ongoing SARS-CoV-2 outbreak and retrospective evaluation of the incidence rate of an outbreak (WHO, 2019). In some instances, where epidemiological data of suspected cases correlates to SARS-CoV-2 infection, the demonstration of fourfold rising antibody titer between acute and convalescent phase sera could support diagnosis of COVID-19 when RT-PCR results are negative (WHO, 2019). In addition, it has been revealed that a significant proportion of COVID-19 patients had tested RT-PCR negative despite having suitable clinical features and radiologic findings highly suspicious of SARS-CoV-2 infection (Xiao, Wu, & Liu, 2020). In most cases, these are termed false negatives which could have been due to wrong sampling where SARS-CoV-2 might have been present in the lower respiratory tracts rather than upper respiratory samples often collected during laboratory diagnosis.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Hence, this poses a challenge in the proper evaluation of SARS-CoV-2 symptomatic patients (Li et al., 2020).

3. Rudiments of smart cities and machine learning

This section discussed the concept of smart city including case studies and provide brief explanation about machine learning. This can help readers new in the domain to comprehend the concept of smart city and machine learning.

3.1 Smart cities

The universally accepted standard definition of a "smart city" is not yet in existence. In order words, there is no existing standard definition of "smart city". However, a smart city, is a city that encourages the prudent utilization of quality resource management as well as the provision of services within a limited time. The information and communication technology (ICT) is one of the major component and integrate element in the smart city projects. The operations in smart city cannot be achieved with ICT in isolation. The state-of-the-art view of city development resulted to the new concept of smart city model. The quality and scale of cities have grown significantly as a result of urbanization after the industrial revolution. The expansion in the urbanization have prompted many challenges including (Smart & Cooperation, 2014):

i. Large scale consumption of resources
ii. The degradation of the environment
iii. Unfair widening of gap between the rich and the poor

The challenges prompted by the urbanization trigger the new concept of city development such as knowledge city, eco city, livable city and low carbon city to solve the challenges of the urbanization as previously listed. At this juncture, the interpretation of smart city as only

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

technological project can be considered as a miss conception of the concept of smart city. However, the smart city is an integral component of the smart city as earlier discussed. To put smart city in its proper position, smart city is the combination of a very large ranges of services that is required by a city and the need to offer the services in such a way that it complies with the current administration requirement through the use of state-of-the-art technology (Smart & Cooperation, 2014).

As such, the goals of a "*good city management*" are ideal for the smart city development, examples of such goals include but not limited to: low emission of carbon and high quality of life. The management of any complex institution such as city requires the processing and exchange of information. Therefore, ICT is the life wire with substantial contribution towards information processing and disseminations in the smart city. The smart city is a new city development build based on the exploration and wide range applications of state-of-the-art ICT technology. The smart city provides new measure and solutions in assisting government transformation of its functions for improving the innovation of social management. The smart city has the potentials for helping the cities to achieve sustainability such as high level efficiency, high economy, improved standard of living for people and beautiful city environment. There are many criteria to assess the smartness of a city as proposed by organizations. These criteria include all or some of the following: smart energy production and conservation, smart mobility, smart economy, smart living, ICT economics, smart environment, smart governance, standard of living and smart society (Smart & Cooperation, 2014).

A city is referred to as smart city if the city has the presence of one or more of the following features: *Digital infrastructure*, for example, High speed broadband, novel ICT infrastructure, fiber optics cable, wireless technologies, etc. *Smart transportation*, for example, real time

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

information about bus scheduling or time table, pools of electrical vehicles, bike schemes etc. *Data*, for example, data collection, storage and analysis to predict and plan to accommodate the future challenges, information processing and service development. *Smart and sustainable building*: motion detectors, smart meters, automatic weather forecast, charging point for electrical vehicles, smart appliances, etc. *Renewable energy and energy efficiency*: Sensors for monitoring traffic congestion, pollution and emissions, smart grids, street lightning, collection systems for waste etc. A comprehensive smart city, integrate the majority of the features to facilitate government operations. A control room (city operating system) is dedicated for controlling the features of the smart city(Smart & Cooperation, 2014).

3.1.1 Case studies of smart cities

Case Study 1: Kuala Lumpur, Malaysia

In Kuala Lumpur, there are many projects in relation to smart city. Many of the projects are in use to optimally utilized the resources in Kuala Lumpur. Kuala Lumpur is currently implementing innovations and new technologies in the area of transportation, renewable energy, trading, environment, security, etc. To reduce traffic congestion, many innovations are implemented in the transport sector. For example, flight check in can be done using smart phones, train and buses schedules are monitored through electronic board place in the train and bus stations in Kuala Lumpur. Smart cards referred to as “touch and go” are commonly use in Kuala Lumpur to avoid long queue in purchasing bus or train ticket. An application referred to as “GrabCar” is used for booking a cap and track the position and estimated time of arrival of the cap to the pick-up position. The weather temperature for the day in different locations in Kuala Lumpur can be monitored through the smartphones. None smoking areas are embedded with sensors to trigger alarm in case of smoking cigarate in the none smoking zone. For the environment, the public transportation

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

mainly used gas for fueling their vehicles to avoid pollution of the environment. Though, the number of electric vehicles in Kuala Lumpur is increasing, thereby trigger the increase of charging points in the city. Many of the electric vehicle in Kuala Lumpur are hybrid (petrol + electric) which enable switching the engine from electric to petrol and back to electric. There are many innovations and expansion works in Kuala Lumpur leading to more smart city.

Case Study 2: Copenhagen, Denmark

In Copenhagen, numerous projects of smart city were analyze from 2 perspective, namely, success factors and the economics. The Boyd Cohen list of smart cities in Europe indicated that Copenhagen is ranked number 8. Copenhagen have the vision of becoming the world pioneer carbon-neural capital by the years 2025. As such, the Copenhagen city is presently implementing innovations in the field of transportation, waste, water, heating and sources of alternative energy to support the target vision of the city by 2025 and enhance sustainability. The Copenhagen has currently expanding its cycle network lanes. The cycle lane is embedded in the broad concept so as to enhance traffic in the Copenhagen city. Examples, switching from bicycles to public transportation as well as providing enough space for parking bicycles (Catriona Manville et al., 2014).

Case study 3: Stockholm, Sweden

The Stockholm is currently implementing smart management and applications to ease traffic and issues related to environment. In Stockholm, several waste collecting vehicles are deployed in the city for the management of waste. Despite the efforts put in place, the Stockholm city is facing challenges of waste transport and city traffic. Therefore, half a million entries of waste fractions, weights, and locations were comprehended. The large amount of data about the waste were used

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

for the collection and analysis of waste management collection data. This allows the identification of the inefficiencies in the waste collection routes in the Stockholm city. Thus, a shared waste management vehicle fleet is suggested as a solution for the waste problem identified (Shahrokni, Van der Heijde, Lazarevic, & Brandt, 2014).

3.2 Machine Learning

Machine learning is a field of science that centers on how a computer learns on data (Abu-Mostafa, Magdon-Ismail, & Lin, 2012). According to (Portugal, Alencar, & Cowan, 2018), "Machine is an algorithm that uses a computer to simulate human learning and allows computers to identify and acquire knowledge from the real-world, and improve the performance of some tasks based on this new knowledge". Machine learning is a discipline that cuts across many fields of studies that correlates with data mining, pattern recognition, computer science (theoretical), artificial intelligence, and statistics (Deo, 2015). In statistics, it seeks to learn the relationship that exists in data whereas, in computer science, it emphasizes the effectiveness of the computational algorithm. The intersection of mathematics and computer science in machine learning is powered by the innovative computational challenge of constructing a statistical model from a huge amount of dataset, ranging from billions to trillions of data points. Machine learning research in computer science examines the algorithm that it can learn from and produces a prediction on data. To achieve that, the input data is employed to construct a model so that a data-driven decision can be made with various static program instructions (Fuyan, 2005).

The machine learning algorithm can be broadly categorized into supervised learning, unsupervised, and semi-supervised learning. Supervised learning (input observation mapped with output observation) is learning where the input observation consists of features and the output observation consists of labels (Hastie, Tibshirani, & Friedman, 2009).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Thus, it constructs a model by utilizing a labelled dataset as input (Mohri, Rostamizadeh, & Talwalkar, 2012) and produces a labeled output data. The primary purpose of supervised learning is to drive a functional correlation from the training data with well-generalized testing data. Some of the examples of supervised learning algorithm are employed in classification and regression problems which includes Naïve Bayes, Decision tree, and Logistic regression. They stand as a basis for other learning algorithms with similar concepts. Unsupervised learning, on the other hand, is a learning algorithm that is employed when there are difficulties in finding the labeled sample since it does not rely on the previous training for mining the data. Thus, there is an existence of only one observation. The primary purpose of unsupervised learning is to find a correlation that exists between the samples behind the observation. One of the notable examples of unsupervised learning is a clustering system. The semi-supervised learning exists in-between the supervised and unsupervised learning. Thus, it is a combination of supervised and unsupervised learning that uses a small amount of labelled data and a huge amount of unlabeled data (Tsur, Davidov, & Rappoport, 2010) during the training process. Semi-supervised learning was created due to the high cost of labeling data in some complex applications. Information recommendation systems and semi-supervised classification are examples of a semi-supervised learning algorithm.

In addition to this categorization, there is one of the subsets of a machine learning algorithm called deep learning. Deep learning is a learning algorithm that employs a neural network to automatically learn from a very large sample of data (Nweke, Teh, Al-Garadi, & Alo, 2018). It is comprised of characteristics that have a similar function to the nerve system of the human brain. The machine learning algorithm can be applied in many fields of research which include natural

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

language processing (Oyelade and Ezugwu, 2020), medical diagnosis, financial data analysis, bioinformatics, and video surveillance.

4. Mathematical modelling of the COVID-19 spread in smart cities

In this section, we develop the proposed model representing the dynamics of COVID-19 in smart cities. For a comprehensive and detailed model representing the transmission dynamics of the novel COVID-19, we divided the total human population existing within a smart space into various compartments and the various intervention strategies adopted for the control of the transmission and management of the disease.

The first compartment in this model is the susceptible group or population. In this population, there are new entrants through birth only into the population (since we assume that restriction of movement is used as a control strategy with in a smart space), and Λ represents it. Note that due to the model looking at COVID-19 over a long period, natural birth rate and death rate cannot be ignored. Furthermore, any person in this compartment will either be vaccinated or unvaccinated (giving a reason for the quest for vaccines as a control strategy). Also, some people in this group can die from natural causes as they are not yet infected. Thus, the susceptible group is further subdivided into the susceptible vaccinated (S_v) and the susceptible unvaccinated groups (S_{uv}).

The susceptible group that has been vaccinated initially may lose the immunity induced by the vaccine and, therefore, can either be infected or move straight back to the susceptible group that is unvaccinated. They may equally die naturally represented by μ. For the susceptible group that is unvaccinated, they will on contact with infectious persons get infected at a rate α_1 and then join the exposed group. They may equally die naturally. Note that every vaccinated or unvaccinated

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

person might not develop or contract the disease as far as they do not have contact with an infectious person. The next group is the exposed group.

