ANALYSIS OF MACHINE LEARNING BASED COVID RISK ESTIMATION

Akansha Chokse[1], Nikhil Patearia [2],
MTech.Scholar[1], Prof. [3]
Department of Computer Science Engineering[1,2,]
Jai Narain College of Technology, Bhopal, India

Abstract

In this study an attempt was made to determine the most general equation for any general patient. This system can help hospitals, medical facilities, and caregivers decide who needs to get attention first before other patients, triage patients when the system is overwhelmed by overcrowding, and also eliminate delays in providing the necessary care. Our study is very helpful to estimate COVID-19 patients' health conditions more effectively and accurately during the pandemic. Our estimated score very helpful to hospitals and medical facilities minimize decision-making time. Our study is to provide an accurate and reliable tool to help medical decision-making and triage COVID-19 patients. We provide automotive tools that reduce the medical staff time and effort to examine COVID-19 patient's condition.

Keywords - *Handling missing values, Eliminating redundant useless, Data elements, Machine learning algorithms, Predict the medical condition*

I INTRODUCTION

The flow flare-up of the novel Covid SARSCoV-2 (Covid infection 2019; already 2019- nCoV), epi-focused in Hubei Province of the People's. The Republic of China has spread to numerous different nations. On 30. January 2020, the WHO Emergency Committee proclaimed a worldwide wellbeing crisis dependent on developing case notice rates in Chinese and global areas. The case discovery rate is changing every day and can be followed in practically ongoing on the site given by Johns Hopkins University [1] and different gatherings. As of the middle of February 2020, China bears the enormous weight of bleakness and mortality, though the occurrence in other Asian nations, in Europe and North America, remains low up until this point. Coronaviruses are a group of related RNA viruses that cause diseases in mammals and birds. In humans and birds, they cause respiratory tract infections that can range from mild to lethal. Mild illnesses in humans include some cases of the common cold (which is also caused by other viruses, predominantly rhinoviruses), while more lethal varieties can cause SARS, MERS, and COVID-19. In cows and pigs they cause diarrhea, while in mice they cause hepatitis and encephalomyelitis. Covids are wrapped, positive single-abandoned enormous RNA infections that taint people, yet additionally a wide scope of creatures. Covids were first portrayed in 1966 by Tyrell and Bynoe, who developed the infections from patients with basic colds [2] The Pandemic has a long history; however, the actual term is yet to be characterized by numerous clinical writings. There have been various critical pandemics recorded in mankind's set of experiences where pandemic-related emergencies contrarily affect wellbeing, economies, and surprisingly public safety around the world.

II Impact of SARS-CoV-2 on Extrapulmonary Organ Systems

Albeit the respiratory framework is the chief objective for SARS-CoV-2 as depicted above, it can influence other significant organ frameworks like the gastrointestinal lot (GI), hepatobiliary, cardiovascular, renal, and focal sensory system. SARS-CoV-2–instigated organ brokenness, all in all, is conceivably clarified by it is possible that one or a blend of the proposed instruments like direct popular poisonousness, ischemic injury brought about by vasculitis, apoplexy, or thrombo-irritation, resistant dysregulation, and renin-angiotensin-aldosterone framework (RAAS) dysregulation.[49]

Cardiovascular framework (CVS): Although the specific system of heart inclusion in COVID-19 is obscure, it is likely multifactorial. ACE2 receptors are likewise shown by myocardial cells involving direct cytotoxicity by the SARS-CoV-2 on the myocardium prompting myocarditis. Proinflammatory cytokines, for example, IL-6 can likewise prompt vascular aggravation, myocarditis, and heart arrhythmias.

Hematological: SARS-CoV-2 significantly affects the hematological and hemostatic framework. The instrument of leukopenia, quite possibly the most widely recognized research facility irregularities experienced in COVID-19, is obscure. A few theories have been proposed that incorporate ACE 2 intervened lymphocyte obliteration by direct attack by the infection, lymphocyte apoptosis because of proinflammatory cytokines, and

conceivable intrusion of the infection of the lymphatic organs.[55] **Focal Nervous System (CNS):** There is arising proof of ACE2 receptors in human and mouse cerebrums, ensnaring the likely contamination of the mind by SARS-CoV-2.[58] The potential courses by which SARS-CoV-2 can attack the focal sensory system are transsynaptic move across tainted neurons through the olfactory nerve, vascular endothelial cell disease, or relocation of leukocytes across the blood-mind barrier.[59]

Gastrointestinal (GI) Tract: The pathogenesis of GI signs of COVID-19 is obscure and is likely viewed as multifactorial because of a few potential instruments that incorporate the immediate ACE 2-intervened viral cytotoxicity of the intestinal mucosa, cytokine-incited irritation, gut dysbiosis, and vascular abnormalities.[60]

Hepatobiliary: Although the pathogenesis of liver injury in COVID-19 patients is obscure, hepatic injury in COVID-19 is likely multifactorial and is clarified by numerous systems alone or in mix that incorporates ACE-2-interceded viral replication in the liver, direct infection intervened harm, hypoxic or ischemic injury, insusceptible intervened incendiary reaction, drug-prompted liver injury (DILI), or deteriorating of previous liver sickness.

Renal: The pathogenesis of COVID-19 related kidney injury is obscure and is likely multifactorial clarified by a solitary or a blend of numerous variables like direct cytotoxic injury from the infection, irregularity in the RAAS, related cytokine-initiated hyperinflammatory state, microvascular injury, and the prothrombotic state related with COVID-19. Different factors like related hypovolemia, expected nephrotoxic specialists, and nosocomial sepsis can likewise possibly add to kidney injury.[61]

Lungs: A multicenter examination of lung tissue got during dissections of patients who tried positive for COVID-19 exhibited ordinary diffuse alveolar harm highlights in 87% of cases. Furthermore, there was an incessant presence of type II pneumocyte hyperplasia, aviation route irritation, and hyaline layers in alveolar zones. 42% of patients were noted to have enormous vessel thrombi, platelet (CD61 positive), and additionally fibrin microthrombi were available in 84% of cases.[64]

Brain A solitary place histopathological investigation of cerebrum examples got from 18 patients who capitulated to COVID-19 showed intense hypoxic injury in the entirety of patients' frontal cortex and cerebellum. Strikingly, no highlights of encephalitis or other explicit cerebrum changes were seen. Also, immunohistochemical investigation of cerebrum tissue didn't show cytoplasmic viral staining.[65]

