

Minimizing the usage of SARS-CoV-2 lab test resources through test pooling enhanced by classification techniques

Ana Cristina Bicharra Garcia
 Univ. Federal do Estado do Rio de Janeiro
crisrina.bicharra@uniriotec.br

Márcio de Oliveira Barros
 Univ. Federal do Estado do Rio de Janeiro
marcio.barros@uniriotec.br

Abstract

Testing is an effective practice to limit the spread of the SARS-CoV-2. PCR is an accurate method to detect SARS-CoV-2 infected individuals, but PCR lab test kits are scarce and expensive resources. Therefore, their usage should be optimized. Testing in batch (pooling) is a procedure that merges individuals' swabs, allowing group diagnosis without affecting the accuracy of the results. Savings on test kits depend on the prevalence of the disease, pool composition and size. We propose a novel approach for optimizing pooling to minimize the usage of lab test kits. We show that estimating the probability of an individual being infected by means of a binary classifier leads to improvements in the efficiency of pooling strategies. We use simulation to select the components of a new pooling strategy based on a classifier and evaluate our approach using a real dataset.

1. Introduction

Testing is an essential component of any strategy to contain the COVID-19 pandemic. Kronbichler et al. [1] have shown that more than 20% of the infected individuals have been asymptomatic. This number is expected to be much higher due to undocumented cases. Asymptomatic infected individuals unwillingly become a contamination vehicle for the SARS-CoV-2 virus. The more individuals are tested and more frequently, the sooner it is possible to identify the contaminated ones and isolate them for containing the disease spread. On the other hand, massive testing for SARS-CoV-2 is expensive.

There are two types of diagnostic tests for detecting the SARS-CoV-2 virus: molecular and antigen tests. In our study, we focus on the molecular test, the Polymerase Chain Reaction (PCR) for being the gold standard in clinical practice [2]. PCR aims to detect the SARS-CoV-2 genetic material in an individual nose or throat swab sample. The swab collected samples are tested by using test kits with PCR reagents.

While an individual test kit may be cheap, performing frequent mass testing becomes an economic challenge. Moreover, there are manufacturing constraints to have the test kits available on a large scale. Even small countries may find it challenging to source the material components required for testing [3] and skilled human resources to apply them, let alone fund this endeavor.

If a solution requires expensive and rare resources, it might benefit from optimization techniques. SARS-CoV-2 testing optimization can be performed in two fronts: (i) selecting the right people to be tested, based on their profiles and interactions; and (ii) grouping individuals into pools and testing the pools instead of the individuals to reduce the consumption of lab time and test kits. This paper addresses the second front. Pooling is a known strategy for optimizing scarce resources, such as staff for performing a service [4], smart city resources [5], and lab test kits for diagnoses [6] [7].

This paper presents a new approach for pooling based on a pre-classification phase to guide the allocation of individuals into the pools during the pooling phase. Tournament pooling was selected as the most effective among a set of alternative strategies as revealed by simulation across several pools sizes and different prevalence of SARS-CoV-2 in the population. We further test the approach under distinct theoretical classifiers through simulation and apply it to a real dataset.

Our contributions are threefold. First, we review pooling strategies currently in use. Then, we demonstrate the usefulness of a binary classifier to support selecting individuals to compose the pools. Finally, we present a case study using our pooling strategy with a public dataset from a Brazilian hospital. Results demonstrate that our approach brings an improvement of up to 14.4% of the number of required lab test kits over the current strategies, even using a classifier trained with a small dataset. Also, the approach can be fine-tuned with new data as it becomes available.

This paper is organized into five sections, starting from this introduction. Section 2 presents background information about pooling strategies and classi-

fiers. Section 3 presents our proposed pooling strategy, while Section 4 presents the case study. Finally, Section 5 presents the conclusions.

2. Theoretical Background

2.1. Pooling strategies

A *pooling strategy* is a method to construct a set of pools [8]. Each pool is comprised of a set of *individuals* from a larger population that needs to be tested. Let the population be denoted by $I = \{I_1, I_2, \dots, I_n\}$, being n the number of individuals. Each individual, I_i , is a swab material either infected or not infected with the SARS-CoV-2 virus. The PCR test can detect evidences of the virus from a collection of swab material grouped in a pool. A pool composed of k individuals requires one test kit. If the test result is negative, the pooling process produces a $k - 1$ savings. On the other hand, if there is at least one individual infected in the pool, the pooling process will require $k + 1$ test kits.