We need to note here that the exposed group (E) comprises of all persons that have had contact with an infectious person (this is where effective contact tracing comes in as a control measure) regardless of the infection status until confirmed to be negative. Note that this model does not include false positives and false negatives. So, those who become infected or had contact with the infections persons and are in the vaccinated or unvaccinated subgroups move into this exposed group. Since person can be infected, if he has contact with infected surfaces, it means such persons also join this group. From this exposed group, we now either self-quarantine, government-quarantine, or the individual develop the symptoms and move straight to the infectious subgroup (I). The personal-quarantine subgroup is represented by P_Q and government-quarantine by G_Q. From these quarantines, one is either confirmed free from the virus and moves back to the susceptible group or confirmed to be infected and moves to join the infectious group. At this level, the only death considered is from natural causes. After these subgroups, we now consider the infectious subgroup represented as (I). This group comprises of those who became infectious as previously described and those who became infectious by inhalation of the droplets of the virus from the air either due to close proximity to an infectious person sneezing, coughing or droplets just in the air as a recent investigation has shown that the droplets can last for more than nine (9) hours in the air (Lewis, 2020) and this is done at a rate σ.

When a person is infectious, an assumption is made that such a person must undergo treatment. The person must self-isolate and get treated (P_T) or be government-isolated and treated (G_T). We note that the infectious persons are joined by those declared as infectious through self-isolation

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

(h_2) or government-isolation (h_1). Also, as noted recently in China, some who had hitherto been declared recovered got re-infectious and then moved back to the infectious subgroup (Wu and McGoogan, 2020; Li, Geng, Peng and Meng, 2020). Those on personal-isolation and treatment (P_T) decreases by those who die either naturally or as a result of the disease. Equally, some persons recover. Similarly, those on government-isolation and treatment will decrease by those who recover, or those who die either naturally or due to COVID-19. Finally, we have the recovered group, and this comprises of those who recovered due to treatment either personally or by government. This population decreased by those fully recovered rejoining the susceptible group of getting re-infected and then infectious and rejoin the infectious group. Based on these explanations, we have the assumptions of the model and then the following flow diagram showing the sequence of development and control of the disease and then the model equations for the COVID-19. We show these as:

Assumptions of the Model

1. Some susceptible persons are vaccinated.
2. There is contact tracing of all persons that had contact with infected persons and all such persons undergo quarantine to confirm their disease status.
3. Personal and government quarantines exist.
4. There is personal or government treatment, but every infectious person must be treated.
5. It is not only through person-to-person contact that infection can occur.
6. Immunity can be lost for some reason over time hence rendering the vaccinated person susceptible or infected.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

7. The COVID-19sometimesinduces death, and so infectious persons can die through COVID-19.
8. Recovered persons can be re-infected or susceptible.
9. One can become infectious by inhalation of the droplets of the virus from the air.
10. Those on quarantine can be confirmed free after the incubation period (2 - 14 days) and can either rejoin the susceptible or the infectious group if re-infected.
11. Some persons may have contact with infected persons and never undergo any quarantine and so they develop the infection and move straight to the infectious subgroup.
12. At the infected level, one can die even before personal treatment or government treatment.

Therefore, using all these assumptions, we have the flow diagram as shown in Figure 2 as follows:

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

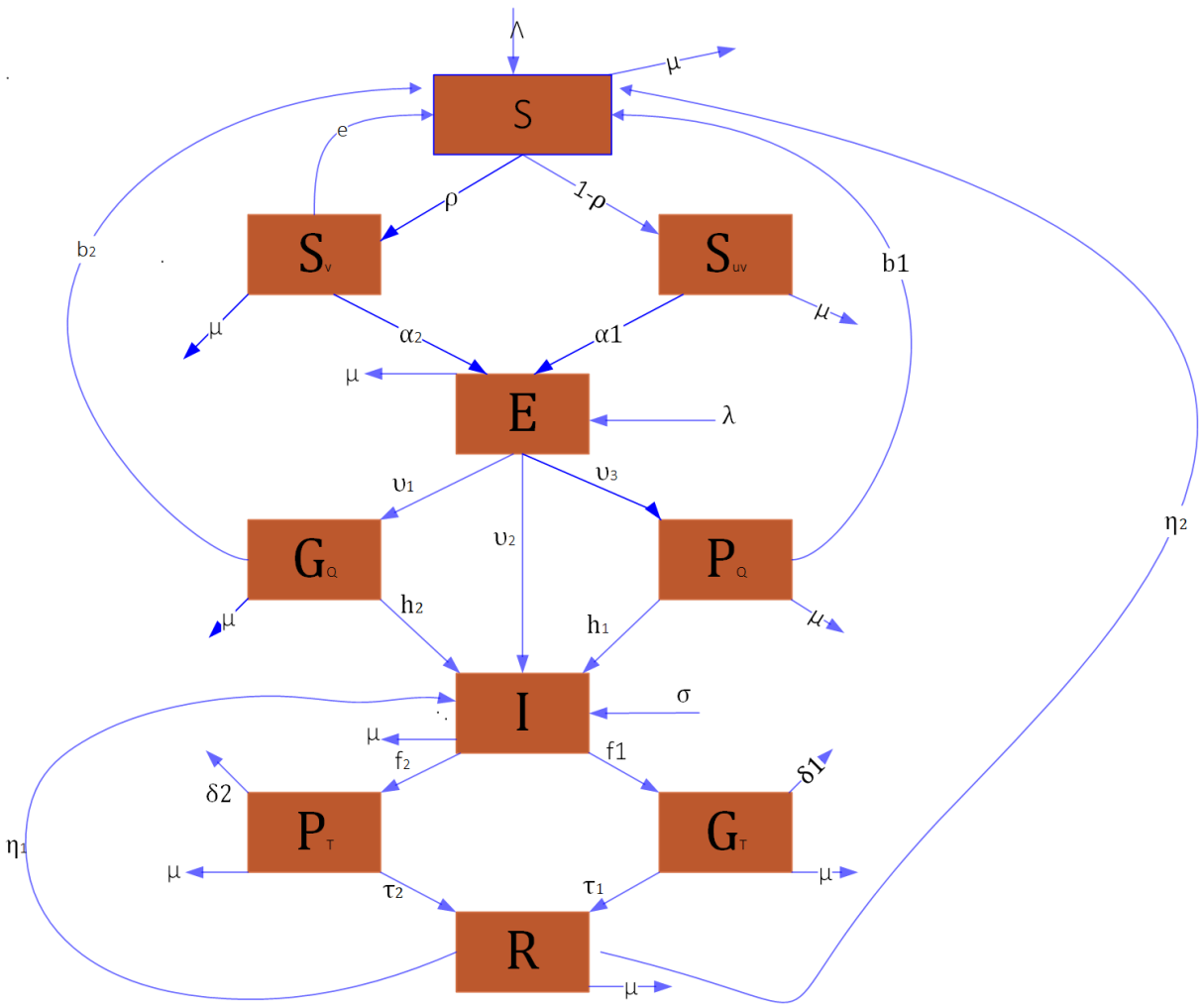

Figure 2: Diagram summarizing the model for COVID-19in a smart city

Definition of the parameters and Symbols used in the Model

Table 1: Summary of parameter definition of the proposed mathematical model

SYMBOL	DEFINITION OF PARAMETERS/SYMBOLS		DEFINITION OF PARAMETERS
S	Susceptible individuals	b_1	The rate at which those on personal quarantine are confirmed free and returns to susceptible class
S_v	Vaccinated Susceptible individuals	b_2	The rate at which those on government quarantine are declared free and return to susceptible class

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

S_{uv}	Unvaccinated susceptible individuals	e	The rate at which vaccinated persons loses immunity ad returns to susceptible class
E	Susceptible individuals that had contact with infectious persons (Exposed)	h_1	The rate at which those on government quarantine develop the infection
G_Q	Persons quarantined by the government	h_2	The rate at which those on personal quarantine develop the infection
P_Q	Individuals exposed and on personal quarantine	λ	The rate at which susceptible persons have contact with surfaces infected with COVID-19
I	Infectious individuals	σ	The rate at which persons become infectious through inhalation of the droplets of COVID-19
P_T	Infectious individuals on personal treatment	f_1	The rate at which infectious persons go for personal treatment
G_T	Infectious individuals on government treatment	f_2	The rate at which infectious persons are treated by the government
R	Recovered individuals	τ_1	The rate at which those treated by government recovers
		τ_2	The rate at which the self-treated group recover
PARAMETER	**DEFINITIONS**	δ_1	The rate of COVID-19 related death among the government treated group
Λ	Persons that join the Susceptible population S	δ_2	The rate of COVID-19 related death among the self-treated group
ρ	The rate at which susceptible are vaccinated	μ	Natural death rate
α_1	The rate at which the unvaccinated persons become infected or had contact with an infectious person	η_1	The rate at which those that recovered again becomes re-infected and infectious.
α_2	The rate at which the vaccinated persons become infected or had contact with an infectious person	η_2	The rate at which those that recovered get well and join the susceptible group again
v_1	The rate at which government quarantines the persons that had contact with infectious persons		

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

v_2	The rate at which exposed persons become infectious
v_3	The rate at which individuals quarantine themselves because they had contact with infectious persons

The corresponding new diagram summarizing the main structure of the above-described model can be presented in Figure 2, and the resulting system of equations describing the model is expressed in the following form:

$$\frac{dS}{dt} = \Lambda - \rho S - (1-\rho)S - \mu S + b_2 G_Q + eS_v + b_1 P_Q + \eta_2 R \tag{1}$$

Equation (1) describes the change in the populations of the susceptible individuals in the populations over time.

$$\frac{dS_v}{dt} = \rho S - (\alpha_2 + e + \mu)S_v \tag{2}$$

Equation (2) describes the change in the population of the vaccinated susceptible individuals over time.

$$\frac{dS_{uv}}{dt} = (1-\rho)S - (\alpha_1 + \mu)S_{uv} \tag{3}$$

Equation (3) describes the change in the population of the unvaccinated persons over time.

$$\frac{dE}{dt} = \alpha_1 S_{uv} + \alpha_2 S_v + \lambda S - (v_1 + v_2 + v_3 + \mu)E \tag{4}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Equation (4) describes the time-change in the number of persons exposed which includes all persons that had contact with infected persons or surfaces.

$$\frac{dG_Q}{dt} = v_1 E - (\mu + h_1 + b_2) G_Q \tag{5}$$

Equation (5) describes the time-change in the population of persons under government quarantine.

$$\frac{dP_Q}{dt} = v_3 E - (\mu + h_2 + b_1) P_Q \tag{6}$$

Equation (6) describes the change in the population of those under personal quarantine over time.

$$\frac{dI}{dt} = v_2 E + h_1 G_Q + h_2 P_Q + \sigma S - (f_1 + f_2 + \mu + \delta) I + \eta_1 R \tag{7}$$

Equation (7) describes the change in the population of the infectious persons over time.

$$\frac{dP_T}{dt} = f_1 I - (\tau_2 + \mu + \delta_2) P_T \tag{8}$$

Equation (8) describes the change in the population of those on personal treatment over time.

$$\frac{dG_T}{dt} = f_2 I - (\tau_1 + \mu + \delta_1) G_T \tag{9}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Equation (9) describes the change in the population of those on government treatment over time.

$$\frac{dR}{dt} = \tau_2 P_T + \tau_1 G_T - (\eta_1 + \eta_2 + \mu)R \tag{10}$$

Equation (10) describes the change in the population of those who recovered over time.

