

Bayesian Reasoning and Artificial Intelligence against COVID-19

Michael Gr. Voskoglou^{1*}, Abdel-Badeeh M. Salem²

¹School of Technological Applications, Graduate Technological Educational Institute of Western Greece, Patras, Greece, E-mail: voskoglou@teiwest.gr

²Faculty of Computer and Information Sciences, Ain Shams University, Abbasia, Cairo, Egypt
E-mail: abmsalem@yahoo.com

*Corresponding author details: Prof. M. Voskoglou, mvoskoglou@gmail.com

ABSTRACT

The article focuses on the potential role of Probability Theory and Artificial Intelligence in the battle against the pandemic of COVID-19, which, starting from China on December 2019, has created a chaos in the world economy and the lives of people, causing hundreds of thousands of deaths until now. After discussing the importance of the reproduction number R_0 of the viruses, the Bayesian Probabilities are used for measuring the creditability of the diagnostic tests for the novel coronavirus. Artificial Intelligence designs are also described which are used as tools against COVID-19 and a Case-Based Reasoning expert system is proposed for the COVID-19 diagnosis.

Keywords: COVID-19, reproduction number R_0 , Bayesian probabilities, Artificial Intelligence (AI), Expert System (ES), Machine Learning (ML), Case-Based Reasoning (CBR).

INTRODUCTION

In December 2019 a cluster of people was detected in a seafood market of the city of Wuhan, China, suffering with pneumonia of unknown case [1]. The medical tests performed indicated that all those patients had been infected by a previously unknown virus, the seventh member of the family of coronaviruses that infect humans (Figure 1).

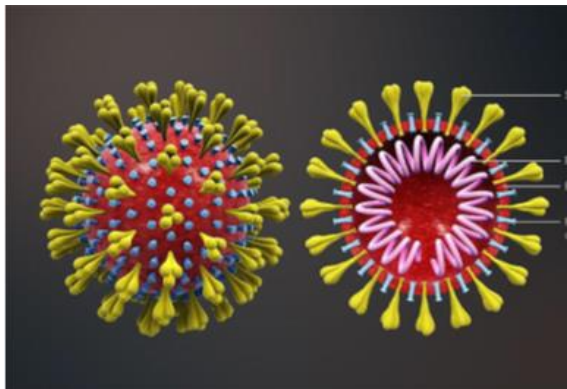


FIGURE 1: Coronavirus as it looks in the electron microscopy

The rest of the story is more or less known to everyone. The disease caused by the new virus, usually referred as COVID-19, was rapidly turned to a pandemic spreading all around the globe, creating a mess to the economies of most countries, affecting dramatically the lives of people and causing a large number of deaths approaching 600000 people until now (Figure 2).



FIGURE 2: Patient with COVID-19 wearing an oxygen mask

The rate of mortality of COVID-19, estimated between 0.5% and 2%, is not actually too high. Nevertheless, the problem is that the virus is transmitted quite easily. In fact, with the help of suitable mathematical models the scientists estimated that the mean value of its *reproduction number* R_0 [2] exceeds 3 (the results vary according to the method used and the existing each time conditions). Consequently, more than three on average other people are infected during the infectious period by one person who has been already infected with the virus. This practically means that the number of infections will be doubled in less than one week.

All the above hold, however, when no one in a population has got immunity from the virus and no restriction measures (e.g. social distancing, travel restrictions, self-quarantine, surveillance of the contacts of the infected persons, etc.) have been taken. In this case R_0 is called the *basic reproduction number*, whereas the *effective reproduction number* is calculated under the real conditions existing at a certain time.

The mathematical models used for calculating R_0 take into account several data, like the number of deaths, the number of hospital admissions, the number of the detected positive cases, etc. In case of COVID-19, however, there exist serious difficulties for accurate calculations, because the infectious period of the virus is not precisely known yet, the number of the asymptomatic cases transmitting the virus cannot be counted easily, there is not known whether or not an infected person who has recovered has got immunity and for how long, etc.

