Research Article

Mustafa Kamal Pasha, Syed Fasih Ali Gardazi, Fariha Imtiaz*, Asma Talib Qureshi, and Rabia Afrasiab

Identification of efficient COVID-19 diagnostic test through artificial neural networks approach – substantiated by modeling and simulation

https://doi.org/10.1515/jisys-2021-0041
received March 16, 2021; accepted May 08, 2021

Abstract: Soon after the first COVID-19 positive case was detected in Wuhan, China, the virus spread around the globe, and in no time, it was declared as a global pandemic by the WHO. Testing, which is the first step in identifying and diagnosing COVID-19, became the first need of the masses. Therefore, testing kits for COVID-19 were manufactured for efficiently detecting COVID-19. However, due to limited resources in the densely populated countries, testing capacity even after a year is still a limiting factor for COVID-19 diagnosis on a larger scale and contributes to a lag in disease tracking and containment. Due to this reason, we started this study to provide a better cost-effective solution for enhancing the testing capacity so that the maximum number of people could get tested for COVID-19. For this purpose, we utilized the approach of artificial neural networks (ANN) to acquire the relevant data on COVID-19 and its testing. The data were analyzed by using Machine Learning, and probabilistic algorithms were applied to obtain a statistically proven solution for COVID-19 testing. The results obtained through ANN indicated that sample pooling is not only an effective way but also regarded as a "Gold standard" for testing samples when the prevalence of the disease is low in the population and the chances of getting a positive result are less. We further demonstrated through algorithms that pooling samples from 16 individuals is better than pooling samples of 8 individuals when there is a high likelihood of getting negative test results. These findings provide ground to the fact that if sample pooling will be employed on a larger scale, testing capacity will be considerably increased within limited available resources without compromising the test specificity. It will provide healthcare units and enterprises with solutions through scientifically proven algorithms, thus, saving a considerable amount of time and finances. This will eventually help in containing the spread of the pandemic in densely populated areas including vulnerably confined groups, such as nursing homes, hospitals, cruise ships, and military ships.

Keywords: COVID-19, COVID-19 testing, binary search methods, machine learning, artificial intelligence

*** Corresponding author: Fariha Imtiaz,** Punjab University College of Pharmacy, University of the Punjab, Allama Iqbal Campus, Lahore, Pakistan, e-mail: farihaaimtiaz@gmail.com
Mustafa Kamal Pasha: Business R&D and Operations, Myst Enterprise, Christchurch, New Zealand, e-mail: mustafa@mystenterprise.com
Syed Fasih Ali Gardazi: School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan, e-mail: fasih.gardazi@seecs.edu.pk
Asma Talib Qureshi: Atta Ur Rahman School of Applied Biosciences, National University of Sciences and Technology, Islamabad, Pakistan, e-mail: asma.q@live.com
Rabia Afrasiab: Department of Medicine, Unit 2, University of Health Sciences, Lahore, Pakistan, e-mail: rabiasims453@gmail.com

Open Access. © 2021 Mustafa Kamal Pasha *et al.*, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

1 Introduction

COVID-19 is a global healthcare problem that has overwhelmed both healthcare and economic sectors on a large scale. If the workforce is required to self-isolate due to contact with a confirmed case of coronavirus, consider ways they can continue to support multidisciplinary team meetings. This may be exacerbated by COVID-19 infection of medical personnel, quarantine requirements, and school closures, all of which may affect the staffing levels and increase stress on the capacity of these institutions [1]. Clinicians and other support staff have to work flexibly to facilitate safe service provision in alternative settings [2]. On one hand, social distancing and separation of clinic workspaces are important steps to reduce the risk of infection [3]. On the other hand, smart, cost-effective methods to identify the spread of viral disease in the community are scarce. To meet the increasing need for personal security devices, including personal protective equipment (PPE), testing kits, and lifesaving drugs, manufacturers utilized their capital exhaustively [1]. They excluded the importance of statistically proven solutions that would help by saving both time and money. Therefore, considering the increasing need of the masses, this study is designed to provide a cost-effective solution to pooled testing. This will not only reduce the cost, amount of workforce, and labor but also would have reduced the burden on the manufacturing sites for the PCR-based testing kits or other novel testing techniques in the market. Similarly, effective measures can be taken to reduce costs in billion dollars and time in days and months.

We started this study by conducting an extensive analysis of current testing techniques that are utilized in the diagnostic centers for COVID-19 testing. We also examined the utilization of artificial intelligence in the domain of healthcare with a particular focus on its role in COVD-19 testing. After a thorough analysis of the literature, we found that all those studies and techniques lacked in improvising the approved testing strategy for COVID-19 testing as they utilized radiographs obtained through CT scans and chest X-rays. According to US FDA, these tests are not approved as diagnostic tests for COVID-19 testing, and only PCR-based tests are regarded as “Gold Standard” for COVID-19 testing. For this reason, this study has been designed to provide a smart, alternative solution to the healthcare sector and enterprises. The principles of artificial neural networks (ANN) were employed in this regard, and the data were collected through bibliometric analysis by using selective keywords. The selected data were then subjected to statistical analysis by applying probabilistic algorithms. The results obtained demonstrated the effectiveness of sample pooling when the disease prevalence is low in the population. Thus, this study provides a cost-effective solution that would help in increasing the testing capacity while reducing the burden of testing and resource availability on the diagnostic centers and manufacturing companies. This ultimately would reduce the disease burden on the healthcare system of low-income countries and will help in robust disease tracking and containment.

1.1 Literature review

Though a considerable progress has been made since the first COVID-19 case reported in Wuhan, China, it still needs our efforts to contain its spread to resume the activities back to normal as early as possible and ultimately lessening the burden on the global economy. Considering the minimal potential to test and handle a massive influx of patients even after a year of the pandemic, we need to devise solutions to increase the efficiency without altering the specificity of testing that will ultimately reduce the burden on pharmaceutical and manufacturing companies and the local healthcare sectors as well [4]. Efforts were being made in this regard at the federal and state levels and were disbursed at a local level to provide better living and healthcare facilities to communities. Strategies were being designed to eliminate the need for extended stay-at-home instructions when enforcing risk-based lockdown, testing, touch-tracking, and surveillance procedures as the economic woes of the COVID-19 pandemic intensify [5]. Given the scarce resources and the critical fiscal, public health, and organizational complexities of the existing 14-days quarantine recommendation, it is important to understand whether it is possible to introduce more efficient

yet similarly effective testing techniques [5,6]. Therefore, it is the need of the time to improve the testing capacity and timely tackle and control COVID-19 cases. This would not only help in resuming the work, reducing the losses by the corporate sector but also can be taken as an opportunity to meet the needs of the shackled healthcare system in this pandemic.