5. Applications of machine learning in fighting COVID-19 pandemic in smart cities

Smart city being an inter-connected urban society and collecting data every moment from several embedded devices, smart cities can effectively work with machine learning approaches during this COVID-19 pandemic. It basically implies that machine learning techniques are hugely dependent on data for better learning and predictive models. Thus, the machine learning techniques have the ability to bring out some intrinsic and useful insights to help smart cities decision makers to take preventive measures during the COVID-19 pandemic. In view of the fact that different machine learning techniques are embraced by Artificial intelligence (AI), which is the model with the ability to self-learn, can also support immensely. It is important to discuss the role of the AI along with the machine learning to combat COVID-19 because first of all the data availability is too limited and we have to deal with real-time streaming of the data. Thus, significance of self-learning systems becomes much more desirable in smart cities. Figure 3 demonstrates the overall flow of how AI & machine learning approaches can help in fighting COVID-19 pandemic in smart cities.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

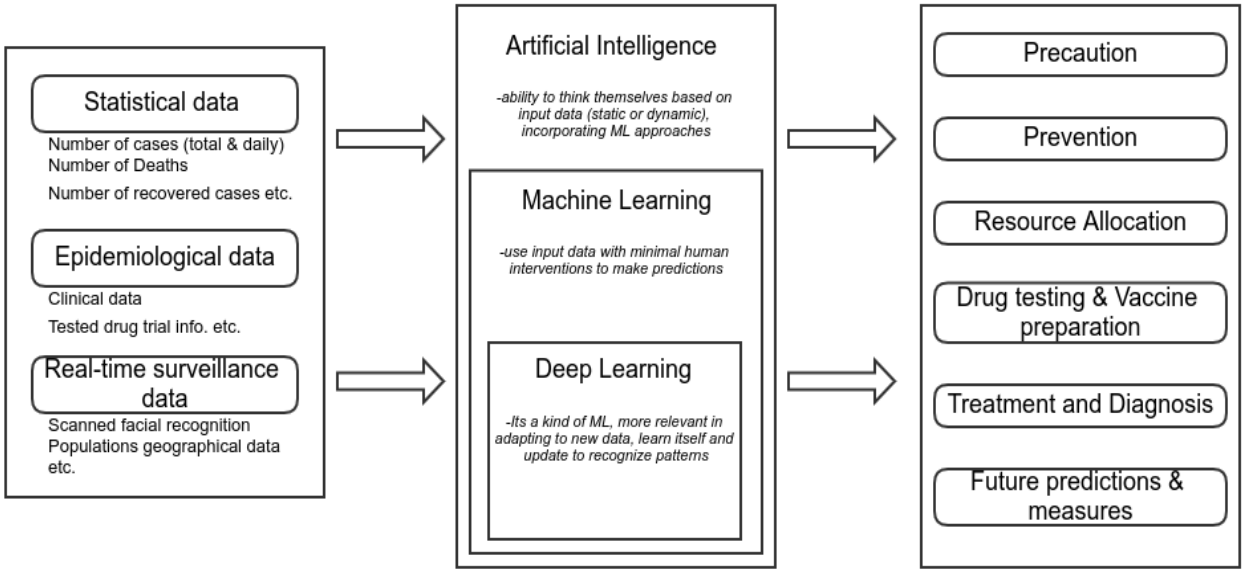

Figure 3: Application and flow of AI & machine learning in combatting COVID-19

As shown in Figure 3, there are several types of data generated from the information and communication technology equipment embedded in the smart cities. These data are as follows:

i. First kind of data is the statistical data that normally contains the daily statistics of number of identified cases, no of positive cases, number of deaths, number of recovered cases, etc. This would help in the prediction of future cases to prepare for emergencies.

ii. Second type of data is the epidemiological data, which majorly contains all of the clinical tests data concerning test results of different medication, various drug trails, patient's medical history, patient's response to several medications, etc.

iii. Third type of data is the real-time surveillance data generated from sensors and cameras in the smart cities. One of the initial detection of COVID-19is based on symptoms is the fever. People body temperature as per facial recognition and other personal information are the type of data which are also useful to prevent spreading of COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The data is processed and analyzed through the machine learning approaches for extracting insights in various applications. Notice that deep learning is shown in Figure 3 inside the machine learning block. It's a type of machine learning algorithm which is much more relevant in adapting to large scale new data, learn itself from the data and update to recognize patterns. The applications of the machine learning in different aspect for combatting COVID-19 are discussed as follows:

5.1 Prevention and precaution

Based on the statistical data, machine learning model can be used to predict the nature of the identified cases to better take the preventive measures. During the pandemic situation, there is a chaos and quick testing of individuals on the large scale which is very challenging. Rather than going door to door to each patient, even a less accurate but faster approach is much more acceptable. Machine learning may help in quickly diagnosing the patients in the smart cities as follows:

i. Facial recognitions with the help of sensors and cameras to scan the patients for body temperature and personal information so that if the particular patient is positive then his/her and his/her nearby individuals can be tested and alarmed.

ii. Helping the patients with the AI powered chatbots for self-awareness and answering the queries which might be impossible for the medical professional to address during this COVID-19 because of very high number of patients.

iii. Use the data from the smartphones and wearable smart watches to monitor the heart rate and daily activity of the citizens.

Although, predictions based on the statistical data may not be 100 % accurate, but they can still enable the decision makers in the smart cities to take some preventive and proactive measures.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

5.2 Prediction models

Medical science (especially dermatology) was one of the real-world fields where AI and machine learning approaches were successfully implemented. Computer vision and machine learning prediction models are enforced for identifying most of the diseases common in the patients by just learning from the images. In case of COVID-19, based on some set of crucial features (set of symptoms), machine learning approaches can help in predicting the following cases:

i. Prediction of a person infected with COVID-19.

ii. Prediction of a positively diagnosed COVID-19patient to be hospitalized.

iii. Scope of certain treatments to be effective while on treatment. This also includes predicting the chances of a COVID-19patient being successfully cured or not survives at all.

(Pourhomayoun & Shakibi, 2020) used the machine learning techniques to predict the mortality rate of the patients infected withCOVID-19 disease. They used the machine learning algorithms such as: random forest, logistic regression, decision tree, support vector machines, artificial neural networks etc. to predict up to 93% total accuracy in the prediction of the mortality rate. Moreover, the study also used the machine learning models to extract the essential and exclusive symptoms and features to detect the virus.

5.2.1 Prediction of COVID-19 pandemic

Different studies have been conducted to predict the likely occurrence of the COVID-19 pandemic (Arkes, 2001).For instance, (Ndiaye, Tendeng, & Seck, 2020) conducted a prediction study on COVID-19 pandemic globally between January and April 2020. The study employed prophet (Taylor & Letham, 2017), a tool for predicting time series data, which depends on the additive

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

model that fits real non-linear trends with daily, weekly, and annual seasonality together with holiday effects. Four countries involving Italy, China, Senegal, and Iran were selected as case studies for the research. However, the predictive performance of the study shows that the COVID-19pandemic in some countries like China under the optimistic estimation will end in few weeks whereas the strike of the anti-pandemic in other countries of the world like Italy, Senegal and Iran will end within the end of April. Similarly, (Wang & Wong, 2020) proposed a COVID-Net, using a convolutional neural network design for the identification of COVID-19 cases using chest X-ray (CXR) images. The study utilized the CXR dataset that comprised of 13,800 chest radiography images obtained from 13,725 patients from three public datasets. The experimental analysis shows that the proposed COVID-Net attained a predictive accuracy of 92.6% on the test data, which indicated the importance of combining human and machine collaborative design strategy for building modified deep neural network architectures in a faster mood fitted around the data, task and working requirements.

In another study, (Yang et al., 2020) proposed a modified Susceptible Exposed Infections Removed (SEIR) and AI prediction of COVID-19pandemics. The study employed the most updated COVID-19 epidemiological data together with population migration data obtained prior and after January 23, 2020, into the SEIR model. Besides, a machine learning approach was employed to train on the 2003 SARS data for the pandemic prediction. The predictive result of the study shows that the pandemic of China is expected to be at peak by late February, which shows a gradual decline by the end of April. However, the cases of the pandemic would have risen higher than expected in Mainland China should the implementation of the proposed model had delayed for 5 days. Thus, the proposed model was effective in predicting the peaks and sizes of the COVID-

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

19pandemic and the implementation of the control precaution performed on 23rd January was important in alleviating the chances of COVID-19 pandemic size.

Progressively, (Gozes et al., 2020) in their study on the outbreak of COVID-19, developed an artificial intelligence-based automated computer Tomography (CT) image analysis tool using a deep learning approach for the detection, tracking, and quantification of COVID-19, which can identify patient infected with COVID-19 and those that were not infected. The study utilized various global datasets that included Chinese disease-infected areas. Various retrospective deep learning experiment was performed to analyze the system performance in the identification of speculated thoracic Computer Tomography features of the COVID-19 for the evaluation of disease evolution in every patient. One hundred and fifty-seven (157) patients were globally selected from the US and China for the testing sets. However, the classification performance of the model attained 0.996% AUC, 92.2% specificity, and 98.2% sensitivity on Chinese control and infected patient datasets. This shows that the proposed model can attain high predictive performance in the identification, tracking, and quantification of the COVID-19 pandemic. Similarly, (Narin, Kaya, & Pamuk, 2020) on their proposal for automatic prediction of COVID-19, employed a deep convolutional neural network built on Chest X-ray image and pre-trained transfer model that includes InceptionV3, Inception-ResNetV2 and ResNet50 models to attain higher predictive performance with negligible X-ray dataset. However, the experimental results attained an optimum result of 98% accuracy on ResNet50 pre-trained model out of the three selected models, which shows that the research can assist doctors on decision making in clinical practice based on the obtained result as it can employ transfer learning to detect the early stage of the COVID-19 on the infected patients.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