Heart: Analysis of cardiovascular tissue from 39 examination instances of patients who tried positive for SARS-CoV-2 showed the presence of SARS-CoV-2 viral genome inside the myocardium.[66]

Kidney: Histopathology examination of kidney examples got from dissections of 26 patients with affirmed COVID-19 showed indications of diffuse proximal rounded injury with loss of brush line, non-isometric vacuolar degeneration, and rot. Furthermore, electron microscopy showed groups of Covid like particles with spikes in the rounded epithelium and podocytes.[67]

GI Tract: Endoscopic examples showed positive staining of the viral nucleocapsid protein in the gastric, duodenal, and rectal epithelium cytoplasm. Various invading plasma cells and lymphocytes with interstitial edema were found in the lamina propria of the stomach, duodenum, and rectum.[68]

Liver: An imminent single-focus clinicopathologic case arrangement study including the posthumous histopathological test of significant organs of 11 expired patients with COVID-19 announced hepatic steatosis discoveries in all patients. The liver examples of 73% of patients exhibited constant clog. Various types of hepatocyte corruption were noted in 4 patients, and 70% showed nodular proliferation.

III METHODOLOGY

The proposed system includes a set of algorithms for pre-processing the data to extract new features, handling missing values, eliminating redundant and useless data elements, and selecting the most informative features. After pre-processing the data, we use machine learning algorithms to develop a predictive model to classify the data, predict the medical condition, and calculate the probability and risk of mortality.

Dataset-we utilized a dataset of more than 117,000 research facility affirmed COVID-19 patients from 76 nations all throughout the planet including both male and female patients with a normal time of 56.6 [3]. The infection affirmed by location of infection nucleic corrosive [3]. The first dataset contained 32 information components from every quiet, including segment and physiological information. At the information cleaning stage, we eliminated pointless and excess information components, for example, information source, administrator id, and administrator name. Then, at that point, Data attribution strategies were utilized to deal with missing qualities. Subsequent to dissecting the information, we discovered that 74% of patients were recuperated from COVID-19. To have an exact and fair model, we ensured that our dataset is adjusted. A offset dataset with equivalent perceptions for both recuperated also, expired patients was made to prepare and test our model. The information perceptions (patients) in the preparation dataset have been chosen haphazardly and they are totally isolated from the testing information. Figure 1 shows a significant level design of our framework.

High-Level System Architecture

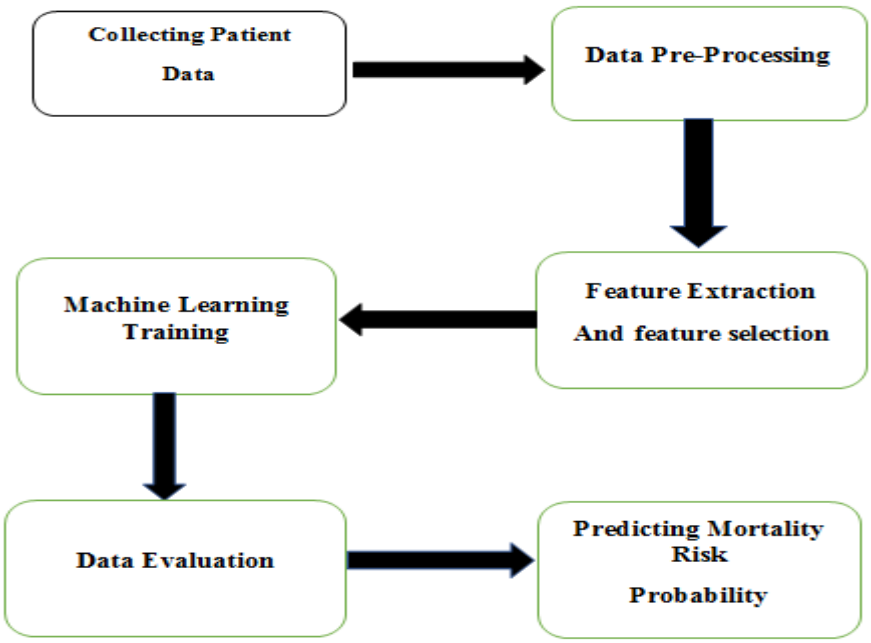

Figure 3.1: High-Level System Architecture

Feature selection-The result name contained various qualities for the patient's wellbeing status. We considered patient that released from emergency clinic or patients in stable circumstance without any manifestations as recuperated patients. An aggregate of 80 highlights were removed from side effects and specialists' clinical notes about the patient's wellbeing status We likewise removed extra 32 highlights from patient's segment and physiological information, made it to add up to 112 highlights. We talked with a clinical group to ensure that the best highlights are separated and chosen. The subsequent stage is including choice. The main role of include choice is to track down the most educational highlights and dispose of excess information to lessen the dimensionality and intricacy of the model [11]. We utilized univariate and multivariate channel technique and covering strategy to rank the includes and select the best element subset [11]. Figure exhibits the means of channel and covering technique that we utilized for include determination. Channel techniques are well known (particularly for huge datasets) since they are generally extremely quick and significantly less computationally serious than covering techniques. Channel strategies utilize a particular measurement to score every individual component (or on the other hand a subset of highlights together). The most mainstream measurements utilized in channel strategies incorporate connection coefficient, Fisher score, shared data, entropy and consistency and chisquare boundaries [11]. In the wake of applying diverse channel and covering techniques, we picked 42 highlights out of 112 highlights. Our last list of capabilities incorporates segment highlights like age, sex, area, country, age, travel history, general clinical data such as comorbidities (diabetes, cardiovascular sickness, …), and additionally persistent side effects, for example, chest torment, chills, colds, conjunctivitis, hack, looseness of the bowels, inconvenience, tipsiness, dry hack, dyspnea, emesis, expectoration, eye disturbance, weariness, heave, cerebral pain, injuries on chest radiographs, little sputum, disquietude, muscle torment, myalgia, obnubilation, pneumonia, myelofibrosis, respiratory indications, rhinorrhea, lethargy, sputum, transient weariness, shortcoming, and so forth Figure 3 shows the Correlation Heatmap for dataset highlights. Figure 3-(a) shows the relationship between's highlights what's more, the result for example mortality hazard, and Figure 3-(b) shows the relationship between's highlights. As Figure 3-(a) delineates, a few highlights like age and ongoing infections (comorbidities) were the top highlights with high relationship to the patient's mortality hazard.