The reproduction number R_0 was introduced by the demographer Alfred Lorde in the 1920's for calculating the rate of reproduction in a given population. The epidemiologist George McDonald used it for first time in the 1950's for measuring the transmission potential of malaria. It is recalled that:

- If $R_0 < 1$, then the corresponding epidemic declines and could eventually die.
- If $R_0 = 1$, then the epidemic remains stable, but no outbreak occurs.
- If $R_0 > 1$, then the transmission rate increases exponentially and an outbreak is possible to appear.

It is of worth noting that, if the $1 - \frac{1}{R_0}$ of the whole

population has been infected, then the epidemic cannot be transmitted easily. This is known as the *herd immunity* [3]. For example, for $R_0 = 3$ the previous formula gives a percentage approximately equal to 67% of the whole population. In case of COVID-19 the herd immunity is practically impossible to be achieved. In Great Britain, for instance, it has been estimated that this would cause 200000 deaths and eight million hospital admissions of patients with serious symptoms! In conclusion, since no vaccine exists yet for COVID-19, the only available strategy for a government in order to reduce the value of R_0 under 1, is to take the proper restriction measures as soon as possible. This of course could create a chaos in the society and the economy of the corresponding country and always involves the danger of a comeback of the virus, when the measures start to loosen. In the meantime, however, the government could have the opportunity to improve the country's healthcare system in order to become able to tackle effectively the new situation.

The present article focuses on the use of *Probability Theory* and *Artificial Intelligence (AI)* as weapons in the battle against COVID-19. The next Section studies, with proper examples and the help of *Bayesian probabilities*, the creditability of the diagnostic tests for the novel coronavirus. In third Section designs of AI are described that are used as tools against COVID-19. The fourth Section proposes a *Case-Based Reasoning (CBR)* expert system for the COVID-19 diagnosis and the article closes with the general conclusions presented in Section 5.

CREDITABILITY OF THE COVID-19 DIAGNOSTIC TESTS

Many epidemiologists argue that the diagnostic tests performed to the general population during the outbreak period of COVID-19 are not useful burdening purposeless the healthcare system of the corresponding country. In this Section we provide, with the help of the *Bayesian probabilities*, examples supporting this view.

The Bayes' rule

It is well known that the *conditional probability* for an event A to happen when another event B has already happened is calculated by

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (1).$$

In the same way we have that

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \text{ or } P(A \cap B) = P(B/A) P(A).$$

Therefore (1) can be written in the form

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \quad (2).$$

Equation (2), which calculates $P(A/B)$ with the help of the inverse in time conditional probability $P(B/A)$, the prior probability $P(A)$ and the posterior probability $P(B)$, is known as the *Bayes' rule*.



FIGURE 3: Thomas Bayes (1701-1761)

The Bayes' rule was first appeared in the work "An Essay towards a Problem in the Doctrine of Chances" of the 18th century British mathematician and theologian Thomas Bayes (Figure 3). This essay was published by Richard Price in 1763, after the Bayes' death, in the "Philosophical Transactions of the Royal Society of London". The famous French mathematician Laplace (1749-1827) was the one who pioneered and popularized the Bayesian probabilities.

Although the Bayes' rule is a simple consequence of equation (1) calculating the value of a conditional probability, it has been proved to be very important for the whole science, while recent researches give evidence that even the mechanisms under which the human brain works are Bayesian! [4-6]. Sir Harold Jeffreys (1891-1989), a British mathematician who played an important role in the revival of the Bayesian view of probability, characterized the Bayesian rule as the "Pythagorean Theorem of Probability Theory" [7].

Diagnostic tests

EXAMPLE 1: The statistical data show that 2% of the inhabitants of a country have been infected by the COVID-19 virus. Mr. B, who has not any symptoms of the disease, makes a diagnostic test. The test, the statistical accuracy of which is 97%, is positive. What is the probability for Mr. B to be a carrier of the virus?