5.2.2 Forecasting of Mortality

Since the outbreak of the COVID-19 at Wuhan city China in December 2019, the number of confirmed dead has risen to over 115,000 deaths, which leads to the doubling of the number of deaths on weekly basis (Brown, Jha, & Consortium, 2020). There has been less bias in mortality compared with the reporting of the cases, which might have been influenced by the test policies. Thus, the daily report of deaths on COVID-19 has been on variance with the actual deaths over time. Accurate death rate estimation is important as it serves as a key factor in concluding if a highly infectious disease is seen as a public concern. Consequently, there is a need for a reliable number of mortality estimates on COVID-19, the peak of deaths date, and the period of high mortality to assist in response towards the present and unforeseen pandemics. (Brown et al., 2020) developed a statistical model; also known as Global-19 Assessment of mortality trends in 12 countries between April 12, 2020, and October 1, 2020. The selected countries include the USA, China (Hubei), Italy, Spain, France, UK, Belgium, Iran, Netherlands, Germany, Canada, and Switzerland. Besides, six US states that include New York, Michigan, California, New Jersey, Washington, and Louisiana were incorporated in the study. The mortality data were collected from the WHO country's daily reports and some obtained online data collection. The results of the prediction in the selected countries showed the estimation of the peak date of the mortality and indicated that the pandemic will be completely wiped off by July 1, 2020. Thus, the model validation provided a reasonable performance with real counts of deaths in various countries. Similarly, Wang et al. (2020) employed the Patient Information Based Algorithm (PIBA) to determine the mortality rate on the COVID-19 pandemic. The algorithm is employed to determine the death rate of the newly infected disease in real-time and forecast future deaths. The data was collected from the 3 public sites of COVID-19 patients in China. The data consists of the newly

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

infected patients with the COVID-19, the patients that died of the infection, the patients with the critical condition and were admitted to the intensive care unit (ICU), daily new cases of the people infected by COVID-19, people with close contacts to the source of infection, and new deaths. The results of the findings show that the average time frame from the beginning of the infection to the time of death is 13 days. This death rate prediction is based on the data collected from Wuhan, the first confirmed city that recorded various death cases related to COVID-19 pandemic.

5.3 Resource Allocation

Resources required to manage COVID-19 pandemic become scares as a result of very high number of people demanding such a resources. These resources ranges from ventilators, mask, testing kids, personal protection equipment, sanitizers, etc. The problem of resource allocation is a NP-Hard problem and it is impossible to solve in a polynomial time. Considering the emergency during COVID-19 pandemic situation, machine learning can be very beneficial in the resource allocation prediction on the basis of linear and logistic regression. Even on the small set of training dataset (as is the case in COVID-19), machine learning model may provide feasibility and resource allocation accuracy as close as possible to the optimal resource allocation.

5.4 Vaccine discovery

The process of discovering new vaccine based on the available clinical data may take a longer time. But with the help of machine learning approaches, the overall process can be reduced significantly without sacrificing the quality of the vaccine. For example, researcher used the Bayesian machine learning model in a study to discover vaccine for Ebola (Ekins et al., 2015). Also, (Zhang et al., 2017) used the random forest for improving the accuracy of the scores while working on the H7N9. However, even if the vaccine will take time to be found, machine learning

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

can help in finding the existing drugs which may be effective in dealing with COVID-19. Machine learning can learn from drug and protein structure and predict the integrations to generate the studies. Currently, there are different efforts from the scientific community applying machine learning to search for COVID-19 vaccine. (Gonzalez-Dias et al., 2020) reported the stages of applying machine learning to predict signature of vaccine immunogenicity and reactogenicity. The stages involve: data preparation, vaccines and relevant gene selection, selecting the suitable machine learning algorithm for modeling and lastly performance evaluation of the predictive model.

5.5 Drug discovery for COVID-19

Since the COVID-19 pandemic, it has become necessary to find the right drugs that can be employed to cure the disease. Various approaches have been attempted in finding the right drugs by either repurposing the existing one (Therapeutic) or discovering the new one. The expansion of machine learning and the development of the new architecture of deep learning has made researchers focus on the application of machine and deep learning models for the discovery of drugs that can bring a cure for COVID-19. A review of some of the studies that applied the machine or deep learning approach for drug discovery and vaccine for COVID-19 are summarized in Table 2.

Table 2: Summary of the application of machine learning in drugs discovery

Reference	Study approach	Advantage	Challenge
(Patankar, 2020)	Deep learning-based computational drug discovery	The generated molecules could serve as a good drug potential for SARs CoV and COVID-19. The result also shows that a deep learning model is important in testing the current compound and obtaining a new	By utilizing a limited data for model training, there could be an atom of error resulting from

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

		molecule for finding COVID-19 drugs.	inaccuracies of parameters.
(Chenthamarakshan et al., 2020)	Target-specific and selective drug design for COVID-19 using Deep generative model	The framework can take care of candidate generations for multiple targets by employing trained mode. Besides, it can also permit explicit accounting of off-target selectivity	Large scale data size is required for effective study

6. Case studies on combatting COVID-19via machine learning in smart cities

In this section, we present use cases where machine learning techniques are applied to help in combating COVID-19 in smart cities. The use case can help readers really understand exactly the way machine learning can provide aid in the era of fightingCOVID-19 pandemic. As such, other nations can share the expertise in fighting the COVID-19 by applying the machine learning approaches. The use cases are discussed as follows:

Case Study: New York City

In New York city, there are heavy cases of COVID-19 patients and those exhibiting the symptoms. The medical staffs in New York hospitals are Overwhelmed in view of the fact that the number of COVID-19 patients is extremely high. As a result of that, the medical staffs are facing difficulties in taken decision on the COVID-19 patient that require emergency treatment and the patient that might have case beyond medical intervention. To speed up the decision of the medical staffs in New York hospitals, machine learning system is developed through training of the system to provide clinical decisions support to triage patients. It is the system that is now in used in the hospitals in aiding clinical decisions (Spectrum, 2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Case Study: China

When it comes to the issue of data, China is well known about it is massive amount of data generated from it is citizens. China installed a network of over 200 million surveillance cameras spread across the country. In addition to video surveillance camera, biometric scanners are installed in doorways of residential complexes. As a form of registration, any resident or person that is leaving the residential building has to scan his face through the biometric scanner at the doorway of the building. After that, the embedded intelligent systems process the data and track the person location through the video surveillance. All the information are stored in a central database in which the machine learning algorithms runs the data to determine the possible social interaction of the person that leaves the residential building (Dingli, 2020).

Case Study: Canada

Migration of human across the globe contributed significantly to the spread of the COVID-19 pandemic throughout the world. BlueDot based in Canada applied machine learning and natural language processing for the tracking, recognition and the reporting of the spread of COVID-19faster than the WHO and center for disease control and prevention in United States of America. It is projected that this machine learning and natural language processing based technology can be leveraged in the future for the prediction of zoonotic infection risk to humans using climate and human activities as variables. The prediction of individual risk profile using the data extracted from social media such as family history and lifestyle as well as clinical, personal and travel data can provide precise and accurate prediction results. However, such technology can trigger privacy concerned (Obeidat, 2020).Similarly, *Virtual healthcare assistant*: the virtual healthcare assistant is a multi-lingual healthcare agent that is developed based on natural language processing. It is a

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

question – answering system that respond to questions related to COVID-19 by delivering information that is trustworthy on COVID-19 guidelines, protection measures, symptoms monitoring and checking as well as providing advised to individuals on the need for screening in the hospital or self-isolation. The healthcare agent is developed by Canada-based stallion (Obeidat, 2020).

Case Study: United States of America

In United States, many medical centers are modifying their existing intelligent system that were purposely meant to predict course of patients' illness. These intelligent systems are now being modified to predict specific type of COVID-19 outcomes like the intubation. The intelligent systems are trained to learned pattern about the illness by feeding the system with thousands of patient records as training data. However, no sufficient data to build entirely new intelligent systems for the prediction of COVID-19. Therefore, researchers are assessing the existing tools with the aim of customizing it to help in the fight against COVID-19 pandemic (Spectrum, 2020).

7. Propose smart city framework integrated with machine learning for combating COVID-19

To address the multiple dimensional challenges outlined in Section 1, as a result of COVID-19, a proposed framework is required within the smart city context to allow decision makers to take a crucial decision on the best way to combat COVID-19 from multiple dimensions. The framework consists of multiple components as shown in Figure 4. Each component has a major impact on enhancing the quality of the analytics to combat COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

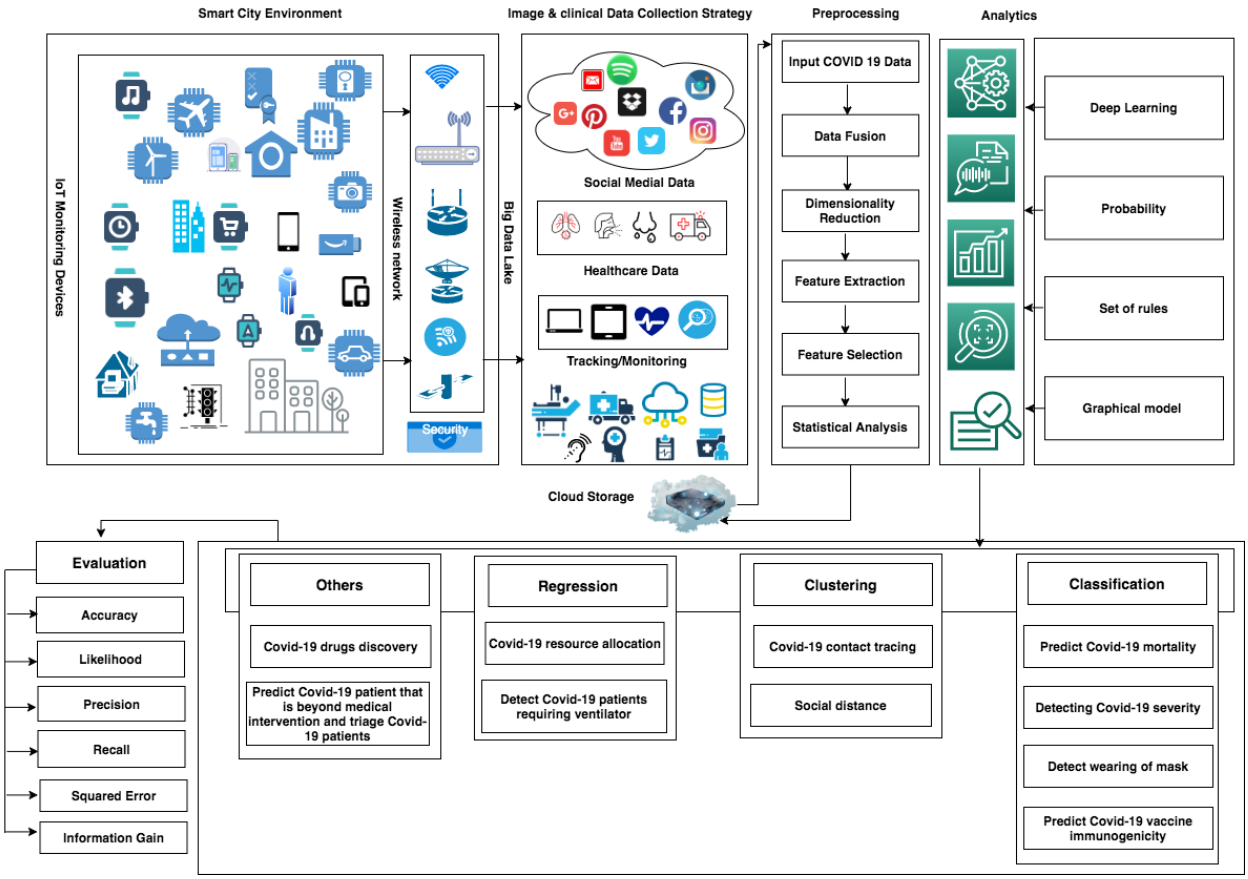

Figure 4: Propose smart city framework integrated with machine learning for fighting COVID-19