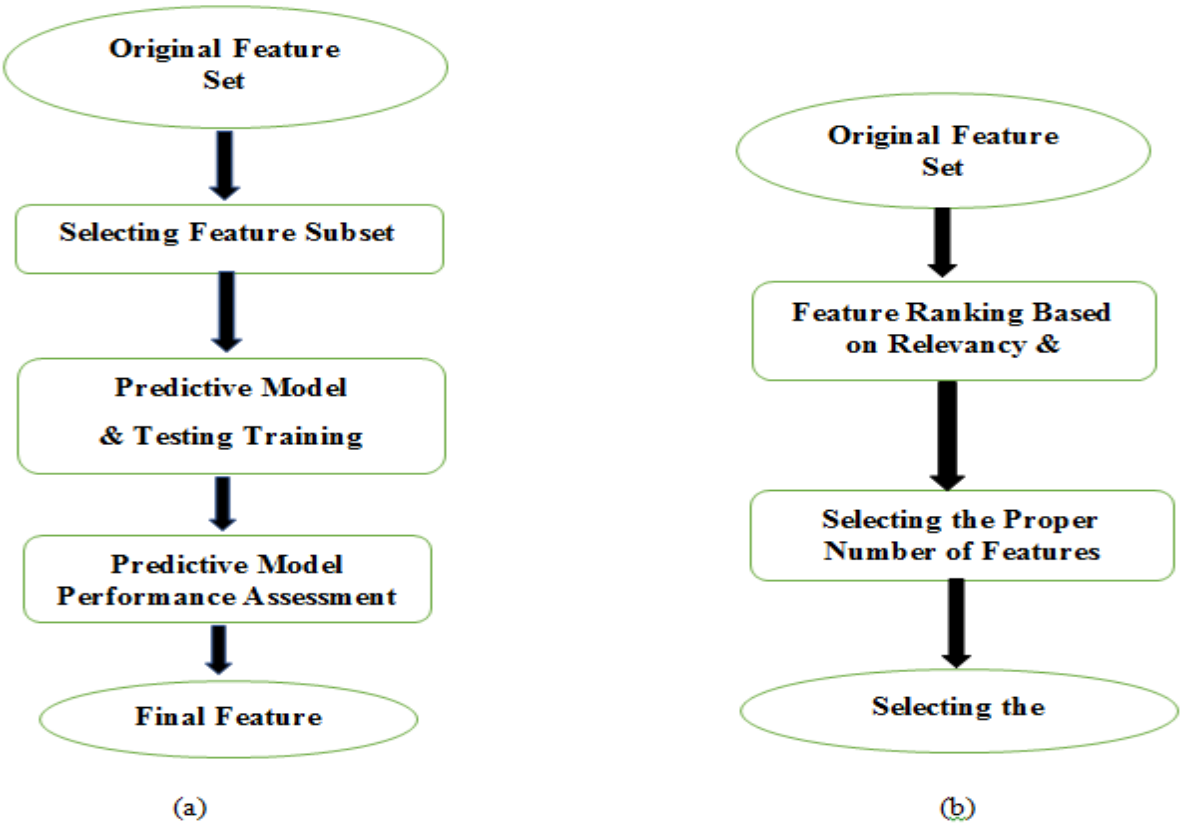

Figure 3.2. Feature Selection:(a) Wrapper method, (b) Filter Method

Linear Regression Algorithms

As per Wikipedia, regression analysis is a measurable methodology for evaluating the connections among variables. It fuses numerous methods for displaying and examining a few variables, when the center is on the relationship between a ward variable and one or more autonomous variables. The focus of estimation is a capacity of the autonomous variables called the regression capacity.

Where,

Xa = cost parameter for patients with comorbidities = (percentage of comorbidities chance * cost parameter for patients without comorbidities)

Xb = cost parameter for patients with hypertension = (percentage of hypertension chance * cost parameter for patients with hypertension)

Similarly apply all feature.

Assessment

We utilized 10-overlay arbitrary cross-approval (with no cover, with no substitution) to assess the created model. We determined the Overall Accuracy for all machine learning calculations to think about. Additionally, we produced Receiver Working Characteristic (ROC) bends for each calculation, what's more, determined the Area Under Curve (AUC) and Confusion Network. Once more, we ensured that there is no cover (no regular patient) among preparing and testing datasets at any level. The following area will give the outcomes and execution of the created framework.

Flow Chart of Linear Regression

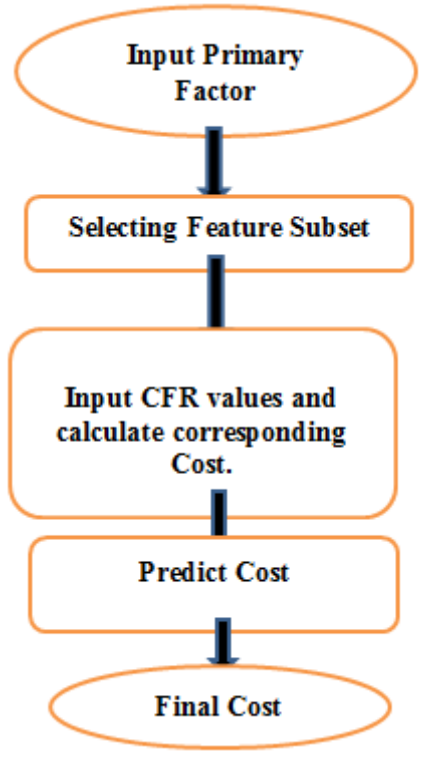

Figure3.3. Feature Selection:(a) Wrapper method

IV SIMULATION ENVIRONMENT

Python may be a high-level, taken, interactive and object-oriented scripting language. Python is intended to be extremely legible. It uses English keywords often whereas different languages use punctuation, and it's fewer grammar constructions than different languages. Python was developed by Guido van Rossum within the late eighties and early nineties at the National analysis Institute for arithmetic and engineering science within the Kingdom of The Netherlands.