Amid this healthcare challenge, out of many hard-hit sectors, the corporate sector was one of the main affectees that has faced a considerable setback due to jamming of all sorts of activities in every domain. COVID-19 has shaken not only the healthcare sector but has caused a considerable damage to businesses around the globe as well. Companies that offer information services continued to survive while manufacturing companies had to either close their workplaces or allow workers conditionally with stringent instructions of maintaining social distancing, avoiding congregations, and proper use of PPE. Under the domain of manufacturing enterprises, only those who were producing basic goods were allowed to operate [7]. And, the effect of these restrictions on all sorts of activities when evaluated was in the form of huge financial losses for small businesses due to disruption of the supply chain, reduced profits, and product demand. Due to these reasons, some enterprises could not survive the lockdown and were permanently closed while others are on the verge of collapsing if it continues for 2 or more months [8]. The main reason was found to be a lack of preparedness toward handling a pandemic by over 83% of these micro, small, and medium enterprises. The ultimate consequences can be observed in terms of economic disparity and increased joblessness in the past few months. Therefore, in some countries, a support package was launched to sustain these businesses [9]. But as the pandemic has struck again vivaciously, providing a support package is not sufficient and requires some preventive and control measures as well. Testing being the cornerstone of identifying COVID-19 positive cases is a need of all the enterprises and of every sector that had to shut down its activities under government instructions. To overcome the adversities of pandemic and to resume the activities back to normal by all the surviving enterprises, it is important to test individuals for COVID-19 before officially restoring the operational activities. But a single PCR-based test costs around $100, so it is economically not feasible for businesses to conduct tests of all the employees after facing an economic blow due to the current pandemic [10].

At present, COVID-19 tests approved by U.S Food and Drug Administration are classified as either diagnostic tests or antibody test both of which detects COVID-19 infection differently. Diagnostic tests can be done by detecting viral genetic material via Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) or it can be performed by detecting viral antigens in the test samples. Antibody test, on the other hand, detects the presence of antibodies against the viral antigen. The diagnostic test is usually performed on fresh swabs taken from individuals to detect viral load while an antibody test is performed on individuals who are already exposed to COVID-19 and have antibodies in the blood against the virus. An antibody test cannot be used to diagnose COVID-19 because antibodies sometimes take a week or more to be synthesized against the virus after initial infection and their levels can be detected even after recovering from COVID-19. Therefore, it cannot tell exactly if the body is still carrying the viral load. The only reliable way to detect the viral load of COVID-19 is through diagnostic tests that can indicate an active infection in the individual [11]. But each of these methods has its own drawbacks as well, depending on the context of testing done. Apart from these, radiology has also found its role in identifying COVID-19 cases by using chest X-rays and CT scans. These radiographs indicate the abnormalities in the lungs but cannot be taken as a way to detect COVID-19 [12]. Out of all these mentioned techniques for COVID-19 testing, the most reliable is RT-PCR-based testing (Table 1). Thus, after identifying the most effective type of diagnostic test, it is important to improve the capacity of already existing techniques so that a maximum number of individuals could be tested for COVID-19 that would ultimately help in identifying disease hotspots and tracing points of contact. As carrying out screening tests on a large scale is not possible with limited resources, Artificial Intelligence and Machine Learning are introduced into the healthcare sector to improve the working capacity with limited available resources.

By using data from diagnostic tests and X-rays or CT scans, AI and ML tools were utilized to improve the detection and tracking of COVID-19 [13–17]. However, these models have limitations in real-world implications despite being them up-to-the mark. Also, most of these studies include a pre-detection technique involved, whether it's an X-ray or CT scan. Mohammed et al. brought a novel strategy of predicting the cases

Table 1: Comparison of different COVID-19 testing techniques [11,12]

Testing method	RT-PCR	Antigen test	Antibody test	CT scan	Chest X-ray
Sample	Nasopharyngeal, nasal, and throat swabs	Nasopharyngeal and nasal swabs	Serological testing	Radiological examination	Radiological examination
Time taken to get the results	Same-day or results can take up to a week in some settings	Depends on the test performed, can take 15–20 min in some instances	Same day or within 1–3 days	40 min	15–20 min
Results indicate	Active coronavirus infection	Active coronavirus infection	Past infection of coronavirus	Lung infection, COVID pneumonia	Lung infection, COVID pneumonia
Cost of testing procedure	Ranges from $50 to $200	Ranges from $5 to $20	Ranges from $40 to $75	Varies from $100 to $1,000	Varies from $5 to $132
Advantages	Highly accurate, the gold standard of testing, does not need to be repeated	Highly accurate, fast detection, low cost	Indicates past infection and the body's response against the infection	Least contamination, results can be obtained on the same day	Least contamination results can be obtained on the same day
Disadvantages	Cannot indicate a past infection, results can take up to a week in some locations, expensive	False-positive results can be generated, needs molecular testing for validation	Cannot detect active virus, cannot be taken as a diagnostic test	Cannot indicate the presence of coronavirus, chances of false-negatives are high, expensive	Cannot indicate the presence of coronavirus, chances of false-negatives are high, expensive

through true positive and negative Mathew's correlation sensitive analysis merged with entropy and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) but again the input data has to have previous reliability on other detection methods, which involves regressive and expensive techniques. Moreover, this study is relied on laboratory- and hospital-based data and feature selection [18]. In previous studies, chest CT scans were integrated with clinical symptoms by using AI that helped in improving the detection of COVID-19 even in cases in which a normal CT scan was obtained [19]. AI also helped in differentiating COVID-19 from community-acquired infections like pneumonia [20]. It also contributed to coupling of self-testing with disease tracking which would help in containing the spread of disease [21]. Many COVID-19 hotspots were detected by using AI and it was also used to predict the future hotspots [22]. All these studies demonstrated the use of AI and ML in the healthcare sector, particularly in the case of a pandemic. It should be noted here that in all these AI- and ML-based studies, the area of focus remained improving radiological findings by utilizing chest X-rays and CT scan images of normal and COVID-positive individuals. These tests are not approved as COVID-19 diagnostic tests, but in case of any emergency situation, they can be used to identify the extent of infection in individuals that when combined with clinical symptoms may indicate COVID-19 infection. Therefore, our study utilizes AI for improving the capacity for RT-PCR that is the gold standard for COVID-19 diagnosis. These findings will provide a cost-effective testing solution for the healthcare sector and manufacturing enterprises as well.

This study analyses the effectiveness of pooled testing of COVID-19 through ANN approach and Machine Learning by using binary search. For validating the hypothesis, ANN and Machine Learning were used as they have shown high performance in obtaining and interpreting huge volumes of data [23]. It works on similar principles as of the human brain in which the input signal after processing gives an output. In the same way, data fed in ANN are interpreted in different layers, and the results are displayed through the output layer. In the simplest model, three layers exist that are input layer, hidden layer, and output layer. But the intermediate hidden layers may increase in number depending on the amount of information and its complexity [24]. ANN though works on similar principles of human brains, but its processing and results are more profound than a human brain. Its capacity to solve problems that are almost impossible for humans make it an efficient tool used in Artificial Intelligence (AI) and is, therefore, used in this study as well to analyze and model the huge volumes of data fed through the input layer [23,25]. Machine Learning tools were then applied to the data generated through ANN, and by using the binary search tool, the data was sorted. The binary search tool has advantages over linear search as it can sort large volumes of data effectively in a short time and, therefore, is more suitable for sorting population-based data.

In this study, we have collected bibliometric data that are analyzed through the ANN approach and by Machine Learning have developed algorithms. These algorithms were used in evaluating the efficiency of sample pooling first with eight samples that were later extended to 16 samples. On this basis, in this study, it is proposed that pooling of samples will not only increase the testing capacity but also work without altering the specificity of the testing procedure that will stay the same 96% even with a pool of 16 samples. This will help in reducing the cost and will prove highly economical in population-based studies where the prevalence rate of diseases is usually low. It will be helpful not only in detecting COVID-19 cases but also in testing both symptomatic and asymptomatic carriers and tracing the points of contact in limited resources [26]. Due to these reasons, it will stand in a more economical position than the current strategy of individual testing that no doubt if has merits, also has demerits of an economic burden on the healthcare sector and an increased consumer demand for manufacturing companies. Therefore, introducing Artificial Intelligence in the healthcare sector not only will provide cost-effective solutions to enterprises but will also provide evidence for further research in the healthcare domain [27]. The ease in scientific data collection by accessing it through devices like antennas and wireless devices has transformed the biotechnology sector by giving smart solutions for healthy lifestyle and wellbeing of people [28,29]. Thus, this study conducted by applying principles of AI will provide grounds to decision-makers to enhance the testing capacity within limited resources and to introduce new strategies to control and confine the pandemic.