7.1 Smart Environment

Smart city technologies have recently demonstrated a major potential in enhancement of citizen's quality of life. Many smart based technologies have benefited from the adaptation of the internet of things (IoT) by witnessing the development of intelligent applications such as smart home, smart grids, smart transportation, smart industry and smart healthcare. The emphasis on smart technologies during the COVID-19pandemic could be of great help to tackle major clinical problems and diseases. Moreover, in recent year's sensors and video cameras surveillance are becoming part of smart city monitoring devices which can lead to an early detection of pandemic. Besides, health agencies may utilize IoT platforms to access data for monitoring the COVID-19

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

pandemic. For example, 'Worldometer' is able to view an instant updates of COVID-19 actual number of cases and deaths for the entire world including daily new cases of COVID-19, distribution of COVID-19 by countries and severity of the disease (Ting et al., 2020). In Figure 4, we demonstrate the smart environment, where various IoT and sensors monitoring devices can possibly be used for healthcare purposes interacting within a limited area to generated numerous clinical signs and symptoms data. These devices are connected via next-generation wireless connectivity which is capable of efficiently transferring the collected data and stored in a big data lake. Big Data plays a critical role in the smart cities because of it is ecosystem of data analytics that can make decision makers take a crucial decision on the best way to further develop strategy to combat COVID-19. With big data, the adequacy of smart city framework implementation, users can be traced at all time with the potential to mitigate any health problems the user may encounter during the movement. Therefore, improve smart city framework efficiency and effectiveness, which in turn improve the life of the citizens living in the smart cities. A literature review was conducted by (Al-Turjman, 2019) the author presents a comprehensive background about 5G standards and it is IoT-specific applications. Furthermore, an overview of recent directions in the utilization of smartphones sensors that can contribute to a scalable operation in smart social spaces is presented.

7.2 Image and clinical data collection strategy

The generated image and clinical data from the smart cities in real-time can be subjected to big data processing to better understand healthcare trends, model risk associations and predict outcomes. The result of the big data lake can be used by the government authorities and private/public healthcare providers to improve healthcare services to the citizens, this process will continue until government and healthcare services providers satisfied the citizens living in the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

smart cities. There are various ways of collecting data at large scale including social media platforms such as (Facebook, Twitter, Google+, Instagram, etc), healthcare services provided through treatment and diagnosis, and tracking monitoring devices such as GPS, vehicle tracking system, smart watches and sensors. All collected data are integrated and stored in a single location within the smart city to be accessed by the authorized entities. The possible technologies used to store such data are Hadoop distributed file system (HDFS) and NoSQL where structured and unstructured data can be stored and processed. In a study by Balduini et al. (2019), the authors proposed a new conceptual framework that put in use a variation of Big Data sources. The framework has a unified approach that make use of spatial and temporal analysis on a heterogeneous stream of data. Based on the results, the study shows generality, feasibility, and effectiveness of the proposed framework through many use cases and examples obtained from real-world requirements collected in many cities.

7.3 Pre-processing

In order to provide an accurate and better input for more reliable results to detect and prevent COVID-19 cases, data pre-processing is considered an important stage. The first steps in pre-processing is to extract all the relevant COVID-19 data from the storage. The second step is to preform data fusion where the collected data are integrated to produce more consistent, accurate, and useful information. The third step during pre-processing of COVID-19 data is to preform dimensionally reduction, in which the number of variables are reduced by extracting a set of main variables. Fourth and fifth steps focus on feature extraction and selection. The two methods are very important because it can be used to filter irrelevant or redundant features from the selected datasets. The possible activities in these steps including Wrappers, Filters, and Embedded. The last

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

step is a basic statistical analysis on COVID-19 data in order to interpret the data before intelligent based algorithms are applied.

7.4 Analytics

Recent years have received an increasing trend in image processing in healthcare due to convolutional neural network which has become a significant approach for handling large amount of images generated from the smart cities. Allam and Jones (2020) discuss the universal data sharing standards coupled with AI to benefit urban health monitoring and management. The proposed framework has the potential to incorporate various machine learning and deep learning algorithms to develop the analytical model. These algorithms range from traditional shallow approach such as neural network, Decision Tree, Naive Bayes, K-nearest neighbor to more sophisticated algorithms such as Convolutional Neural Networks, Deep Generative Modeling, Deep Belief Network and Deep Recurrent Neural Networks. These algorithms can be applied to run on COVID-19 dataset to help in the combatting COVID-19 in hospitals in smart cities across the world. These applications include detect, prevent the spread of COVID-19, forecast next epidemic, diagnose cases, monitor COVID-19 patient, suggest vaccine development, track potential patient, help COVID-19 drug discovery and provide better understand of the COVID-19virus in smart cities.

7.4.1 Social media information verification

The COVID-19 pandemic has come with the challenges of fake news associated with it including conspiracy theories. Since the COVID-19 pandemic start spreading across the globe, there has been a lot of fake news regarding the COVID-19 origin, cure, mode of spread, treatment, and many other myths. This is very common on the social media platforms such as Facebook, Twitter,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Instagram, YouTube, etc. In the smart city, citizens voice out opinion on social media regarding COVID-19, as such generate a lot of unstructured data. It is reported by (Obeidat, 2020) that no systematic quantitative study has been conducted to ascertain the magnitude of the myths regarding COVID-19 on the social media but certainly the figures of the misinformation about COVID-19is significant. The fake news regarding COVID-19 can come in the form of manipulated content, misleading content, satire, false context, malicious account, fabricated content, false connection and imposter content. Therefore, machine learning or deep learning algorithms can be applied to detect fake news regarding COVID-19 on the social media and alerts the citizens living in smart cities.

7.4.2 Prediction of Future pandemic

Although being the worst phases in recent times, this pandemic has come in the times of digital age. Therefore, every aspect of analysis is now being captured including macro-level analysis, logistics, biological etc., in terms of data which will definitely be fruitful for predicting the new and unknown (possibly sources) future pandemic. For example, with the help of machine learning approach (random forest), (Eng, Tong, & Tan, 2014) were able to predict the possible zoonotic strains of influenza i.e. some virus which could only affect animals can also be dangerous to humans. It only implies that machine learning can help in predicting future pandemic which may come from any species.

The only limitation being the data could be from different domain i.e. the source of COVID-19 is possibly from ‘bats’ and in the past, sources of pandemic are different (say, different gnome structure etc.). And traditional machine learning requires data distribution to be from the same domains in training and testing. However, Transfer Learning (TL), a part of machine learning, can effectively handle such situations where training and testing data could be from different data

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

distribution. That is, the knowledge learned from the past pandemics could be used in other situations with new domain even with fewer amounts of data. Such scenario is shown in Fig. 4a, where the pre-trained model from current COVID-19 pandemic (with large data and labels) could be used to predict future pandemics, prepare the smart city for that situation and help overcome quickly in case of the spread (with very less amount of data and labels).

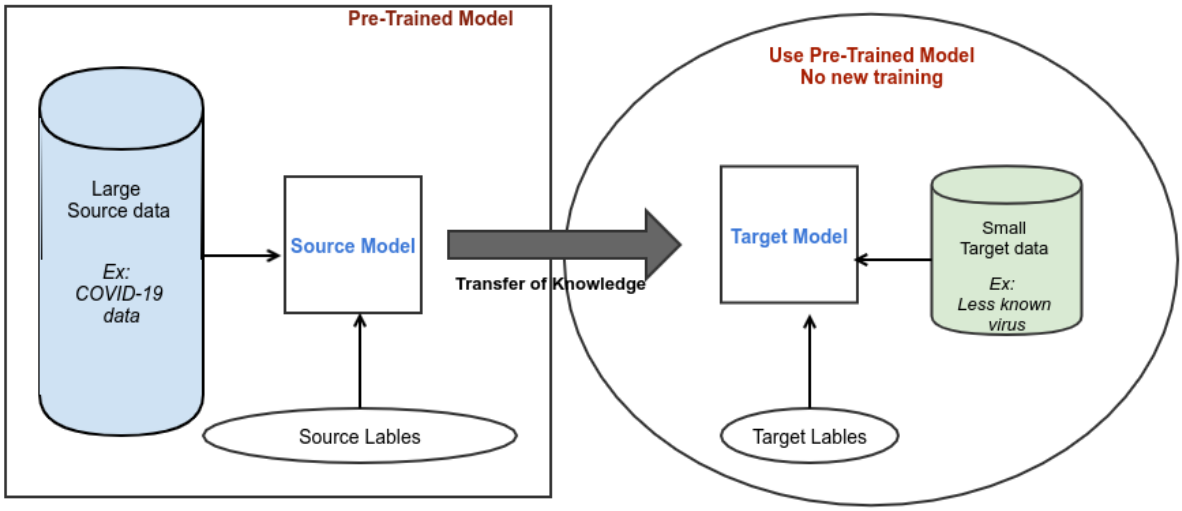

Figure 4a: Typical transfer learning model for future pandemic predictions

7.6 Containment framework of COVID-19 in smart cities: Contact tracing and basic preventive measures

In this section, we display the proposed framework. As at the time of this study, there are no defined set of procedures or approved drugs that can provide complete cure to patients affected by the recent COVID-19 pandemic. However, as temporary preventative measures, usage of face masks and shields, limiting the movement of the hand to the face, social distancing, washing of the hands vigorously with soap, water, and alcohol-based sanitizers have all been recommended for the general population. Also, for people who have been exposed to COVID-19 or display

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

symptoms, measures like usage of face masks, using personal items alone, self-quarantine, and government quarantine are some of the measures put in place. Similarly, new machine learning techniques and application such as the use of facial recognition systems embedded in different location in the smart spaces e.g. bus station, subway station, airports, etc can be developed to detect those with facial mask, or gesture recognition system to enforce social distances. However, it is crucial to note that due to the scale of the pandemic and the lack of fore-knowledge to most of the populace, obtaining adequate containment measures can be cumbersome. In this regard, scientists have made several efforts with the help of valuable artificial intelligence strategies to curtail possible escalation of the pandemic, through continuous monitoring of the COVID-19 spread. In addition, machine learning based algorithms have equally been developed and used in making efficient classifications and predictions that could suggest preventive measures of the COVID-19virus from further spread. A typical example is the application of machine learning techniques in COVID-19 contact tracing (CT), such as facial recognition and entity resolution analysis, to track persons of interest across various signal-based media and social network platforms.