- Python is taken − Python is processed at runtime by the interpreter. you are doing not ought to compile your program before executing it. this is often just like PERL and PHP.
- Python is Interactive − you'll be able to sit at a Python prompt and move with the interpreter on to write your programs.
- Python is Object-Oriented − Python supports Object-Oriented vogue or technique of programming that encapsulates code at intervals objects.
- Python may be a Beginner's Language − Python may be a nice language for the beginner-level programmers and supports the event of a good varies of applications from easy text process to WWW browsers to games.

History of Python

- Python was developed by Guido van Rossum within the late eighties and early nineties at the National analysis Institute for arithmetic and engineering science within the Kingdom of The Netherlands.
- Python comes from several different languages, as well as ABC's, Modula-3, C, C++, Algol-68, SmallTalk, and UNIX shell and different scripting languages.
- Python is proprietary. Like Perl, Python ASCII text file is currently obtainable beneath the wildebeest General Public License (GPL).

Python with Anaconda

Anaconda is an open source distribution of the Python and R programming languages and it is used in data science, machine learning, deep learning-related applications aiming at simplifying package management and deployment.Anaconda Distribution is used by over 7 million users, and it includes more than 300 data science packages suitable for Windows, Linux, and MacOS.

Anaconda Download

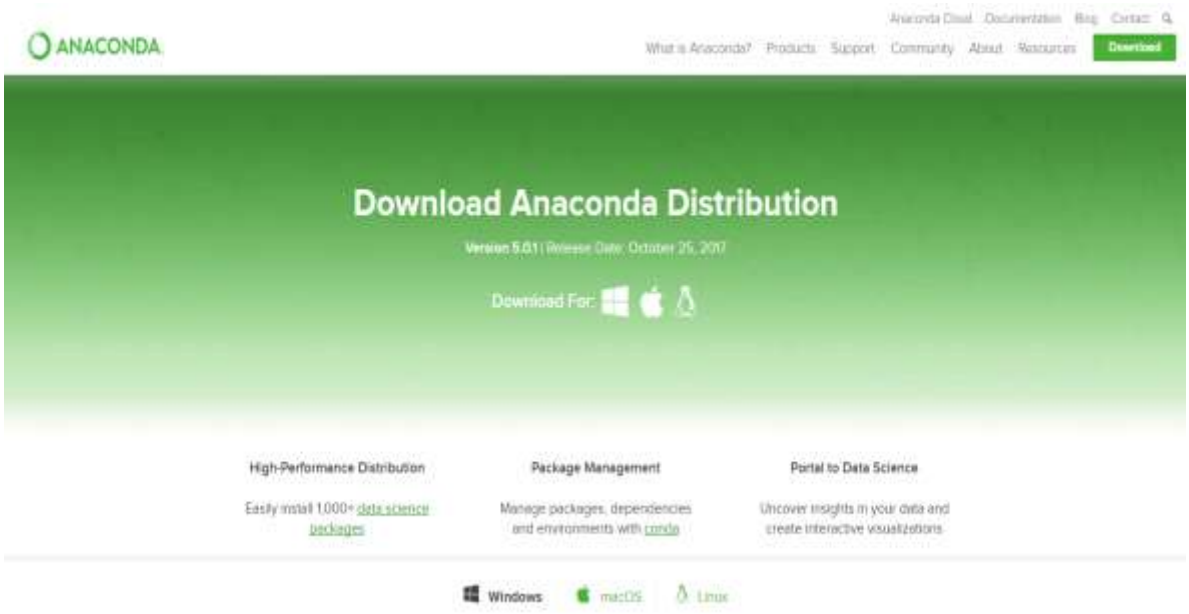

Figure 4.1: Anaconda Download Page

- Go to Download page and select option according to our OS and select.
- Select windows version(windows 7/8/10) and select OS Architecture (32bit/64bit).
- Select and download and then click on download file.

Figure 4.2: Anaconda Installation Page-1

- Click next
- Next page they show two option two path environment variable.
- First option for advance setting they give permission to set path to user self-according to his requirement.
- 2nd option is by default option anaconda self-set path and by default set all Path and environment.