2 Materials and methods

This section introduces the dataset used in this study along with the proposed methodology for conducting this research.

2.1 The dataset used

A dataset containing citation information must be collected before starting any bibliometric analysis. Therefore, data were retrieved from databases including PubMed, Scopus, and Web of Science by using keywords related to COVID-19 and was further evaluated.

2.2 Proposed methodology

The available information from different Internet sources was obtained for bibliometric analysis. It was further evaluated by using the ANN approach. The output information from ANN was obtained in the form of a map that was constructed by VOSviewer software. The output was further analyzed via MATLAB where probabilistic algorithms were applied to find out the probability of effectiveness of our proposed solution. A collective summary of all the components of the Artificial Intelligence and Machine Learning model adopted for this research is schematically depicted in Figure 1.

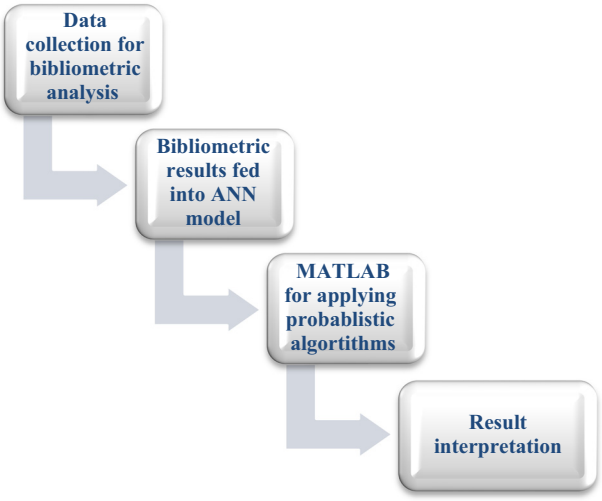

Figure 1: Workflow of study.

2.2.1 ANN model for bibliometric analysis

ANN approach was used in this study because of its effectiveness in processing huge volumes of data. The results obtained via ANN are more reflective, and therefore, it has also been employed in conducting this research. Furthermore, we used the training set to determine average evaluation for model parameters (studies including the ANN, Artificial intelligence, and Machine learning before covid). We then trained the model for 10 runs. The observed data (training data) contains “testing studies prior to covid and their

relevance implication" and the test data represented the studies after the global pandemic and their effective evaluation on the testing system.

We have utilized the results that were analyzed by constructing a map that represents the relationship of different nodes joined together by lines. In our study, nodes represent the research domains, their size represents the amount of work being done while the lines represent the interconnectedness of these research domains with COVID-19.

2.2.2 VOSviewer for constructing maps

VOSviewer is a computer program that helps in constructing and viewing maps generated with bibliometric data. As the name indicates, it constructs maps based on the Visualization Of Similarities (VOS). It has the advantage that it can run on multiple hardware devices as well as operating systems and even can be applied directly onto the Internet source. It generates two types of maps based on either distance between two entities that hold significance in terms of strength of the relation between them or graph-based maps in which the distance holds no significance and does not indicate any strength in their relation. Generally, constructing maps via VOSviewer needs three steps to be performed by calculating similarity matrix based on co-occurrence in the first step followed by applying the VOS mapping technique on similarity matrices in the second step and that is subjected to translation, rotation, or reflection in the last step. These steps are briefly discussed in the following sections.

2.2.3 Calculating similarity matrix

Similarity matrix serves as an input to VOS mapping technique and is obtained by normalizing co-occurrence matrix. The similarity between two entities i and j is given by the following formula,

$$S_{ij} = \frac{C_{ij}}{W_i W_j}.$$

Here, C_{ij} represents the co-occurrence of i and j and W_i and W_j represent a total number of occurrences of these entities.

2.2.4 VOS mapping technique

VOSviewer constructs a 2D map of n number of items based on a similarity matrix so that the distance between them determines the similarity of the items. This implies that items that are closely linked are mapped in close proximity while those sharing less similarity are placed in a distant position. The VOS mapping technique works on the concept of minimizing a weighted sum of squared Euclidean distances, and therefore, it assigns the average distance of one between two closely related entities. It is mathematically represented as follows:

$$V(x_1,\ldots, x_n) = \sum_{i<j} S_{ij}\|x_i - x_j\|.$$

Here, vector x_i represents the location of i on a 2-D map and $||\bullet||$ represents Euclidean norm

Minimization is performed by the following formula:

$$\frac{2}{n(n-1)} \sum_{i<j} \|x_i - x_j\| = 1.$$

In the first step, constrained optimization is changed into an unconstrained optimization problem that in the second step is subjected to majorization algorithms for further solving.

2.2.5 Translation, rotation, and reflection

As no unique and optimal solution exists for the optimization problem that is addressed in the previous step and consistent results must be produced every time when the same co-occurrence matrix is used for constructing maps via VOSviewer; therefore, three transformations of translation, rotation, and reflection should be applied [30].

3 Experiments and results

The experimental setup of this study and the results obtained are discussed in this section.

3.1 Experimental setup

3.1.1 Bibliometric analysis

We started this study by performing bibliometric analysis on a list of keywords that were most commonly used for retrieving information on COVID-19 testing. The keywords used for this research include, COVID-19 testing, COVID-19 testing through binary search, ANN approach for COVID-19 testing, AI in COVID-19 testing, etc. Based on these keywords, data were retrieved from different search engines including ScienceDirect, Web of Science, Scopus, and PubMed. Testing technologies published in top-tier journals were collected, and bibliometric analysis was performed to identify new research domains by carefully tracking the progress made toward COVID-19 diagnostics. The result of this bibliometric analysis was used as the input parameter for the ANN model.

3.1.2 ANN model

Multilayer ANN works by comparing different testing methodologies in three different layers: input layer, hidden layer, and output layer. The aim of using this complex network model was to find out the best testing approach for COVID-19 detection. The ANN model was used to identify the domain carrying benefits for COVID-19 testing by using VOSviewer software. This software made a map based on bibliometric information that shows different domains explored under COVID-19 testing. The weight of each input is represented by the size of the node, and the output is represented in the form of a network.

3.1.3 Mathematical algorithms

Information obtained from the ANN model was further analyzed. It was subjected to MATLAB that helped in statistically validating the probability of our proposed solution. Therefore, we started this study by taking samples in power of two, so that there can be a total of $l = \log_2 N$ levels. The model generated was elaborated by using a group of eight individuals. Figure 2 represents all eight samples arranged in four levels. These samples from S1 to S8 belong to eight different individuals and represent the lowest level, level 4. These samples are then paired to give a higher level, level 3, in which four pairs will be formed from eight samples.