Solution: Consider the following events:

- A: The subject is a carrier of the virus.
- B: The test is positive.

On the basis of the given data it turns out that $P(A) = 0.02$ and $P(B/A) = 0.97$.

Among 100 inhabitants of the country, 2 on average are carriers and 98 are noncarriers of the virus. Assuming that all those people make the test, we should have on average $2 \times 97\% = 1.94$ positive tests for the carriers and $98 \times 3\% = 2.94$ positive tests from the noncarriers of the virus, i.e. 4.88 in total positive tests. Therefore, $P(B) = 0.488$. Replacing the values of $P(A)$, $P(B/A)$ and $P(B)$ in equation (2) one finds that $P(A/B) \approx 0.398$.

Therefore, the probability for Mr. B to be a carrier of the virus is only 39.8% and not 97%, as it could be thought through a first, rough estimation! This means that Mr. B has to make a second test to see what really happens with his health condition. Further, if the second test is negative, a third test will be also required. At the same time, however, there is an urgent need for other people to make the test. This becomes evident by the next example.

EXAMPLE 2: Modifying the previous example, assume that Mr. B has some suspicious symptoms and that 85% of the people presenting such symptoms have been infected by the virus. Mr. B makes the test, which is positive. What is now the probability for Mr. B to be a carrier of the virus?

Solution: Let A and B be the events defined in Example 1. Here we have that $P(A) = 0.85$ and $P(B/A) = 0.97$. Further, assuming that 100 people having suspicious symptoms make the test, we shall have on average $85 \times 97\% = 82.45$ positive tests from the carriers and $15 \times 0.3\% = 0.45$ from the noncarriers of the virus, i.e. 82.9 in total positive tests. Therefore, $P(B) = 0.829$. Replacing the values of $P(A)$, $P(B/A)$ and $P(B)$ in equation (2) one finds that $P(A/B) \approx 0.995$.

In this case, therefore, the probability for Mr. B to be a carrier of the virus is 99.5%, i.e. exceeds the statistical accuracy of the test!

In general, the sensitivity of the solution of the problem is great, depending on the values of the prior probability $P(A)$. The greater the value of $P(A)$, the higher is the accuracy of the test. Our third and last example examines what happens, if the test is negative.

EXAMPLE 3: Assume that Mr. B makes a diagnostic test for COVID-19, which is negative. Find the probability to be a carrier of the virus: i) Under the conditions of Example 1, and ii) under the conditions of Example 2.

Solution: Consider the following events:

- A: The subject is a carrier of the virus.
- B: The test is negative.

i) In this case we have $P(A) = 0.02$ and $P(B/A) = 0.03$. Assuming that 100 people make the test, we shall have on average $98 \times 97\% = 95.06$ negative tests from the noncarriers and $2 \times 3\% = 0.06$ from the carriers of the virus, i.e. an average of 95.12 in total negative tests. Therefore, $P(B) = 0.9512$ and replacing the values of $P(A)$, $P(B)$ and $P(B/A)$ to equation (1) one finds that $P(A/B) \approx 0.0006$. Therefore, the probability for Mr. B to be a carrier of the virus is only 0.06%.

ii) Here we have $P(A) = 0.85$ and $P(B/A) = 0.03$. Further, assuming that 100 people make the test, we shall have on average $15 \times 97\% = 14.55$ negative tests from the noncarriers and $85 \times 3\% = 2.55$ from the carriers of the virus, i.e. an average of 17.1 in total negative tests. Therefore, $P(B) = 0.171$ and replacing the values of $P(A)$, $P(B)$ and $P(B/A)$ to equation (1) one finds that

$P(A/B) \approx 0.1491$. Therefore, the probability for Mr. B to be a carrier of the virus is 14.91%.