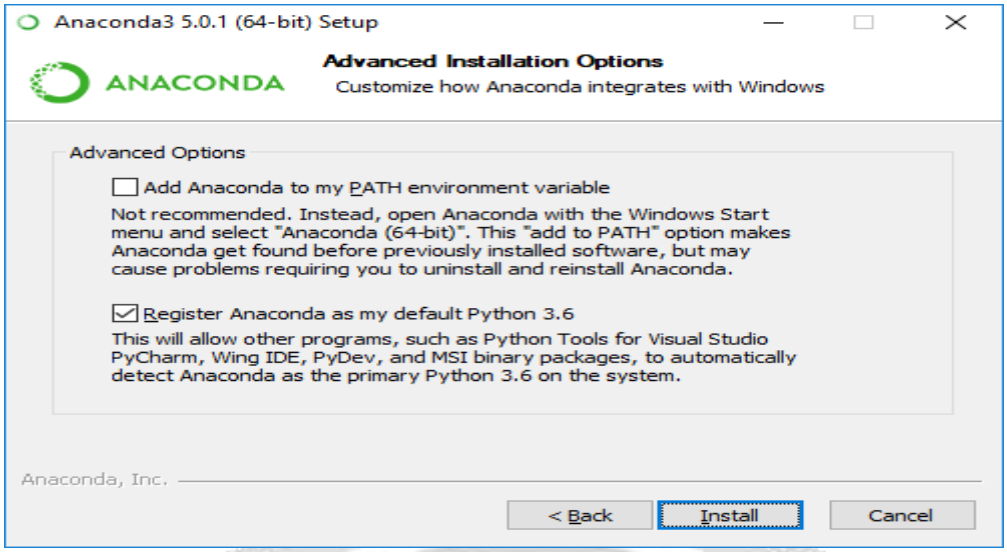

Figure 4.3: Anaconda Installation Page-2

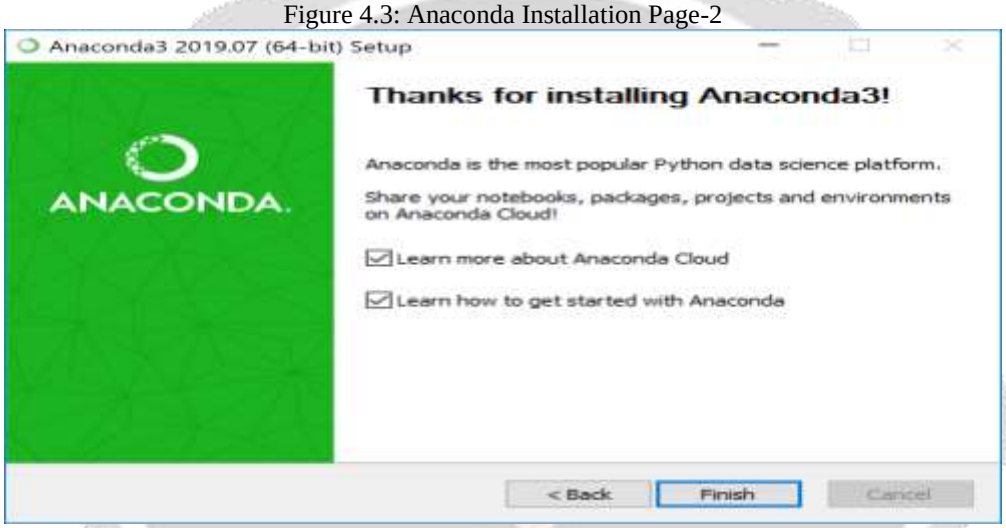

Figure 4.4: Anaconda Installation complete

- Start Search Bar and Search Anaconda.
- Click search Jupyter Notebook (Anaconda3) and click