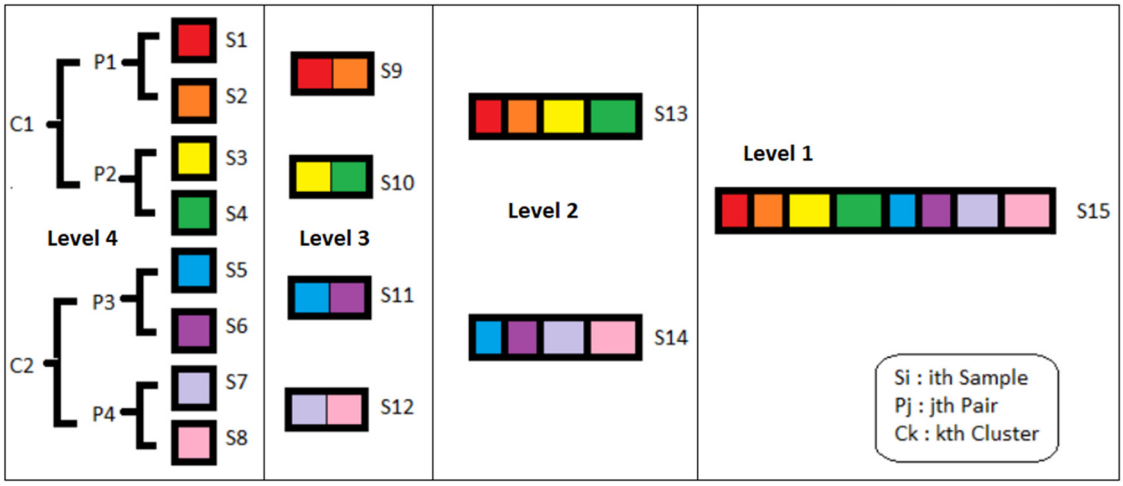

Figure 2: Sample pooling of eight samples in four levels.

Pairs from level 3 will again cluster to give level 2 with four samples in each cluster that is ultimately pooled up to give level 1 with all eight available samples. From level L_1 to L_{l-1}, each sample is formed by adding two child samples of succeeding level such that

$$S(i) = S(i - 2^L) + S(i + 1 - 2^L),$$

where $S(i - 2^L)$ and $S(i + 1 - 2^L)$ become child samples of $S(i)$

In this approach instead of testing all the samples in a linear fashion, we started from the highest level 1 with all samples, i.e., $S(2^L - 1)$ pooled, and if sample $S(i)$ comes out positive, both child samples of $S(i)$ will be tested. By this approach, a possible number of tests to diagnose eight people are t such that $t \in \{1, 7, 9, 11, 13, 15\}$. All the possible values of t are elaborated in the given figures.

As depicted in Figure 3, if all the samples pooled in $S(15)$ are COVID-19 negative, they will be tested by only a single test performed on this pool of samples. This will be the most effective way in which a minimum number of tests will be performed.

Figure 3: One test to be performed for testing eight individuals.

Figure 4 shows the number of tests to be performed if either both or one sample of a pair is positive for COVID-19. In such instances, it can be evaluated by performing seven tests starting from level 1 to level 4.

Figure 4: Seven tests to be performed for testing eight individuals.

Figure 5 represents a scenario in which nine tests will be performed to test eight individuals when either both samples or one sample from both pairs of the same cluster are detected positive.

Figure 5: Nine tests to be performed for testing eight individuals.

Figure 6 represents the situation when either both samples or samples from a single pair of both clusters are positive and can be tested by performing a total of 11 tests.

Figure 7 reflects the number of tests to be performed when one or both samples of three pairs from both the clusters are COVID-19 positive will be 13. A total number of tests performed will increase to 15 tests in situations when either both or a single sample from each pair of both clusters is positive. This is shown in Figure 8 and will give the maximum number of tests that will be performed from pooling samples from eight individuals.

All this information on sample pooling suggests that sample pooling for eight individuals will be valuable only when the probability of occurrence of positive cases is low. Therefore, it will be beneficial in epidemics or pandemics when the prevalence rate is low, and there are more chances of getting a negative result than a positive result. If p is the probability of a positive outcome of a test, the probability of the possible number of tests then becomes

Figure 6: Eleven tests to be performed for testing eight individuals.

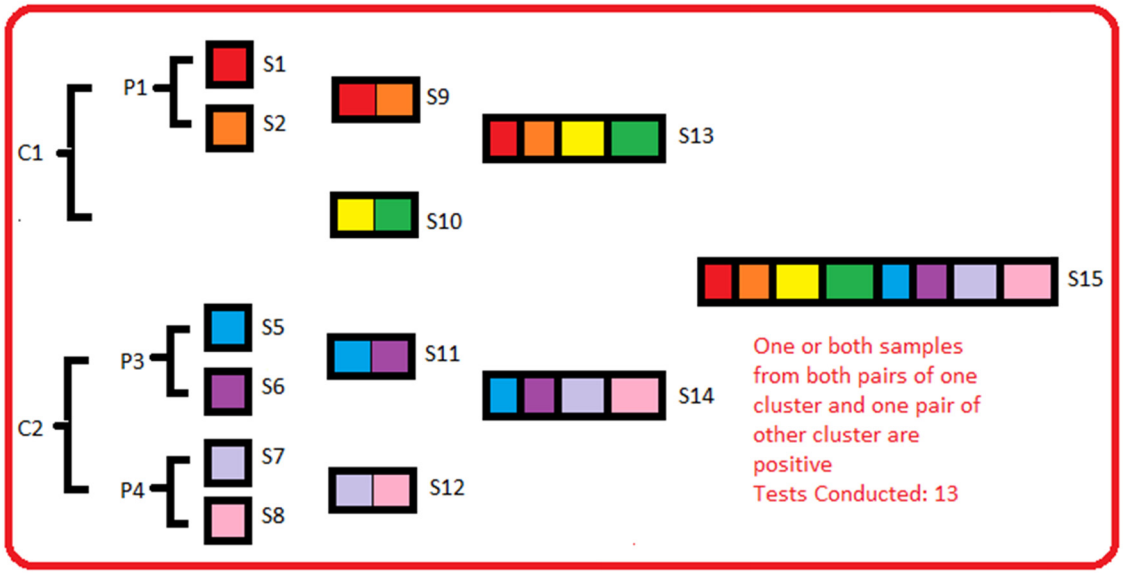

Figure 7: Thirteen tests to be performed to test eight individuals.

$$P[t] = \begin{cases} (1-p)^8; & t = 1 \\ 4(2p-p^2)(1-p)^6; & t = 7 \\ 2(2p-p^2)^2(1-p)^4; & t = 9 \\ 4(2p-p^2)^2(1-p)^4; & t = 11 \\ 4(2p-p^2)^3(1-p)^2; & t = 13 \\ (2p-p^2)^4; & t = 15. \end{cases}$$