Concerning the sensitivity of the solution in this case, one observes that the greater the value of the prior probability $P(A)$, the smaller the accuracy of the test.

AI IN THE FIGHT AGAINST COVID-19

Artificial Intelligence

AI is the area of Computer Science that deals with the theory and practice of constructing smart machines, which think and behave like humans [8]. Although the roots of AI can be traced much earlier (e.g. Alan Turing's theoretical learning machine in 1936), the term was coined by John McCarthy in 1956, when he organized the first conference on the subject in Dartmouth College, USA [9].

Machine learning (ML) is the branch of AI that focuses on enabling computers to learn from data and make improvements without dependence on any program's commands. This could eventually help computers in building models for predicting the weather, for medical purposes, etc. MATLAB applications let you see how different algorithms work with your data, while WEKA is considered to be the workbench for ML Python ML modules have also become very popular, some of them specialized to technologies such as the artificial neural networks. The Scikit-learning module, since its release in 2007, has been widely used as a general-purpose, open-source module both in industry and academia.

AI, being a synthesis of ideas from mathematics, engineering, technology and science, constitutes a main component of the forthcoming fourth industrial revolution. This new revolution is characterized by the Internet of Things and Energy (IoT&E) and by the cyber-physical systems controlled through this advanced form of the Internet [10]. As a result, AI finds nowadays a continuously increasing number of applications to many sectors of the human activity.

Expert Systems

A *system* is a set of interacting components forming an integrated whole. A system involves a process or a collection of processes transforming inputs to outputs; e.g. the output in the system of a bus is the movement of people to their destinations.

An *expert system* is a system developed with the help of the AI theories and methodologies [11]. The application of AI techniques enables the expert systems to accept knowledge from human input and using it to solve problems with the help of simulated reasoning methodologies. Those methodologies include automated, rule-based, case-based, model-based, geometric, fuzzy and several other modes of reasoning.

The successive steps for the development of an expert system involve:

- Construction of the knowledge base including collection, acquisition and representation of the required knowledge.
- Choice of the appropriate reasoning and inference methodology.
- Selection of the suitable authoring shells, which allow the system's user to easily enter the knowledge base without requiring computer programming skills

Expert systems against COVID-19

The AI companies are using their cutting-edge technologies to help fight the COVID-19 outbreak [12, 13]. Sense Time, a Chinese AI pioneer, has allocated 20 GPU server nodes to help Luo Haibin, a professor at the School of Pharmaceutical Sciences of Sun Yat-Sen University, with research on large scale screening of the drugs against new coronaviruses.

Sense Time has also constructed powered thermal cameras that can rapidly detect people who have fever (body temperature 37.3°C , or higher) among those in a crowd. A facial algorithm and a proper thermal imaging temperature measurement are used for this purpose (Figure 4). The cameras can be deployed in airports, in train and subway stations, in shopping centers and in several other crowded places. The company has currently more than 700 customers, partners and alliances including MIT, Honda, Alibaba, Suning, Wanda, Xiaomi, China Mobil, Union Ray, etc.



FIGURE 4: Thermal imaging and fever detection

Intelligent medical imaging evaluation systems have been also developed to help doctors diagnose pneumonia caused by the novel coronavirus (Figure 5). Those systems are able to perform a quantitative analysis of the lungs and suggest therapy options.



FIGURE 5: Intelligent medical imaging evaluation system for pneumonia

The recently developed by the Shanghai-based Yitu Technology Co Ltd evaluation system has the ability to classify pneumonia according to its severity, precisely calculate the burden on the lungs and provide 4D dynamic pathology comparisons. Traditional quantification analysis performed by humans usually requires up to 6 hours, whereas the new AI system performs the same analysis in less than 3 seconds!

Those are some of the most important contributions of AI to the battle against COVID-19, but the catalogue does not stop here. The battle is continued!