The CT has been used as a method of monitoring and controlling the pandemic contagious diseases, such as Ebola (Classen et al., 1999; Browne, et al., 2015; Webb et al., 2015), Tuberculosis (Begun, Newall, Marks, & Wood, 2013), and Middle East Respiratory Syndrome (MERS) virus (Breban, Riou, & Fontanet, 2013). CT is the process of identifying and following up on people who may have come into contact with an infected person (Fisman, Khoo, & Tuite, 2014; Salathé et al., 2020; Webb et al., 2015). In Figure 6, we present an explicit illustration of a typical CT process flow diagram for inhibiting the spreading of COVID-19. Accurate modelling of CT requires precise information about the disease transmission pathways from each individual, and hence the network of contacts. Since contact tracing takes place as a process over the system of

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

interactions between hosts, it is natural to consider network-based models for this process. The CT model presented in this paper utilizes the peer-to-peer network architecture that centers on the query routing strategy for identifying infected individuals or potential carriers of COVID-19. Similarly, CT can in this case be supplemented with machine learning algorithms to develop technologies that alerts to higher levels of immediate risk and supports the rapid facilitation of practical interventions such sanitization equipment and deep cleans.

We present a peer-to-peer network-based framework for implementing an effective CT process in a smart city. In particular, the main contributions of this section is the design of an intelligent machine learning based CT strategy. The proposed framework has three main goals:

i. It takes into account the dynamic characteristics of the human network when performing CT process
ii. Communication and query referral between communities hosting suspected COVID-19persons is considered so efficient CT can be performed and tracked, and
iii. The strategy demonstrates scalability, making it suitable for adoption into other related pandemic cases that would require both human and financial resources on a large scale.

7.6.1 Peer-to-Peer Network Model

In smart spaces relative to smart city, networked system of places and individuals provide us with resourceful platforms that focus on global information sharing, searching, and replicated categorization of directories (Khambatti, Ryu, and Dasgupta, 2002; Khambatti, et al., 2003). In this instance, much of the interactions among interconnected peered network is directed at forming an overlay network that is based on cluster identifiers and using the overlay network to perform searches on a content addressable space. In the current proposed smart city framework, we extend

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

the aforementioned peer-to-peer network concepts to allow the notion of smart communities or connected society. Smart communities are useful in structuring the information storage space, discovering resources, and pruning the search space. It also aids in better dissemination of helpful information (Khambatti et al., 2003).

A peer-to-peer network consists of a large number of nodes that can join or leave the system at any time. For illustration, consider people living in smart spaces that comprises of households, neighborhoods, and communities. Here, each peer (or node) connects to a relatively small set of neighbors, who, in turn, are connected to larger peers (or smart spaces of community) such as the case of peer C in Figure 5. Similarly, the neighbors of peer A are peers' B, and C. Formation and discovery of peers belonging to some specific communities are significantly dependent on how peers declare and use their common "interests" (Khambatti et al., 2002). Firstly, we define attributes as a method of declaring an interest, and then we use these attributes to discover affected communities. Secondly, we also define another set of attributes as a method of identifying peers infected with COVID-19 viruses, and we use these attributes to trace all possible contacts made by an individual peer within the designated smart spaces of community. Therefore, two things are involved in the peer-to-peer network machine learning based CT model, first is the identification of communities where COVID-19 pandemic has occurred, and second is the identification of infected COVID-19 virus peers together with all their contacts.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 5: Regular connections amongst peers of human network

The adopted machine learning based smart city CT strategy conforms with the architectural perspective of the well-known Hop-Count Routing Indices (HRI) search mechanism for peer-to-peer network systems. The HRI algorithm discussed in this paper is assumed to have a deep learning features and thus our notion for machine learning based CT concept. Therefore, the hybrid machining learning based HRI algorithm is different from the classical HRI algorithm. In Figure 6, we present a model for the HRI contact tracing strategy that unveils the dynamic interactions of the susceptible and infected COVID-19populations based on smart space community or regional distribution of the COVID-19 virus. The machine learning based HRI algorithm incorporates the main features of contact tracing strategies, which include; the number of contacts per identified infectious case of COVID-19, the likelihood that a traced contact is infected, and the efficiency of the contact tracing process.

The initial identification of communities where peers could fall into based on their common interest and the community classification into either COVID-19 free or epicenter zones is vital to

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

the implementation of the proposed HRI-contact tracing algorithm. The identification process would help in pruning the search space during the CT process. Since it would only be the affected communities or epicenter zones that would be contact traced during the CT process.

7.6.2. Attribute-Based Community Networks

In this section, we present a community formation framework that determines the membership of a peer in one or more smart spaces or communities. For identification as part of a community, a peer is required to possess or declare attributes that are usually identified with members of that community. For example, in the smart spaces, several common interests and attributes exists in the formation of a community. These shared interests could come from the following: culture, religion, peer or age group, social activities, tribe, etcetera. For instance, a peer is expected to have originated from one ethnic group by birth. The ethnic group has its unique cultural backroad and probably its accepted religion or beliefs, as is the case with most tribal communities in some countries.

A similar use of peer-attributes in network-formulation is detailed in (Khambatti et al., 2003). (Khambatti et al., 2002) stated that attributes could be either explicitly provided by a peer or implicitly discovered from past queries. The authors further went on to classify peer attributes into three categories, namely, personal, claimed, and group. The main reason for these classifications is as a result of privacy and security concerns. A full set of peer attributes refers to a collection of personal peer attributes, and subsets of the peer's personal attributes are referred to as the peer claimed attributes. The personal attributes disclose complete attributes information about the peer, while the asserted attributes only disclose those attributes that the peer wishes to make public. The group attribute is a location or affiliation oriented and is needed to form a physical basis for communities. Table 3 provided examples of these three attribute classes.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 3: Peer attribute formation

Personal attributes	Claimed attributes	Group attributes
Phone number	Pet name	Tribe/Language
Date of birth	Religious denomination	culture
Full name	Occupation	State/province
Residential address	Gender /sex	Profession
Health status/ Blood group	Hobby	Religion
Next of Kin/Contact of relatives	Marital status	Race

The group-based attribute is relevant since it associates a peer to a specific community. In other words, every peer belongs to one pre-determined group and has a group attribute that identifies the peer as a member of that community.

7.6.3. Population Structure

The population structure for the peer-to-peer community network model presented in this paper follows a group hierarchical population structure in (Kasaie, Dowdy, & Kelton, 2013). The fabric covers the following settings;

i. The infected peer's immediate family members
ii. The infected peer's neighbors or the immediate community hosting the infected peer's together with the household members
iii. Finally, the communities that contain the family members, neighbors, and other extended neighboring communities.

The machine learning based CT process usually would first start from the COVID-19infected peer household, move on to the closest neighbors, before proceeding to neighbor communities for some possible contacts. The sizes of the mixing group peers (or community) are chosen arbitrarily in such a way that it would enable the modelling of different disease transmission routes associated with each mixing peers. The peers presented in Figure 6, only provides a realistic representation

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

of the population structure, which would aid the simulation implementation of the proposed framework.

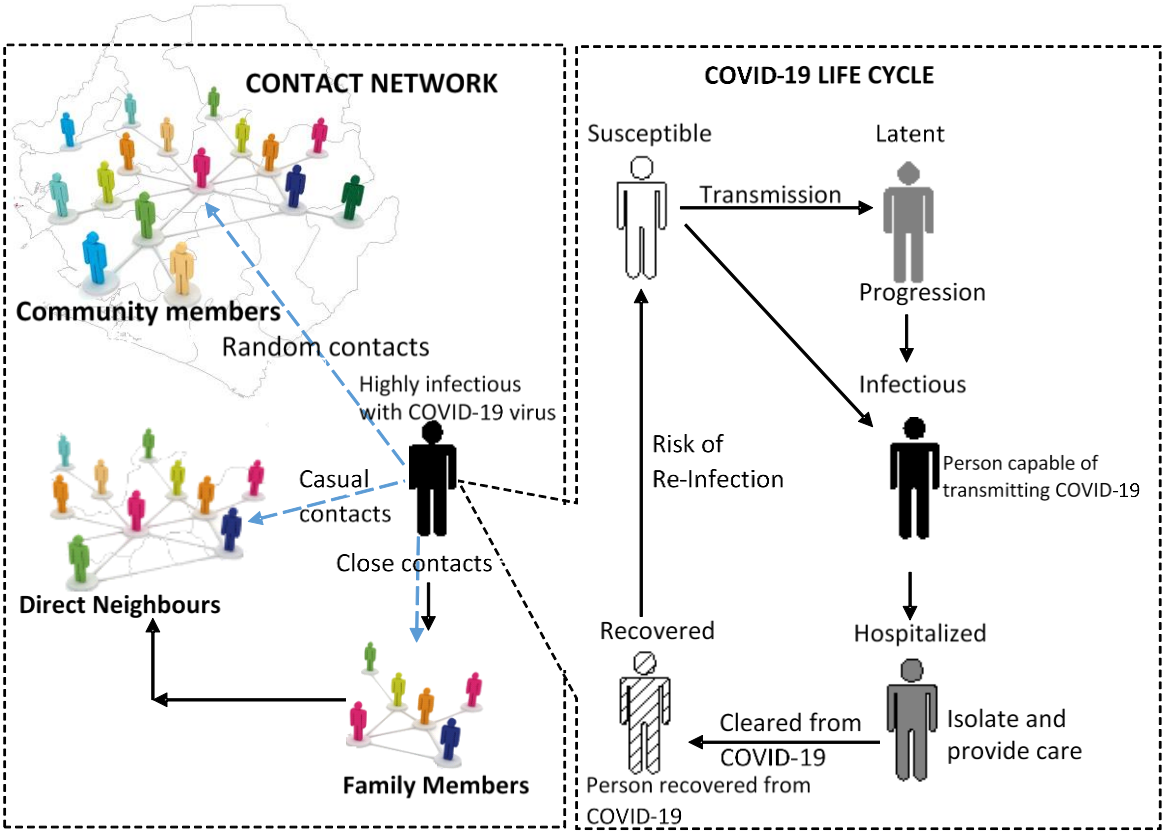

Figure 6: Proposed machine learning based CT model outlines for smart spaces in a smart city.

Note that the notion of smart spaces is used here to denote spaces occupied by households or family members, neighbourhood, community, etc. The COVID-19 life cycle considered in this paper follows the mathematical model sequence presented in section 4 above, which consists of the following compartments;

i. "*S*" the number of susceptible individuals, who can be infected
ii. "*E*" the number of exposed individuals, who have been infected but not yet infectious
iii. "*I*" the number of infectious cases in the community, who are capable of transmitting the disease

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

iv. "*H*" the number of individuals who are in the hospital, and

"*R*" the number of individuals removed from the chain of transmission (cured or dead and buried), respectively. A person is assumed to be born in full health and Susceptible to COVID-19. Upon successful transmission of the disease, the person enters the Latent or incubation state, which varies between 2 to 21 days, in severe forms of the disease, death may occur within an average of 14 days after onset of illness (Ivorra, Ferrández, Vela-Pérez, & Ramos, 2020; Kucharski et al, 2020).