The average number of tests $E[t_8]$ that this strategy takes as a function of probability p for a pool of eight people is given as follows:

$$E[t_8] = -2p^8 + 16p^7 - 56p^6 + 112p^5 - 144p^4 + 128p^3 - 88p^2 + 48p + 1.$$

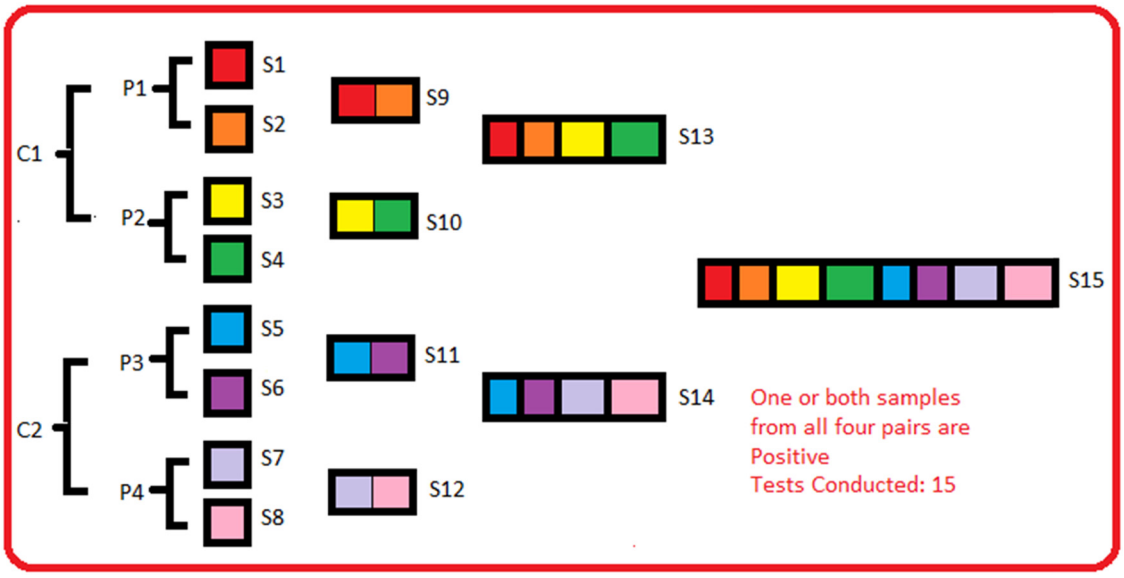

Figure 8: Fifteen tests to be performed to test eight individuals.

As discussed earlier, the number of samples can be taken in the power of two. Therefore, if the number of samples is increased from 8 to 16 samples, then a total of five levels will be made to evaluate the number of tests performed. Therefore, the number of tests performed for all negative cases will be one, and the maximum possible tests performed when either one or two samples of all the pairs are positive will be 31.

Hence, the maximum possible number of tests for 16 samples pooling (t_{16}) is given by

$$\in \{1,\ 9,\ 11, 13, 15, 17, 19,\ 21,\ 23,\ 25,\ 27,\ 29,\ 31\}.$$

The probability for the possible number of tests performed by pooling samples of 16 individuals is discussed by the given algorithm.

$$P[t_{16}] = \begin{cases} (1-p)^{16}; & t = 1 \\ 8(2p-p^2)(p-1)^{14}; & t = 9 \\ 4(2p-p^2)^2(p-1)^{12}; & t = 11 \\ 8(2p-p^2)^2(p-1)^{12}; & t = 13 \\ 8(2p-p^2)^3(p-1)^{10} + 4(2p-p^2)^2(p-1)^{12}; & t = 15 \\ 2(2p-p^2)^4(p-1)^8 + 4(2p-p^2)(2p-p^2)(p-1)^{10}; & t = 17 \\ 4(2p-p^2)^4(2p-p^2)^8 + 8(2p-p^2)^2(2p-p^2)(p-1)^{10}; & t = 19 \\ 16(2p-p^2)^4(p-1)^8 + 8(2p-p^2)^3\,4(2p-p^2)(p-1)^8; & t = 21 \\ 16(2p-p^2)^5(p-1)^6 + 16(2p-p^2)^4(p-1)^8 + 2(2p-p^2)^4\,4(2p-p^2)(p-1)^6; & t = 23 \\ 4(2p-p^2)^6(p-1)^4 + 32(2p-p^2)^5(p-1)^6; & t = 25 \\ 24(2p-p^2)^6(p-1)^4; & t = 27 \\ 8(2p-p^2)^7(p-1)^2; & t = 29 \\ (2p-p^2)^8; & t = 31. \end{cases}$$

Mean for 16 person's sample becomes

$$E[t_{16}] = \begin{cases} (p-1)^{16} + 232(2p-p^2)^7(p-1)^2 + 748(2p-p^2)^6(p-1)^4 + 1,168(2p-p^2)^5(p-1)^6 \\ + 814(2p-p^2)^4(p-1)^8 + 120(2p-p^2)^3(p-1)^{10} + 148(2p-p^2)^2(p-1)^{12} \\ + 15(2p-p^2)^2(p-1)^{12} + 31(2p-p^2)^8 + 72(2p-p^2)(p-1)^{14} + 46(2p-p^2)^{16}(2p-p^2)(p-1)^6 \\ + 168(2p-p^2)^3 4(2p-p^2)(p-1)^8 + 220(2p-p^2)^2 4(2p-p^2)(p-1)^{10}. \end{cases}$$

By using these above-mentioned algorithms, the probability of effectiveness of our proposed solution was evaluated for sample pools of eight individuals and 16 individuals.

4 Result analysis

4.1 Bibliometric analysis and ANN model

Bibliometric data fed into the ANN model were obtained for analysis through the output layer in the form of a map by using VOSviewer software. The map shows the most searched topics with reference to COVID-19 and is given in Figure 9.

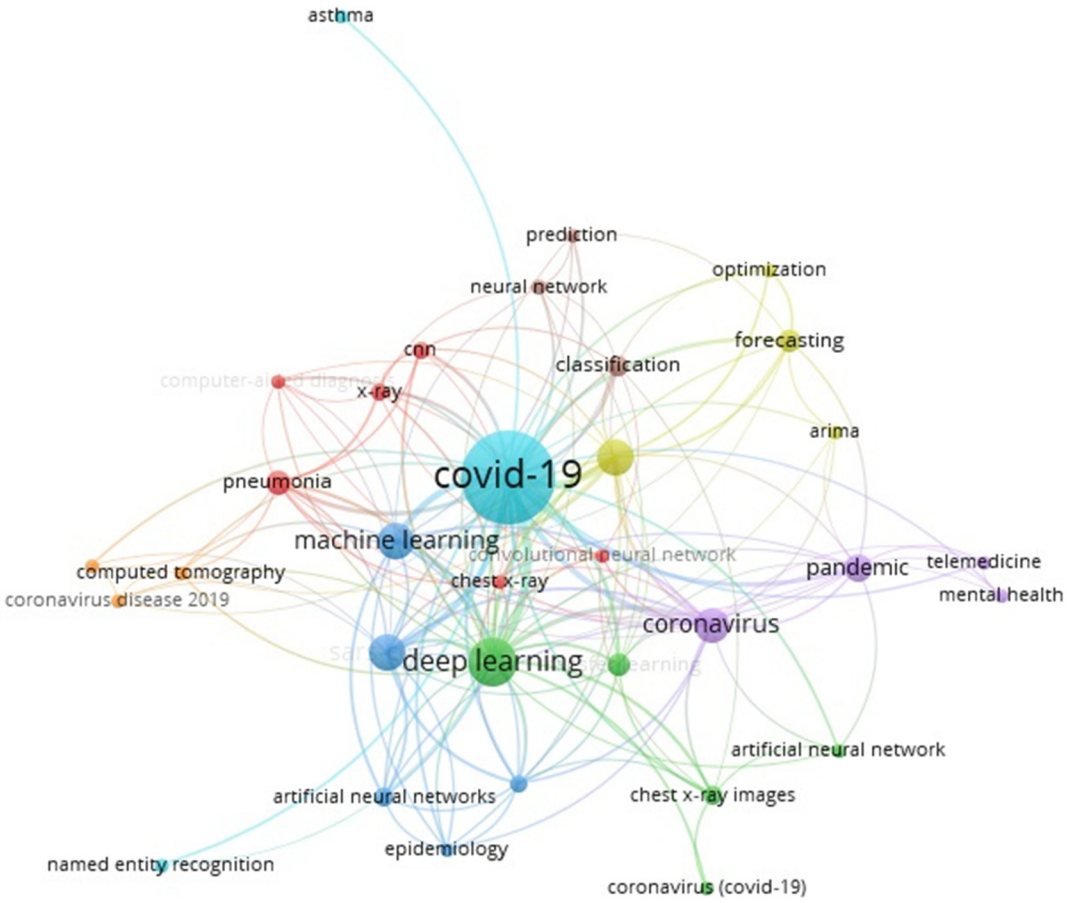

Figure 9: Keywords found in the literature related to COVID-19 testing.