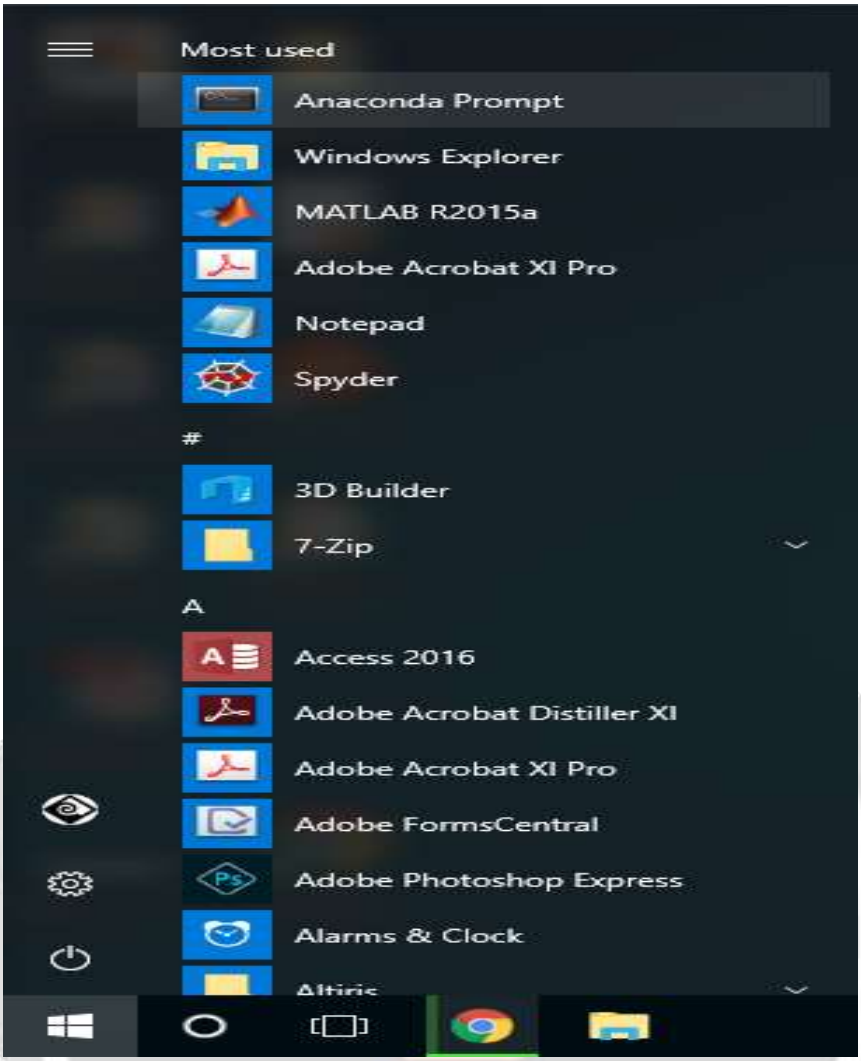

Figure 4.5: Start-up Window

- Click search Jupyter Notebook (Anaconda3) and click

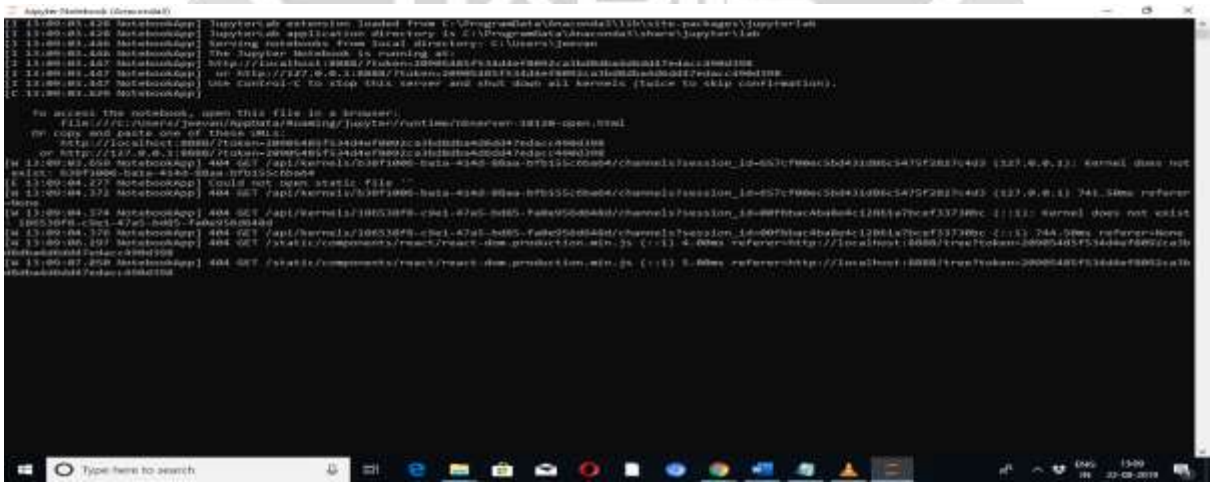

Figure 4.6: Start-up WindowJupyter Notebook

- Launch window of Jupyter

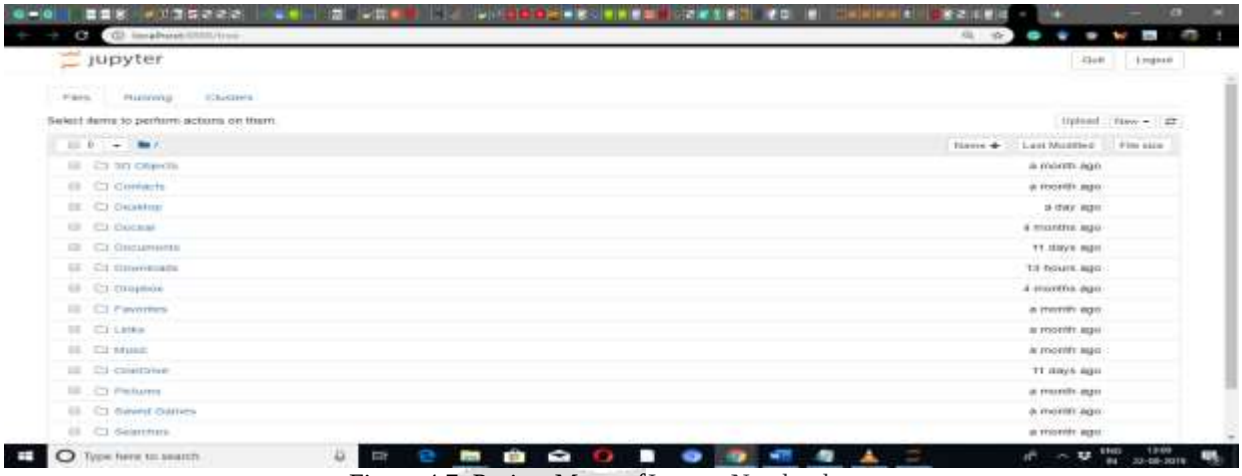

Figure 4.7: Project Menu ofJupyter Notebook

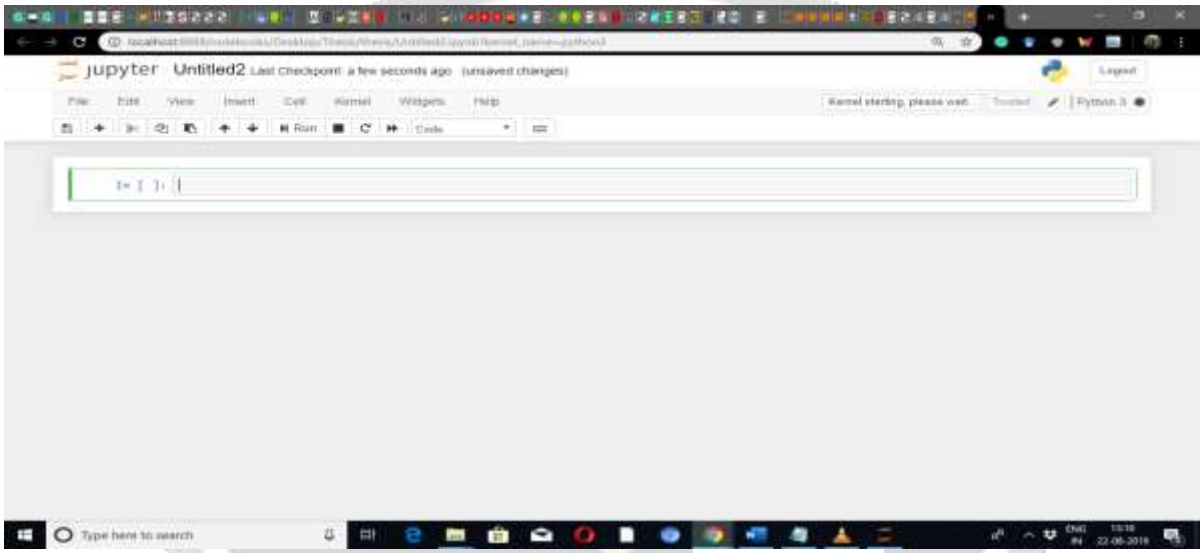

Figure 4.8: Programming window ofJupyter Notebook

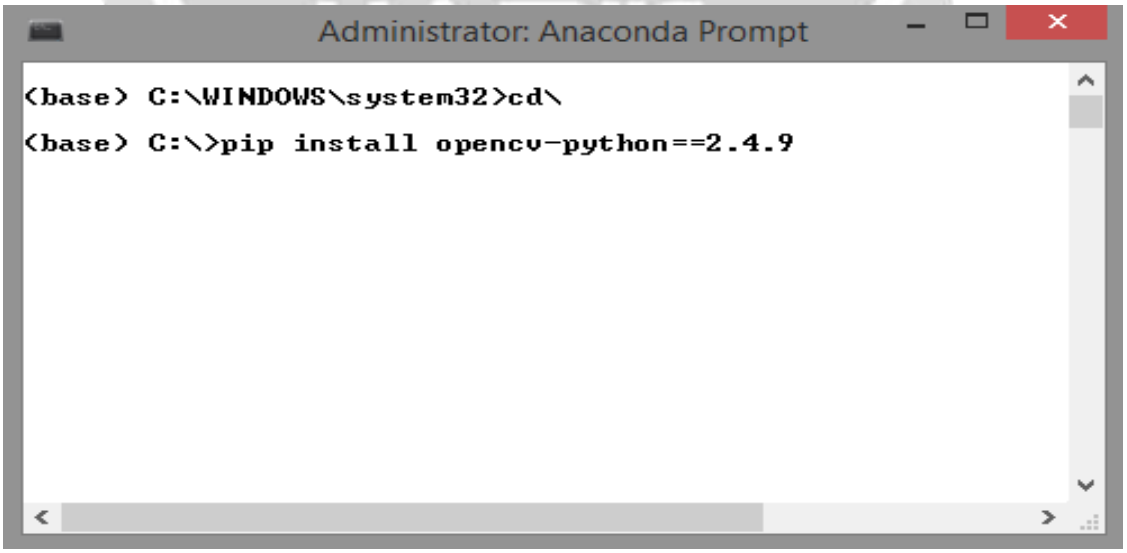

Figure 4.9: Anaconda prompt

V COMPUTATIONAL RESULTS AND DISCUSSION

The purpose of this study is to create a predictive algorithm to help hospitals and medical facilities maximize the number of survivors by providing an accurate and reliable tool to help medical decision making and triage COVID-19 patients more effectively and accurately during the pandemic.
We found that age is the variable that presents higher risk of COVID-19 mortality, where 60 or older patients have an OR = 18.8161 (CI95% [7.1997; 41.5517]). Regarding comorbidities, cardiovascular disease appears to be the riskiest (OR= 12.8328 CI95% [10.2736; 15.8643], along with chronic respiratory disease (OR=7.7925 CI95% [5.5446; 10.4319]). Males are more likely to die from COVID-19 (OR=1.8518 (CI95% [1.5996; 2.1270]).