7.6.4 Smart City Contact Tracing Network Structure

In modelling the smart city CT strategy, a three-layer contact network in connection with the population structure referred to as close, casual, and random contact is considered for the peer-to-peer hop-count routing indices algorithm. The three contact network structures represented in the model presented in Figure 7 illustrate the social relationships that could be established or exist between any peer with the rest of the peers' population. Following the three identified group-mixing, that is the family member group, neighborhood group, and community group mixing, we defined each of the contact types to fall under each of them. For instance, the close contact type will fall under the family member group mixing. In contrast, the causal contact type will fall under the neighborhood group mixing, and the random contact type will fall under the community group-mixing. Structuring the population sizes of each group mixing will differ from places to places or country to country, depending on some parameters that might be linked to either social, cultural, or religious beliefs. Table 4 shows three contact types for the three group mixing.

Table 4: Peer-to-peer contact types within a smart space

Peer Groups	Family	Neighbour	Community
Peer Contact Types	**Close Contact**	**Casual Contact**	**Random Contact**

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Population Group Mixing	Family members	Neighborhood playground Worship centers School Restaurants	Recreational centers Bus stop Conferences

In view of the possible individual network contact types within a confined smart space, the close contacts type defines the first level, which is the most recurrent interaction type among family members within a confined smart space. On the one hand the casual contacts which is the second level type of contact, represent a social relationship that often occurs among friends at a smart recreational center, sit out spaces, neighbors at a shopping malls, or children at learning institutions such high schools and universities. However, these types of contacts are less regular and cherished as compared to the first level type of contact. More so, it more restricted to a specific network of related individuals that are connected via a predefined smart spaces, including friends, coworkers, etc. (Classen et al., 1999; Raffalli, Sepkowitz, & Armstrong, 1996). For the proposed machine learning contact tracing model, we limit the domain of level one type of contact to each neighborhood's residents and assume a limited period of twelve months for the duration of the interactions. Finally, the third type of contact or interaction, which is the random contacts are used to represent interactions among individuals at smart spaces, for example, at smart bus or train stations, institutions, etc. This type of contact network has a very limited duration of interaction that would last for a period of one month. This type of contact usually accounts for the potential risk of COVID-19 transmission among non-related people in the whole smart community.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

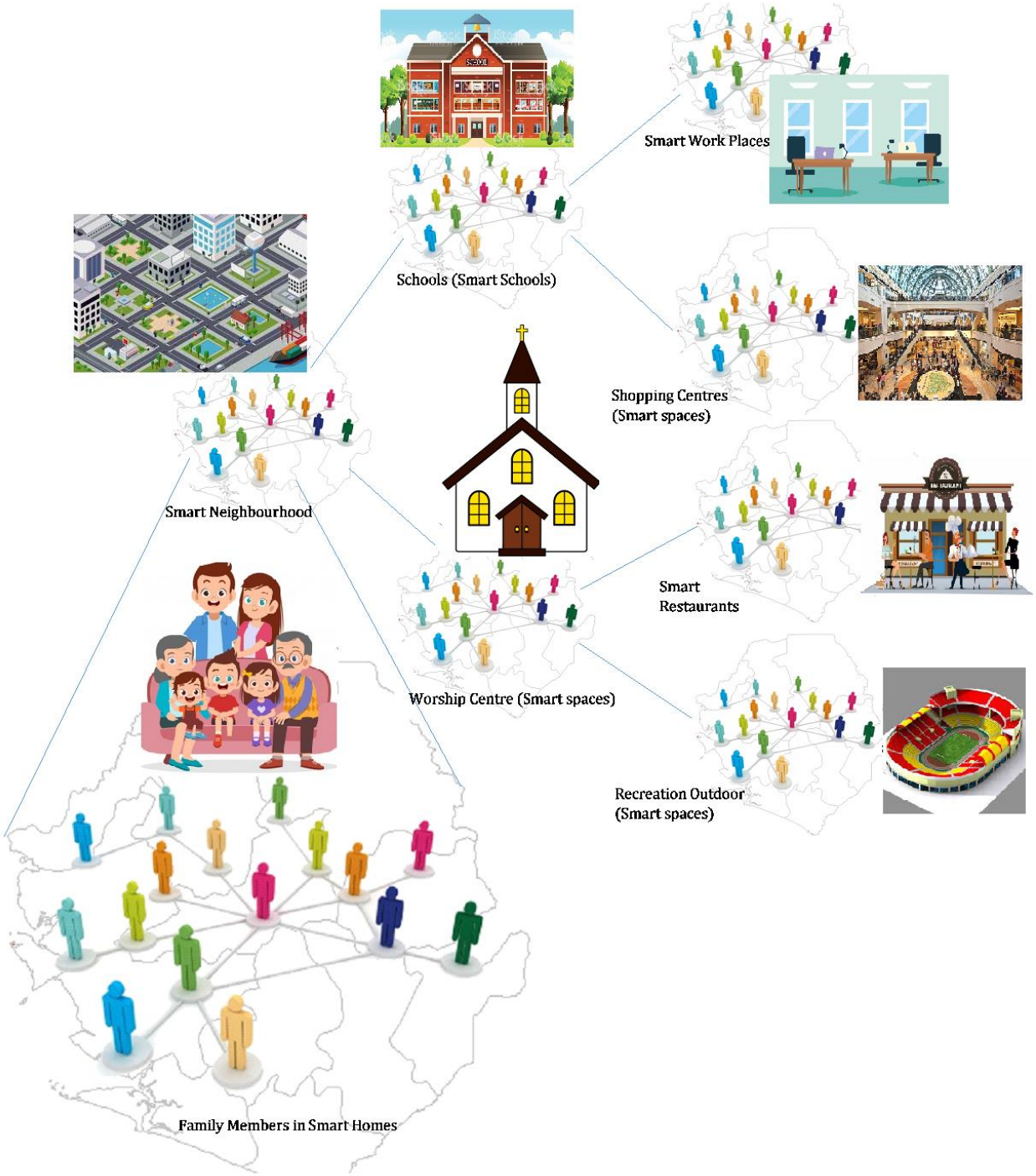

Figure 7: Cluster-based peer-to-peer smart city CT architecture

We anticipate the proposed machine learning based CT framework to be an essential tool in slowing down and stagger the spread of the current COVID-19 pandemic or any other pandemic. The structure, when implemented, can be used as an efficient and effective process of identifying

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

people who may have come into contact with an infected person and then amassing further information about the people who are potentially exposed. Indisputably, the new machine learning based CT model, if adequately implemented, can curtail the spread of infection in a vulnerable population. It can be employed by the government and policymakers to help people be aware of their infection status by being alerted. It also offers a diagnosis for a potentially vulnerable population, gain insights into the path of the spread of the disease to aid in further preventive measures. Finally, it allows the uninfected population to move around freely, provided the identified contacts have quarantined themselves.

8. Conclusions

In this paper, we propose a solution framework based on machine learning for integration in smart cities to fight COVID-19. The propose machine learning solution framework integrated different task to fight COVID-19 in the smart cities from different dimension such as predicting COVID-19 vaccine immunogenicity, detecting COVID-19 severity, predicting COVID-19 mortality, COVID-19 resource allocation, COVID-19 drugs discovery, COVID-19 contact tracing, social distance, detecting wearing of mask, detecting COVID-19 patients requiring ventilator, predicting COVID-19 patient that is beyond medical intervention and triage COVID-19 patients. The paper presented a comprehensive guide for implementing the machine learning framework in smart cities. We believed that the solution framework has the potential to automate the measures of fighting the COVID-19 pandemic in smart cities from multiple dimension to ease the fatigue of the healthcare workers due to very high number of COVID-19 patients requiring medical attention simultaneously and provide widespread access to quality healthcare system. In addition, we hope that the proposed smart city machine learning based framework for combatting COVID-19 will serve as essential guide to the research community in developing a more compartmentalized

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

forecasting and analyzing tools with prospect of mitigating the spread of the ongoing COVID-19 pandemic and the reoccurrences of any similar future pandemic disease. For future work, it will be interesting to see the real world application of the proposed framework to further investigates the model practicality and efficiency in smart city environments.

References

Abu-Mostafa, Y., Magdon-Ismail, M., & Lin, H. (2012). Learning from data vol. 4: AMLBook New York. *NY, USA*.

Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. (2020). Artificial intelligence and machine learning to fight COVID-19: American Physiological Society Bethesda, MD.

Angeletti, S., Benvenuto, D., Bianchi, M., Giovanetti, M., Pascarella, S., & Ciccozzi, M. (2020). COVID-2019: the role of the nsp2 and nsp3 in its pathogenesis. *Journal of medical virology*.

Allam, Z., & Jones, D. S. (2020, March). On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. In *Healthcare* (Vol. 8, No. 1, p. 46). Multidisciplinary Digital Publishing Institute.

Arkes, H. R. (2001). Overconfidence in judgmental forecasting *Principles of forecasting* (pp. 495-515): Springer.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Al-Turjman, F. (2019). 5G-enabled devices and smart-spaces in social-IoT: an overview. *Future Generation Computer Systems*, *92*, 732-744.

Bassetti, M., Vena, A., & Giacobbe, D. R. (2020). The novel Chinese coronavirus (2019-nCoV) infections: Challenges for fighting the storm. *European journal of clinical investigation*, *50*(3), e13209.

Balduini, M., Brambilla, M., Della Valle, E., Marazzi, C., Arabghalizi, T., Rahdari, B., & Vescovi, M. (2019). Models and practices in urban data science at scale. *Big Data Research*, *17*, 66-84.

Begun, M., Newall, A. T., Marks, G. B., & Wood, J. G. (2013). Contact tracing of tuberculosis: a systematic review of transmission modelling studies. *Plos one*, *8*(9).

Breban, R., Riou, J., & Fontanet, A. (2013). Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *The Lancet*, *382*(9893), 694-699.

Brown, P., Jha, P., & Consortium, C. C. M. (2020). Mortality from COVID-19 in 12 countries and 6 states of the United States. *medRxiv*.

Browne, C., Gulbudak, H., & Webb, G. (2015). Modeling contact tracing in outbreaks with application to Ebola. Journal of theoretical biology, 384, 33-49.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Catriona Manville, Gavin Cochrane, Jonathan Cave, Jeremy Millard, Jimmy Kevin Pederson, Rasmus Kåre Thaarup, . . . Bas Kotterink. (2014). Mapping smart cities in the EU. Retrieved 4/11/2016, 2016, from http://www.smartcities.at/assets/Publikationen/Weitere-Publikationen-zum-Thema/mappingsmartcities.pdf

Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the Total Environment*, 138882.