DEVELOPING A CBR EXPERT SYSTEM FOR THE COVID-19 DIAGNOSIS AND TREATMENT

The CBR methodology

CBR is an analogical reasoning method from experiences, usually referred as *past cases*, in an effort to solve problems, critique solutions and explain complicated situations [14]. This explains, for example, why a patient feels more comfortable with older doctors, who have seen and treated more patients who had illnesses similar to his own.

Each case is composed of three major parts:

- The problem's description.
- The solution procedure, including an explanation of why its steps are correct and a list of methods for verifying it.
- The outcomes

CBR provides both a cognitive model for problem solving and learning in computers and people [14], as well as a very popular nowadays reasoning methodology for constructing expert systems [16]. The use of computers enables a CBR system to maintain a continuously increasing library of past cases, usually referred as the *case-memory*, and to retrieve from it each time the most similar cases that could lead to the solution of a new problem.

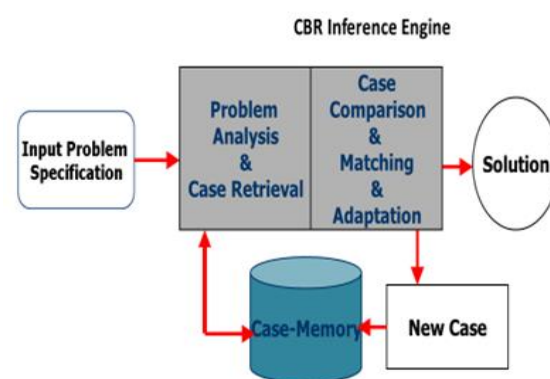


FIGURE 6: CBR methodology

The CBR methodology, which is graphically illustrated in Figure 6, addresses problems analogous to those found in developing rule-based ES, namely knowledge acquisition, adaptation, performance, and maintenance.

From the computational intelligence and ML perspective, CBR refers to a number of concepts and techniques (e.g. data structures and intelligent algorithms) that can be used to perform the following processes:

- Record and index cases.
- Search cases to identify the ones that might be useful in solving new cases when they are presented.
- Modify earlier cases to better match new cases.
- Synthesize new cases when they are needed.

CBR is a preferred method in dynamically changing situations and in other cases where solutions are not clearly cut. This happens because knowledge acquisition for a classical rule-based system would be extremely difficult in such situations, and is likely to produce incomplete or inaccurate results. The main CBR applications include domains such as diagnosis, help-desk, assessment, decision support, design, etc.

The proposed hybrid ES for COVID-19

Currently, CBR methodology is used in developing medical ES reasoning in analogy to the past medical cases. A hybrid ES with CBR reasoning methodology is proposed here for the COVID-19 diagnosis and treatment (Figure 7)

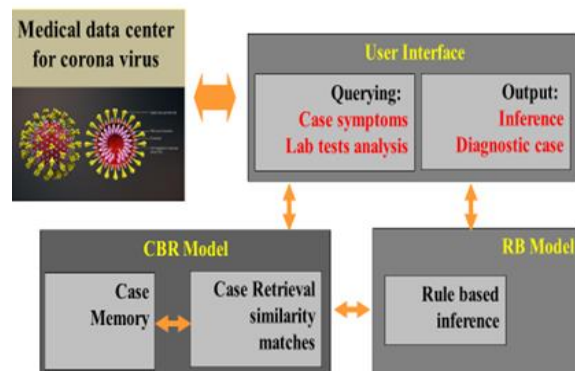


FIGURE 7: Architecture of the proposed hybrid ES for COVID-19 diagnosis

The general lines of the proposed CBR system are the following: The COVID-19 disease has several initial signs and symptoms. If any of them is detected, the patient should perform complete clinical examinations. The patient's medical case (input data) is analyzed in the system's CBR unit by matching it to previous similar cases. The CBR unit is connected to a rule-based unit and the two units collaborate to produce the final diagnosis and the treatment suggestions.