The nodes on this map depict that the most repetitive keywords found in the literature retrieved through different search engines were COVID-19, machine learning, deep learning, ANN, pandemic, chest X-ray, computed tomography, etc. This shows the data collected for this research revolved around the COVID-19 testing and involves the application of AI and ML in controlling the pandemic. It can be seen in Figure 9 that CT scans and chest X-rays were already being under research for their role in identifying COVID-19. Therefore, after careful analysis, it was decided to improve the working capacity of a gold

standard of COVID-19 testing that is RT-PCR. Though area of research of many artificial intelligence-based studies revolved around chest X-rays and CT scans, but due to their limitations and inefficiency in correctly detecting COVID-19, we evaluated a different testing strategy focusing on pooling swabs for COVID-19 diagnosis.

4.2 Comparison and analysis of COVID-19 linear and pooled testing

Probabilistic results were obtained by applying the above-mentioned algorithms. The results obtained are discussed here.

Figure 10a demonstrates the average tests that need to be performed for both linear and pooled testing, when the probability of getting a positive result increases. For linear testing, it can be seen that number of tests is neither increasing nor decreasing for both eight individuals testing and 16 individuals testing. However, a rising curve is obtained for pooled testing that increases sharply in the beginning and then after a certain point attains a steady state. This shows that when the probability of getting a positive result is as low as 20%, the average number of tests to be performed on a pooled sample is lower than the number of tests performed for linear testing. After 20%, it can be seen that number of tests for pooled testing increases than for linear testing, making it the least viable solution in cases where the probability of getting a positive sample is higher.

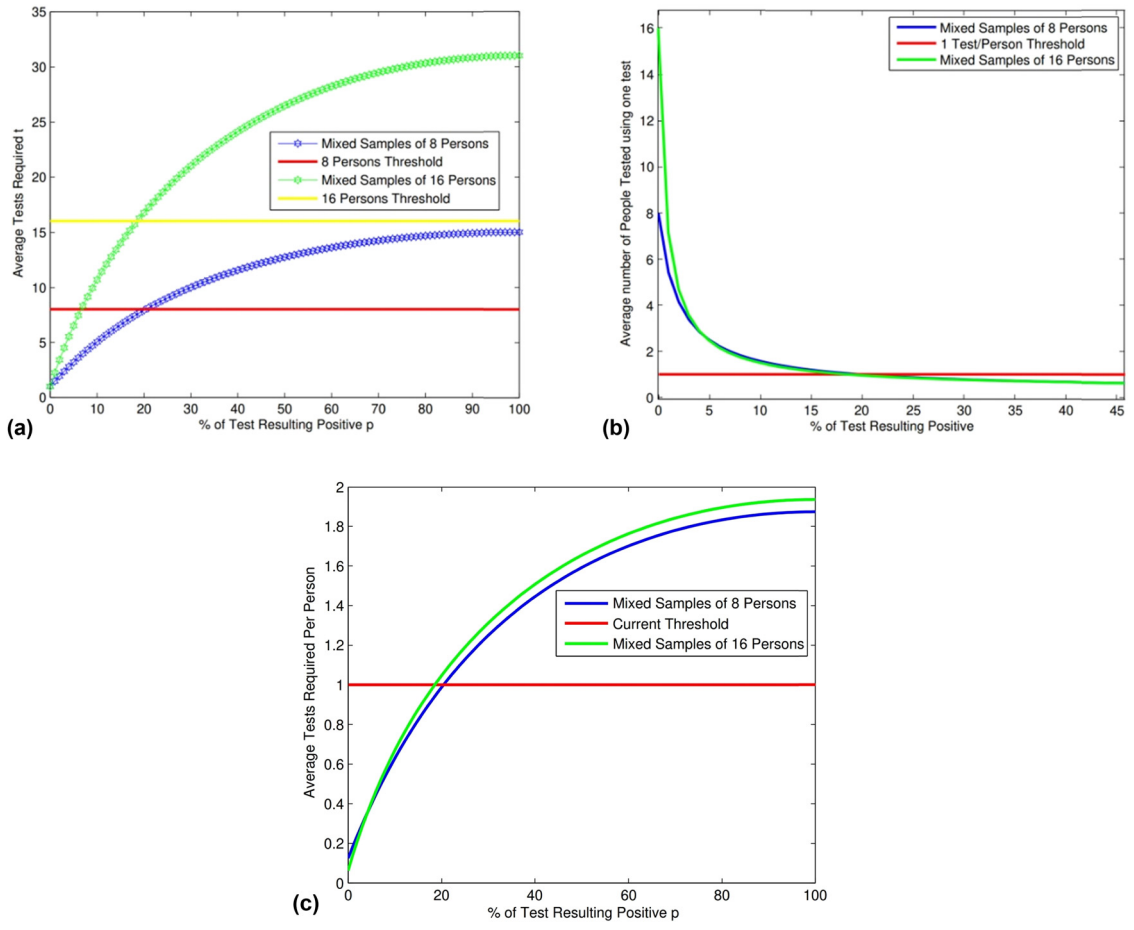

Figure 10: (a) Average tests are required for linear and pooled testing. (b) An average number of people tested with one test. (c) The average number of tests required per person for linear and pooled testing.

Figure 10b represents the number of people that can be tested by using only one test when the probability of getting a positive result increases. It can be seen that for linear testing a straight line is obtained showing that only one person will be tested with one test. However, for pooled testing, a sudden decline in the curve can be seen that shows that when the probability of getting a positive is as low as 0%, one test is sufficient to test a pooled sample of either 8 individuals or 16 individuals. However, as the probability starts increasing, the number of tests that should be performed to test samples would also increase. The probability at the point of intersection of all three lines indicates that sample pooling would be an effective and time-saving technique before that point of 19% probability. But, if probability will increase beyond that point, sample pooling will lose all of its effectiveness, and therefore, it will no longer be a recommended testing approach.

Figure 10c compares a number of tests required per person for linear testing with pooled testing. A straight line for linear testing represents that one test will be required to test one person irrespective of the probability of getting a positive or negative result. However, for pooled samples, a rising curve is obtained that rises sharply at the beginning when the probability of getting a positive result is low. These rising curves show that the average number of tests per person performed on pooled samples will be low when probability will be low and as soon as the probability increases to 19%, after that, it is no longer significant and the average test per person will start to increase in that case. The point of intersection of these lines demonstrates that sample pooling will be effective only when the probability of getting a positive is low.

All these results suggest that sample pooling holds potential benefits over linear testing, but only in instances when the probability of getting a positive result is low. This suggests that for diseases having a low prevalence rate in different communities, sample pooling would be an effective technique for detecting the disease and taking timely actions. As the prevalence of COVID-19 is low in many countries, this study provides a statistically proven solution, demonstrating the effectiveness of sample pooling in maximizing the number of individuals to be tested.