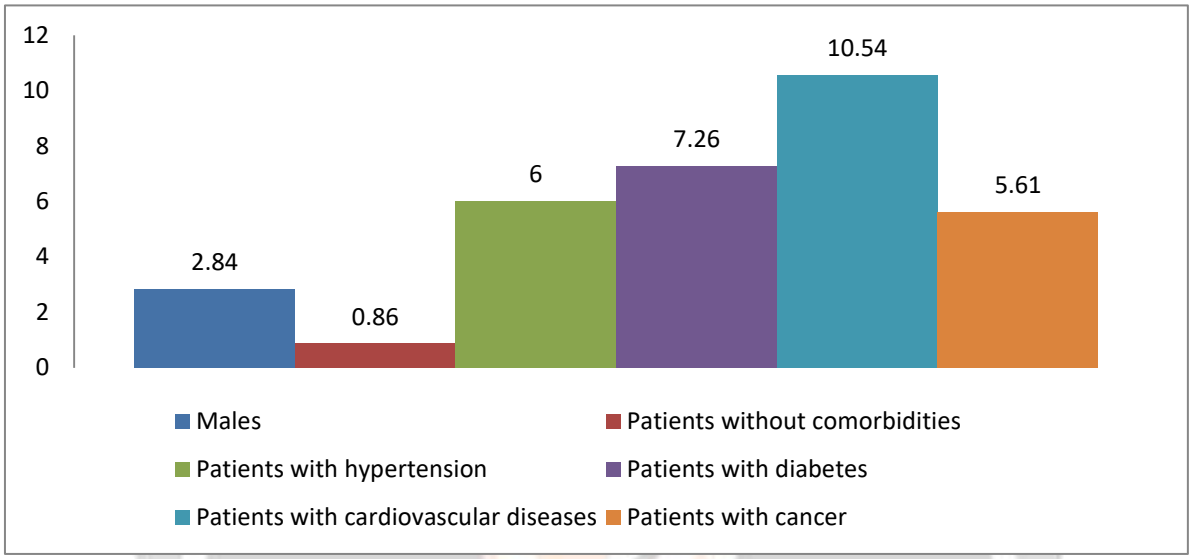

Figure 5.1 Relative frequencies and corresponding confidence intervals for 1000 simulated datasets and the actual values

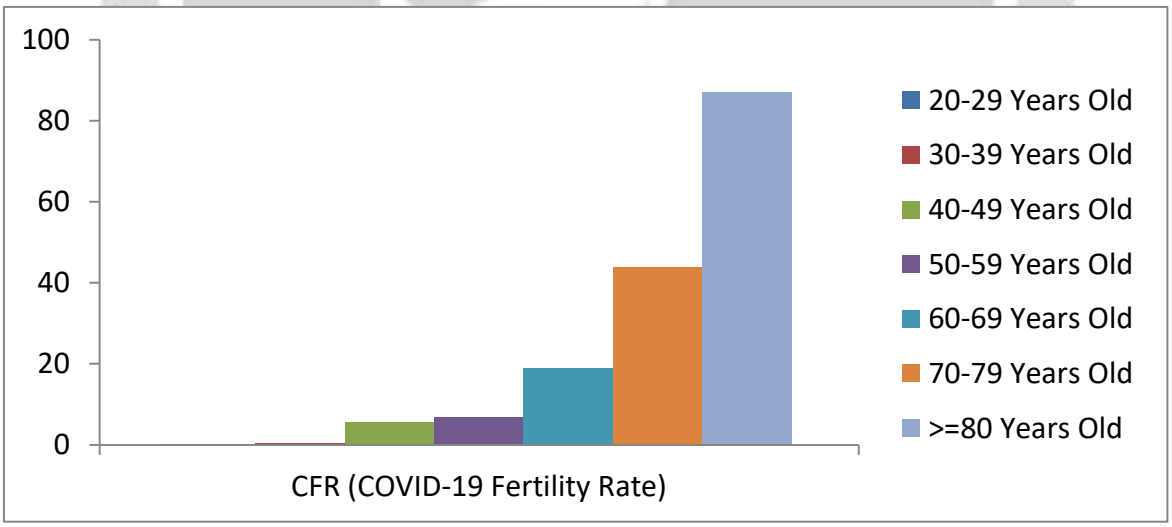

Figure 5.2 shows the statistics according to patients age and the corresponding values from the China CDC report.