The proposed CBR system for COVID-19 could help the young physicians, especially those working in remote and isolated areas of a country (islands, small villages on the mountains, etc.), to check their diagnosis and suggest the proper treatment on the basis of the recorded experience from older similar cases. It could be also used as a learning device for medical students.

The cooperation between experienced physicians and AI experts is essential for the development of the technical details of the proposed ES that could guarantee its effective function in practice.

CONCLUSIONS

The discussion performed in the present paper leads to the following general conclusions:

- The continuous detection of the values of the reproduction number R_0 is very important and useful for predicting the evolution of the COVID-19 pandemic on the purpose of taking the required each time measures.
- The use of the Bayesian probabilities plays a significant role for checking the creditability of the viruses' diagnostic tests.
- AI in general and ML in particular have been proved to be valuable tools in the battle against the COVID-19 pandemic,
- A CBR-based ES has been proposed for the COVID-19 diagnosis and treatment that could essentially help doctors working in remote and isolated from the center places of a country.

REFERENCES

- [1] Na Zhu, et al., A Novel Coronavirus from Patients with Pneumonia in China, 2019, The New England Journal of Medicine, 382 (8), 727-733, 2020.
- [2] Gallagher, J., Corona: What is the R_0 number and how is it calculated? May 2020, available in [bbc.com/news/health-52473523](https://www.bbc.com/news/health-52473523).
- [3] Fine, P., Eames, K., Heymann, D.L., Herd Immunity: A rough guide, Clinical Infections Diseases, 52(7), 911-916, 2011
- [4] Hoghan, J., Bayes' Theorem: What is the Big Deal? January 2015, available in <http://blogs.scientificamerican.com/cross-check/bayes-s-theorem-what-s-the-big-deal>.
- [5] Bertsch McGrayne, S., The Theory that would not die, Yale University Press, New Haven and London, 2012.
- [6] Athanassopoulos, E. and Voskoglou, M.Gr., A Philosophical Treatise on the Connection of Scientific Reasoning with Fuzzy Logic, Mathematics, 8, article 875, 2020.
- [7] effreys, H., Scientific Inference, 3d Edition, Cambridge University Press, UK, 1973.
- [8] Mitchell, M., Artificial Intelligence: A Guide for Thinking Like Humans; Parrar, Straus and Gtraux, NY, USA, 2019.
- [9] Moor, J., The Dartmouth college artificial intelligence conference: The next fifty years, Artificial Intelligence Magazine, 27, 87-91, 2006.
- [10] Schwab, K., The Fourth Industrial Revolution, Crown Publishing Group: NY, USA, 2016.
- [11] Salem, A.-B.M. and Voskoglou, M.Gr., Reasoning Methodologies in for Intelligent Information Systems, International Journal of Applications of Fuzzy Sets and Artificial Intelligence, 5, 111-127, 2015.
- [12] Nguyen, T.T, AI in the Battle against Coronavirus(COVID-19): A Survey and Future Research Directions, preprint, 2020, DOI:10.13140/RG.22.36491.23846
- [13] Randhawa, G.S., Soltysiak, M.P.M., El Roz, H., de Souza, C.P.E., Hill, K.A., Kari, L., Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Corona case study, Plos One, 15(4):e0232391, April 2020.
- [14] Voskoglou, M.Gr. and Salem, A.-B.M., Analogy Based and Case-Based Reasoning: Two Sides of the Same Coin, International Journal of Applications of Fuzzy Sets and Artificial Intelligence, 4, 5-51, 2014.
- [15] Voskoglou, M.Gr., Case-Based Reasoning: A Recent Theory for Problem-Solving and Learning in Computers and People, Communications in Computer and Information Science, 19, 314-319, Springer-Verlag, 2008.
- [16] Voskoglou, M.Gr. and Salem, A.-B.M., Benefits and Limitations of the Artificial with Respect to the Traditional Learning of Mathematics, Mathematics, 8, article 611, 2020.