4.3 Case study based on our proposed algorithm

Adopting this approach, a case study was conducted in the Punjab region when positive detection probability was reported to be 6%. In this case study, several tests were conducted with the group pooling of 16 samples in each test. The abovementioned algorithms were applied, and the probability against each case was evaluated. The graph in Figure 11 shows percentages of tests to be performed for each sample pool.

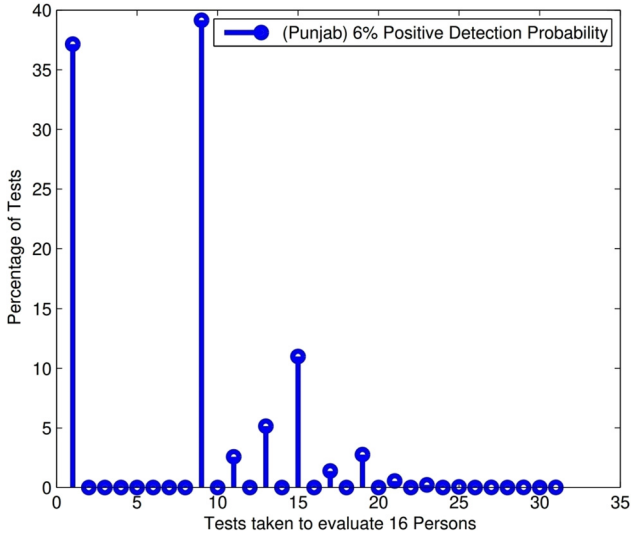

Figure 11: Population-based study with seven sample pools of 16 individuals each.

All these results obtained from applying algorithms to find out the efficiency of our proposed solution suggest that for population-based studies for diseases that have a low prevalence rate, sample pooling is an effective means of increasing the capacity of testing with minimum available resources. As COVID-19 prevalence is low in many communities, our study will provide scientific grounds to incorporate better and improved solutions in the healthcare sector.

4.4 Application of the algorithm on other countries

Evidently, the probability of the number of tests increases with the Covid positive occurrence for each state mentioned (Figure 12). There is a 98% probability for New Zealand and Australia to have detected a Covid case with the sample pooling suggested above. With the increase in Covid Positive ratio, there is a decline in the percentage of positive test detection. The minimum number of tests performed on a high positive ratio for 16 people pool is 10 with the probability of approx. 40% for USA. For Germany, there need to be 10 repetitions of the test to reach a 30% probability. The recent data available from the source mentioned in Figure 12 can be imported into the supporting file to extrapolate results.

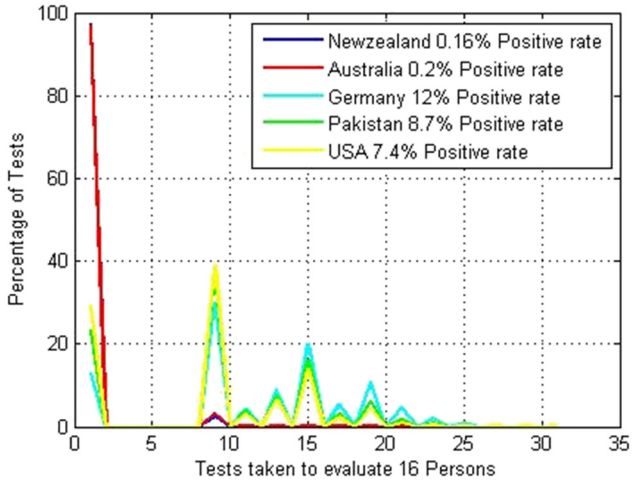

Figure 12: Test predictions for five major countries hit as of 21/04/2021(The COVID +ve ratio is adapted from https://ourworldindata.org/coronavirus-testing).

5 Discussion

Though pandemics are challenging and require a substantial relocation of resources, energies, and capital, but they also bring with them an opportunity for resilience and innovation. As science has progressed rapidly since the last pandemic, this has improved healthcare facilities as well. But this must not undermine the adversities of a challenging situation. Therefore, it is advisable for healthcare systems and enterprises to timely take measures and devises strategies to combat unpleasant circumstances that may last for months or even years in some instances. COVID-19 has made us evaluate our current standing in a pandemic and also how current interdisciplinary knowledge can bring efficient yet simple solutions. Testing, being the first step of identifying and controlling the pandemic, needs a special focus so the concerned authorities could timely take action even with the least available resources [31]. Therefore, in the quest to save time and improve testing capacity, this study was conducted using the ANN approach, and by applying Machine Learning tools, we have found that disease burden would reduce considerably if testing will be done by

sample pooling with either 16 samples or eight samples for diseases that have low prevalence rate in different populations.

Statistical algorithms were employed in this regard to evaluating the effectiveness of our hypothesis, and the results supported our proposed solution. It has shown that testing would be reduced to only five levels in a pool of 16 individuals and four levels with eight individuals in each pool. As in most of the communities, the spread of COVID-19 and its, prevalence is low, therefore, this will support both the healthcare sector and government regulatory bodies by performing a greater number of tests with minimal testing resources available saving time and finances, and making tracking of COVID-19 and containing more robust. Though sample pooling will be an effective strategy in increasing the testing capacity, before integrating it with healthcare settings, it is important to evaluate few negative aspects of this technique as well. The testing specificity must be validated via conducting tests on sample pools from community-acquired samples. Though we found that sample pooling can be done when samples are taken in the power of two, but it should also indicate the maximum limit for sample pooling where a positive case can be detected without diluting out. Supporting our findings, a study was performed which demonstrated that the specificity of testing did not alter even with sample pools of 32 and 64 individuals [32], but it might need some extra amplification cycles. Sample pooling must be done carefully as a viral load from recently acquired COVID-19 individuals or those recovered from COVID may not be detected due to low levels giving a false-negative. All these concerns must be carefully addressed before opting sample pooling technique that no doubt has lots of benefits, but the associated potential risks cannot be neglected. Therefore, we can say that the use of Artificial Intelligence and Machine Learning in biological research during a pandemic will provide a basis for answering research questions and giving smart solutions for improved healthcare facilities but these solutions must be validated in labs for their effectiveness before actually implementing them. This will provide safe grounds, increased working capacity, and reduced financial losses to the biological community in such unprecedented times.

In the end, all these benefits of sample pooling do not lie only for the healthcare sector, but they are projected toward enterprises as well. By using such algorithms, they can enhance their own testing capacity and can provide smart solutions to hospitals, nursing homes, prisons, and cruise ships, where due to confined space and insufficient resources for testing, the probability of contracting COVID-19 increases considerably. The increased demand for reagents used in PCR-based testing, pharmacological solutions, and PPE and their burden on manufacturing and pharmacological industries would be reduced, and these companies will be able to meet public demands with the least available resources. Moreover, smart lock-downs and increased testing capacity will serve both local communities and the corporate sector in resuming the functional activities and overcoming the losses of the previous fiscal year. This study provides a scientifically proven solution that, if adopted, will serve all the sectors of society that are hit hard by the pandemic.

This study is based on the principles of Artificial Intelligence and provides the probable solution for the COVID-19 testing; therefore, it holds the limitation of not incorporating the results of wet lab work. Due to this reason, it is advisable to validate it through a pooling of community samples for PCR-based testing before implementing at mass level testing.