Chenthamarakshan, V., Das, P., Padhi, I., Strobelt, H., Lim, K. W., Hoover, B., . . . Mojsilovic, A. (2020). Target-specific and selective drug design for covid-19 using deep generative models. *arXiv preprint arXiv:2004.01215*.

Classen, C. N., Warren, R., Richardson, M., Hauman, J. H., Gie, R. P., Ellis, J. H., . . . Beyers, N. (1999). Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. *Thorax, 54*(2), 136-140.

Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature reviews Microbiology, 17*(3), 181-192.

Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920-1930.

Dingli, A. (2020). How technology is helping China crush the Coronavirus. from https://www.businesstoday.com.mt/people/people/799/how_technology_is_helping_china_crush_the_coronavirus#.XsAGX8Ao9Pa

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Duffy, S. (2018). Why are RNA virus mutation rates so damn high?. *PLoS biology*, *16*(8), e3000003.

ECDC. (2020). European Centre for Disease Prevention and Control from https://www.ecdc.europa.eu/en/2019-ncov-background-disease

Ekins, S., Freundlich, J. S., Clark, A. M., Anantpadma, M., Davey, R. A., & Madrid, P. (2015). Machine learning models identify molecules active against the Ebola virus in vitro. *F1000Research*, *4*.

Eng, C. L., Tong, J. C., & Tan, T. W. (2014). Predicting host tropism of influenza A virus proteins using random forest. *BMC medical genomics*, *7*(S3), S1.

Fisman, D., Khoo, E., & Tuite, A. (2014). Early epidemic dynamics of the West African 2014 Ebola outbreak: estimates derived with a simple two-parameter model. *PLoS currents*, *6*.

Fuyan, L. (2005). *An attribute selection approach and its application.* Paper presented at the 2005 International Conference on Neural Networks and Brain.

Gao, L., Jiang, D., Wen, X.-s., Cheng, X.-c., Sun, M., He, B., . . . Qin, S. (2020). Prognostic value of NT-proBNP in patients with severe COVID-19. *Respiratory research*, *21*(1), 1-7.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Ge, Y., Tian, T., Huang, S., Wan, F., Li, J., Li, S., . . . Yuan, E. (2020). A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *bioRxiv*.

Gonzalez-Dias, P., Lee, E. K., Sorgi, S., de Lima, D. S., Urbanski, A. H., Silveira, E. L., & Nakaya, H. I. (2020). Methods for predicting vaccine immunogenicity and reactogenicity. *Human Vaccines & Immunotherapeutics, 16*(2), 269-276.

Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., . . . Siegel, E. (2020). Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. *arXiv preprint arXiv:2003.05037*.

Grijalva, C. G., Goeyvaerts, N., Verastegui, H., Edwards, K. M., Gil, A. I., Lanata, C. F., & Hens, N. (2015). A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. PloS one, 10(3).

Guan, W.-j., Ni, Z.-y., Hu, Y., Liang, W.-h., Ou, C.-q., He, J.-x., . . . Hui, D. S. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine, 382*(18), 1708-1720.

Hao, P., Zhong, W., Song, S., Fan, S., & Li, X. (2020). Is SARS-CoV-2 originated from laboratory? A rebuttal to the claim of formation via laboratory recombination. *Emerging microbes & infections, 9*(1), 545-547.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*: Springer Science & Business Media.

Ivorra, B., Ferrández, M. R., Vela-Pérez, M., & Ramos, A. (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation*, 105303.

Kampf, G., Todt, D., Pfaender, S., & Steinmann, E. (2020). Persistence of coronaviruses on inanimate surfaces and its inactivation with biocidal agents. *Journal of Hospital Infection*.

Kasaie, P., Dowdy, D. W., & Kelton, W. D. (2013). *An agent-based simulation of a tuberculosis epidemic: understanding the timing of transmission.* Paper presented at the 2013 Winter Simulations Conference (WSC).

Khambatti, M., Ryu, K. D., & Dasgupta, P. (2002). *Peer-to-peer Communities: Formation and Discovery*. Paper presented at the IASTED PDCS.

Khambatti, M., Ryu, K. D., & Dasgupta, P. (2003). *Structuring peer-to-peer networks using interest-based communities.* Paper presented at the International Workshop On Databases, Information Systems, and Peer-to-Peer Computing.

Lewis, D. (2020). Is the coronavirus airborne? Experts can't agree. *nature*, *580*(7802), 175.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., . . . Wong, J. Y. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England journal of medicine*.

Li, X., Geng, M., Peng, Y., Meng, L., & Lu, S. (2020). Molecular immune pathogenesis and diagnosis of COVID-19. Journal of Pharmaceutical Analysis.

Lin, L., Lu, L., Cao, W., & Li, T. (2020). Hypothesis for potential pathogenesis of SARS-CoV-2 infection–a review of immune changes in patients with viral pneumonia. *Emerging microbes & infections*, *9*(1), 727-732.

Lippi, G., & Plebani, M. (2020). Laboratory abnormalities in patients with COVID-2019 infection. *Clinical Chemistry and Laboratory Medicine (CCLM)*, *1*(ahead-of-print).

Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., . . . Jervey, S. (2020). Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases: ACS Publications.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., ... & Bi, Y. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565-574.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Metsky, H. C., Freije, C. A., Kosoko-Thoroddsen, T.-S. F., Sabeti, P. C., & Myhrvold, C. (2020). CRISPR-based COVID-19 surveillance using a genomically-comprehensive machine learning approach. *bioRxiv*.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). Foundations of Machine Learning. Adaptive computation and machine learning. *MIT Press*, *31*, 32.

Narin, A., Kaya, C., & Pamuk, Z. (2020). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849*.

Ndiaye, B. M., Tendeng, L., & Seck, D. (2020). Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting. *arXiv preprint arXiv:2004.01574*.

Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, *105*, 233-261.

Obeidat, S. (2020). How Artificial Intelligence Is Helping Fight The COVID-19 Pandemic from https://www.entrepreneur.com/article/348368

Oyelade, O. N., & Ezugwu, A. E. (2020). COVID19: A Natural Language Processing and Ontology Oriented Temporal Case-Based Framework for Early Detection and Diagnosis of Novel Coronavirus. Preprints, doi: 10.20944/preprints202005.0171.v1.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Pandey, R., Gautam, V., Bhagat, K., & Sethi, T. (2020). A machine learning application for raising wash awareness in the times of covid-19 pandemic. *arXiv preprint arXiv:2003.07074*.

Patankar, S. (2020). Deep learning-based computational drug discovery to inhibit the RNA Dependent RNA Polymerase: application to SARS-CoV and COVID-19.

Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications, 97*, 205-227.

Pourhomayoun, M., & Shakibi, M. (2020). Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making. *medRxiv*.

Prompetchara, E., Ketloy, C., & Palaga, T. (2020). Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pac J Allergy Immunol, 38*(1), 1-9.

Raffalli, J., Sepkowitz, K. A., & Armstrong, D. (1996). Community-based outbreaks of tuberculosis. *Archives of internal medicine, 156*(10), 1053-1060.

Randhawa, G. S., Soltysiak, M. P., El Roz, H., de Souza, C. P., 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of autoimmunity*, 102433.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., . . . Guggemos, W. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England journal of medicine*, *382*(10), 970-971.

Salathé, M., Althaus, C. L., Neher, R., Stringhini, S., Hodcroft, E., Fellay, J., . . . Wilder-Smith, A. (2020). COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss medical weekly*, *150*(11-12), w20225-.

Shahrokni, H., Van der Heijde, B., Lazarevic, D., & Brandt, N. (2014). *Big data GIS analytics towards efficient waste management in Stockholm.* Paper presented at the Proceedings of the 2014 conference ICT for Sustainability.

Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., . . . Melino, G. (2020). COVID-19 infection: the perspectives on immune responses: Nature Publishing Group.

Smart, E.-C., & Cooperation, G. C. (2014). Comparative study of smart cities in Europe and China. *Current Chinese Economic Report Series, Springer*.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Spectrum, I. (2020). AI Can Help Hospitals Triage COVID-19 Patients. from https://spectrum.ieee.org/the-human-os/artificial-intelligence/medical-ai

Sullivan, F. (2008). Strategic Opportunity Analysis of the Global Smart City Market Smart City Market is Likely to be Worth a Cumulative $1.565 Trillion by 2020. *"The Growth Partnership Company", vol.19 pg. 1-19*

Taylor, S. J., & Letham, B. (2017). prophet: Automatic Forecasting Procedure, R package version 0.2.

Ting, D. S. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). Digital technology and COVID-19. *Nature medicine*, *26*(4), 459-461.

Tsur, O., Davidov, D., & Rappoport, A. (2010). *ICWSM—a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews.* Paper presented at the fourth international AAAI conference on weblogs and social media.

van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., . . . Boshier, F. A. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*, 104351.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

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

Webb, G., Browne, C., Huo, X., Seydi, O., Seydi, M., & Magal, P. (2015). A model of the 2014 Ebola epidemic in West Africa with contact tracing. *PLoS currents, 7*.

WHO. (2019). MINIMUM REQUIREMENTSfor infection prevention and control programmes. from https://www.who.int/infection-prevention/publications/MinReq-Manual_2019.pdf?ua=1

WHO. (2020). Coronavirus disease (COVID-2019) situation reports. Geneva. from https://www.who.int/emergencies/diseases/novel-coronavirus-2019?gclid=Cj0KCQjwnv71BRCOARIsAIkxW9HqYxLRYA3R2Sz_jVGDPecH0r-RV5Fe7dd4XAO5ObXn9kYYivrYKnIaApArEALw_wcB

Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. Jama, 323(13), 1239-1242.

Xiao, S. Y., Wu, Y., & Liu, H. (2020). Evolving status of the 2019 novel coronavirus infection: proposal of conventional serologic assays for disease diagnosis and infection monitoring. *Journal of medical virology*, *92*(5), 464-467.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105577; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Yan, L., Zhang, H.-T., Xiao, Y., Wang, M., Sun, C., Liang, J., . . . Xiao, Y. (2020). Prediction of survival for severe Covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan. *medRxiv*.

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., . . . Mai, Z. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease, 12*(3), 165.

Yu, S., Wang, C., Liu, K., & Zomaya, A. Y. (2016). Editorial for IEEE access special section on theoretical foundations for big data applications: challenges and opportunities. *IEEE Access, 4*, 5730-5732.

Zhang, L., Ai, H.-X., Li, S.-M., Qi, M.-Y., Zhao, J., Zhao, Q., & Liu, H.-S. (2017). Virtual screening approach to identifying influenza virus neuraminidase inhibitors using molecular docking combined with machine-learning-based scoring function. *Oncotarget, 8*(47), 83142.

Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D. S., . . . Orekhov, P. (2020). Potential COVID-2019 3c-like protease inhibitors designed using generative deep learning approaches. *Insilico Medicine Hong Kong Ltd A, 307*, E1.