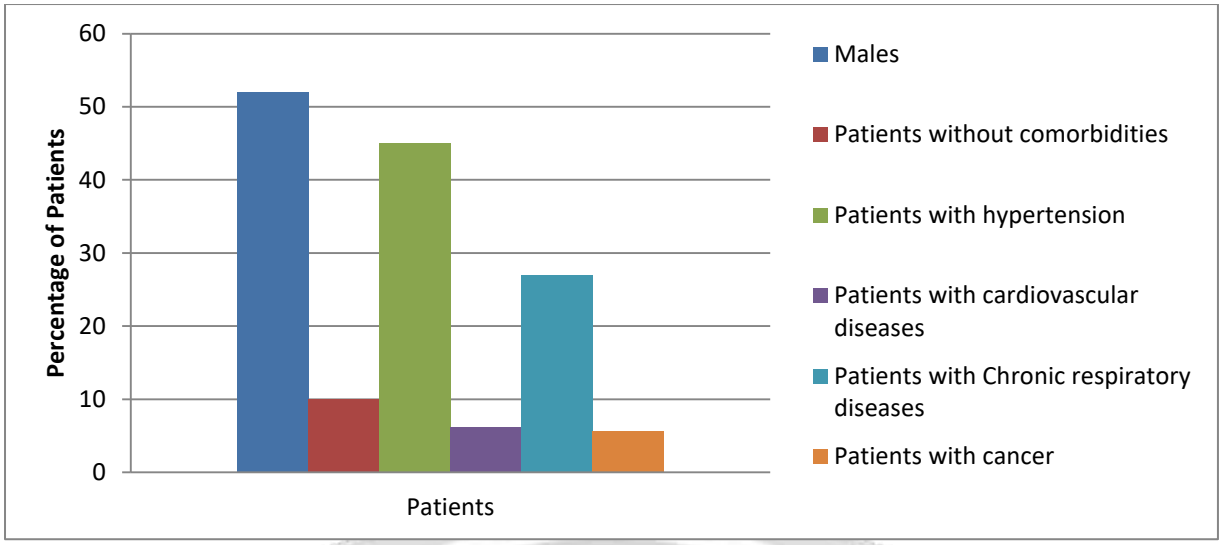

Figure 5.3 Relative frequencies and corresponding percentage of patient's percentage values.

V CONCLUSION

In this study an attempt was made to determine the most general equation for any general patient. This system can help hospitals, medical facilities, and caregivers decide who needs to get attention first before other patients, triage patients when the system is overwhelmed by overcrowding, and also eliminate delays in providing the necessary care.

Our study is very helpful to estimate COVID-19 patients' health conditions more effectively and accurately during the pandemic. Our estimated score very helpful to hospitals and medical facilities minimize decision-making time. Our study is to provide an accurate and reliable tool to help medical decision-making and triage COVID-19 patients. We provide an automotive tool that reduces the medical staff time and effort to examine COVID-19 patient's condition

References

1. Coronavirus 2019-nCoV, CSSE. Coronavirus 2019-nCoV Global Cases by Johns Hopkins CSSE .https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6)
2. Tyrrell DA, Bynoe ML. Cultivation of viruses from a high proportion of patients with colds. Lancet 1966: 1: 76–77.
3. Qiu, W., et al. "The pandemic and its impacts." Health, culture and society 9 (2017): 1-11.
4. CoopersmithCM,AntonelliM,BauerSR,DeutschmanCS,EvansLE,FerrerR,HellmanJ,JogS,KesecioglU, KissoonN,Martin-LoechesI,NunnallyME,PrescottHC,RhodesA,TalmorD,TissieresP,De Backer D, The Surviving Sepsis Campaign: Research Priorities for Coronavirus Disease 2019 in Critical Illness. Critical care medicine. 2021 Apr 1 [PubMed PMID: 33591008]
5. Huang C,WangY,LiX,RenL,ZhaoJ,HuY,ZhangL,FanG,XuJ,GuX,ChengZ,YuT,XiaJ,WeiY,WuW,XieX,YinW, LiH,LiuM,XiaoY,GaoH,GuoL,XieJ,WangG,JiangR,GaoZ,JinQ,WangJ,Cao B, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet (London, England). 2020 Feb 15; [PubMed PMID: 31986264]
6. Liu H,WeiP,ZhangQ,ChenZ,AviszusK,DowningW,PetersonS,ReynosoL,DowneyGP,FrankelSK,KapplerJ, MarrackP,Zhang G, 501Y.V2 and 501Y.V3 variants of SARS-CoV-2 lose binding to Bamlanivimab {i}in vitro{/i}. bioRxiv : the preprint server for biology. 2021 Feb 16 [PubMed PMID: 33619479]
7. HuaA,O'GallagherK,SadoD,Byrne J, Life-threatening cardiac tamponade complicating myo-pericarditis in COVID-19. European heart journal. 2020 Jun 7 [PubMed PMID: 32227076]
8. GuoT,FanY,ChenM,WuX,ZhangL,HeT,WangH,WanJ,WangX,Lu Z, Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). JAMA cardiology. 2020 Jul 1 [PubMed PMID: 32219356]
9. Libby P,LoscalzoJ,RidkerPM,FarkouhME,HsuePY,FusterV,HasanAA,Amar S, Inflammation, Immunity, and Infection in Atherothrombosis: JACC Review ANALYSIS OF MACHINE LEARNING

BASED COVID RISK ESTIMATION of the Week. Journal of the American College of Cardiology. 2018 Oct 23 [PubMed PMID: 30336831]

10. Abou-Ismail MY,DiamondA,KapoorS,ArafahY,Nayak L, The hypercoagulable state in COVID-19: Incidence, pathophysiology, and management. Thrombosis research. 2020 Oct [PubMed PMID: 32788101]
11. AmgalanA,Othman M, Exploring possible mechanisms for COVID-19 induced thrombocytopenia: Unanswered questions. Journal of thrombosis and haemostasis : JTH. 2020 Jun [PubMed PMID: 32278338]
12. Ni W,YangX,YangD,BaoJ,LiR,XiaoY,HouC,WangH,LiuJ,YangD,XuY,CaoZ,Gao Z, Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. Critical care (London, England). 2020 Jul 13 [PubMed PMID: 32660650]
13. ZubairAS,McAlpineLS,GardinT,FarhadianS,KuruvillaDE,Spudich S, Neuropathogenesis and Neurologic Manifestations of the Coronaviruses in the Age of Coronavirus Disease 2019: A Review. JAMA neurology. 2020 Aug 1 [PubMed PMID: 32469387]
14. Patel KP,PatelPA,VunnamRR,HewlettAT,JainR,JingR,Vunnam SR, Gastrointestinal, hepatobiliary, and pancreatic manifestations of COVID-19. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2020 Jul [PubMed PMID: 32388469]
15. 1] GabarreP,DumasG,DupontT,DarmonM,AzoulayE,Zafrani L, Acute kidney injury in critically ill patients with COVID-19. Intensive care medicine. 2020 Jul [PubMed PMID: 32533197]
16. BorczukAC,SalvatoreSP,SeshanSV,PatelSS,BusselJB,MostykaM,ElsoukkaryS,HeB,DelVecchioC,Fort arezzaF,PezzutoF,NavalesiP,CrisantiA,FowkesME,BryceCH,CalabreseF,Beasley MB, COVID-19 pulmonary