6 Conclusion

Artificial Intelligence and Machine Learning have contributed significantly to providing efficient solutions to healthcare units even in pandemics. Integrating AI-based solutions into healthcare facilities will improve the working capacity of these units. Therefore, research has been going on in this domain to provide smart solutions to both healthcare settings and enterprises. This study was also conducted to utilize the effectiveness of AI and ML solutions in the healthcare sector and by using the approach of ANN, found sample pooling to be an effective strategy for COVID-19 testing when available resources are limited. Our results demonstrated sample pooling would be a cost-effective COVID-19 testing solution as the testing resources

are scarce in many communities. The low prevalence rate of COVID-19 in many communities makes our solution easy to adopt because chances of getting a positive case are low when carrying out a mass testing program. Therefore, sample pooling would help in improving testing capacity that would ultimately help in detecting and tracking more robust. Moreover, our findings will also help in minimizing the utilization of testing reagents that would release the burden of increased consumer demand from manufacturing enterprises as well. Our study provides scientifically proven algorithms for COVID-19 testing that if implemented would help in containing the spread of the pandemic and ultimately resuming the activities back to normal. However, these algorithms further need to be tested on community-acquired samples before finally implementing in laboratories. This study can further be improved with the implementations of current data into the proposed methodology and running large-scale simulations. We recommend more methodological work to address the strategic public behavior and involvement with State entities and economic consideration to see the impact of the research.

Acknowledgments: We express special appreciation to the industry and the authors for their help with the illustration, latest tools in the market, Programing of the simulations, and platform to perform this research presented in this article.

Funding information: Funding for this study was provided by the Myst Enterprise of Canterbury Province, Christchurch, New Zealand to Mustafa Pasha. Mustafa Pasha received the funding (MYS/0022/20) for the purposes to be used in man-hours and collaborative meetings from three different regions of the world including, Australia, New Zealand, Pakistan, and China.

Conflict of interest: There are no potential conflicts of interest found in the making of this article.

References

[1] Joglekar N, Parker G, Srai J. Winning the race for survival: how advanced manufacturing technologies are driving business-model innovation. Switzerland: World Economic Forum; 2020. Available at SSRN 3604242.
[2] Yawson R. Strategic flexibility analysis of HRD research and practice post COVID-19 pandemic. Hum Resour Dev Int. 2020;23:406–17.
[3] Spurk D, Straub C. Flexible employment relationships and careers in times of the COVID-19 pandemic. J Vocat Behav. 2020;119:103435.
[4] Solis J, Franco-Paredes C, Henao-Martínez AF, Krsak M, Zimmer SM. Structural vulnerability in the US revealed in three waves of COVID-19. Am J Trop Med Hyg. 2020;103:25–7.
[5] Wells CR, Townsend JP, Pandey A, Moghadas SM, Krieger G, Singer B, et al. Optimal COVID-19 quarantine and testing strategies. Nat Commun. 2021;12:1–9.
[6] Munawar HS, AAwan AA, Khalid U, Munawar S, Maqsood A. Revolutionizing telemedicine by instilling H. 265. Int J Image Graph Signal Process. 2017;9:20–7.
[7] Seetharaman P. Business models shifts: impact of Covid-19. Int J Inf Manag. 2020;54:102173.
[8] Shen W, Yang C, Gao L. Address business crisis caused by COVID-19 with collaborative intelligent manufacturing technologies. IET Collab Intell Manuf. 2020;2:96–9.
[9] Bartik AW, Bertrand M, Cullen Z, Glaeser EL, Luca M, Stanton C. The impact of COVID-19 on small business outcomes and expectations. Proc Natl Acad Sci. 2020;117:17656–66.
[10] Scheiber N. The New York Times website [Online]. https://www.nytimes.com/2020/09/15/business/economy/employers-coronavirus-testing.html?auth=link-dismiss-google1tap. (The New York Times September 16, 2020. [Cited: December 21, 2020]).
[11] Chang L, Yan Y, Wang L. Coronavirus disease 2019: coronaviruses and blood safety. Transfus Med Rev. 2020;34:75–80.
[12] Cleverley J, Piper J, Jones MM. The role of chest radiography in confirming covid-19 pneumonia. BMJ. 2020;370:24–6.
[13] Abdulkareem KH, Mohammed MA, Salim A, Arif M, Geman O, Gupta D, et al. Realizing an effective COVID-19 diagnosis system based on machine learning and IOT in smart hospital environment. IEEE Internet Things J. 2021;1–8.
[14] Al-Waisy A, Mohammed MA, Al-Fahdawi S, Maashi M, Garcia-Zapirain B, Abdulkareem KH, et al. COVID-DeepNet: hybrid multimodal deep learning system for improving COVID-19 pneumonia detection in chest X-ray images. Comput Mater Contin. 2021;67(2):2409–29.

[15] Al-Waisy AS, Al-Fahdawi S, Mohammed MA, Abdulkareem KH, Mostafa SA, Maashi MS, et al. COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. Soft Comput. 2020;1–16.
[16] Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. Chaos Soliton Fract. 2020;139:110059.
[17] Mohammed MA, Abdulkareem KH, Garcia-Zapirain B, Mostafa SA, Maashi MS, Al-Waisy AS, et al. A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of covid-19 based on x-ray images. Comput Mater Contin. 2020;66:3289–310.
[18] Mohammed MA, Abdulkareem KH, Al-Waisy AS, Mostafa SA, Al-Fahdawi S, Dinar AM, et al. Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and TOPSIS methods. IEEE Access. 2020;8:99115–31.
[19] Mei X, Lee H-C, Diao K-y, Huang M, Lin B, Liu C, et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. Nat Med. 2020;26:1224–8.
[20] Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology. 2020;296(2):1–15
[21] Mashamba-Thompson TP, Crayton ED. Blockchain and artificial intelligence technology for novel coronavirus disease-19 self-testing. Switzerland: Diagnostics, MDPI; 2020.
[22] Vaishya R, Javaid M, Khan IH, Haleem A. Artificial intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab Syndr Clin Res Rev. 2020;14:337–9.
[23] Munawar HS. Flood disaster management: risks, technologies, and future directions. Mach Vis Inspect Syst Image Process Concept Methodol Appl. 2020;1:115–46.
[24] Maind SB, Wankar P. Research paper on basic of artificial neural network. Int J Recent Innov Trends Comput Commun. 2014;2:96–100.
[25] Mair C, Kadoda G, Lefley M, Phalp K, Schofield C, Shepperd M, et al. An investigation of machine learning based prediction systems. J Syst Softw. 2000;53:23–9.
[26] Sunjaya AF, Sunjaya AP. Pooled testing for expanding COVID-19 mass surveillance. Disaster Med Public Health Preparedness. 2020;14:e42–3.
[27] Munawar HS, Qayyum S, Ullah F, Sepasgozar S. Big data and its applications in smart real estate and the disaster management life cycle: a systematic analysis. Big Data Cognit Comput. 2020;4:4.
[28] Munawar HS. An overview of reconfigurable antennas for wireless body area networks and possible future prospects. Int J Wirel Microw Technol. 2020;10:1–8.
[29] Munawar HS. Applications of leaky-wave antennas: a review. Int J Wirel Microw Technol. 2020;10(4):56–62.
[30] Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84:523–38.
[31] Rajan S, Cylus JD, Mckee M. What do countries need to do to implement effective 'find, test, trace, isolate and support'systems? J R Soc Med. 2020;113:245–50.
[32] Yelin I, Aharony N, Tamar ES, Argoetti A, Messer E, Berenbaum D, et al. Evaluation of COVID-19 RT-qPCR test in multi sample pools. Clin Infect Dis. 2020;71:2073–8.