Some pooling strategies involve redundant tests for some or even all individuals, allowing them to identify false-positive and false-negative cases. A false-negative occurs if an individual that should be classified as positive is tested negative. False-negatives are observed due to human errors, using low-quality reagents, material collection errors, and low concentration of the expected substance in the collected material. False-positives occur if an individual that might be classified as negative is tested positive. Besides collection and processing errors, false-positives occur when the testing procedure confounds the substance that characterizes the disease with another reactant in the collected materials.

The pooling strategies below were evaluated because they are the most well-known approaches for pooling [9] or represent state-of-the-art procedures [8]. Tournament pooling is also used for SARS-CoV-2 testing in Brazilian labs. A recent article shows how these strategies can be combined and used in SARS-CoV-2 testing [10].

Tournament testing: in this strategy, the individuals to be tested are distributed into non-overlapping pools of fixed size [9]. If a pool tests negative, its individuals are marked negative. Otherwise, they are undetermined. Several tournaments rounds, with constant or decreasing pool size, may be executed in sequence to handle the undetermined individuals. Figure 1 shows two tournament rounds performed to find a positive individual. This strategy does not rely on redundancy and cannot identify nor correct testing errors. Hogan et al. [11] report that a two-round tournament strategy was used in the early days of the SARS-CoV-2 in the San Francisco bay area to save resources. Abdalhamid et al. [12]

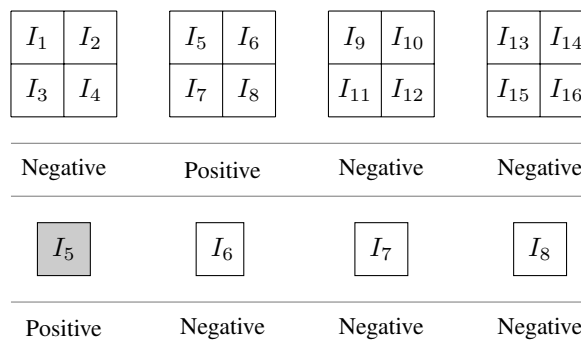


Figure 1. A tournament pooling strategy example. Sixteen individuals are distributed into four pools of equal size. Tests reveal that some pools are negative and the remaining individuals are retested in isolation to identify the real positive cases.

also used a tournament strategy with retesting in isolation to find optimal parameters for pooling testing for SARS-CoV-2. Regen et al. [13] also discuss the optimum pool size for detecting SARS-CoV-2 as a function of the population's prevalence.

Full grid testing: this strategy separates the individuals into pools of fixed size s , where $s = l^2$, $l \in \mathbb{Z}$. The individuals on each pool are organized in a square matrix of l rows. One test is performed for each line, as well as for each column of the matrix. Thus, the material collected from each individual is used twice. An individual is negative if both pools in which he/she participates are negative. An individual is positive if both pools on which he/she participates are positive and all other individuals in these pools are negative. Otherwise, the individual is undetermined. The redundancy incurred by using each individuals' materials twice allows detecting errors in the tests. For instance, if a false-negative case occurs in the first pool containing individual I_1 but not in his second pool, he will be undetermined. Verdun et al. [14] report that grid testing can be more efficient than tournament strategies in a context with a high prevalence of SARS-CoV-2. Mutesa et al. [15] applied a full grid strategy in positive pools after tournament testing instead of retesting these individuals in isolation to increase the savings in lab test kits in Rwanda.

Grid border testing: this strategy separates the individuals into pools of fixed size s , where $s = 4 \times (l - 1)$, $l \in \mathbb{Z}$. The individuals on each pool are organized in the borders of a square matrix of l rows. One pool is created for the first row of the matrix, one for the last row, one for the first column, and one for the last column. An individual is negative if all pools in which he/she participates are negative. An individual is positive if all pools

on which he/she participates are positive and all other individuals in these pools are negative. Otherwise, the individual is undetermined. Four out of s individuals participate in two pools, while the remaining ones take part in a single one. Therefore, redundancy is limited in this strategy, as only the individuals in the corners of the matrix are tested twice. We did not find applications of grid border testing in the context of SARS-CoV-2.

Shifted Transversal Design (STD): this strategy is based on the concept of layers [8] and minimizes the number of times a pair of individuals appear together in a pool while maintaining the number of individuals on each pool roughly equal [16]. STD is defined by three parameters, $STD(n, q, k)$, that denote the number of individuals to be tested (n), the number of pools on each layer (q), and the number of layers (k). Each individual is tested once for each layer. The number of pools in a layer must be a prime number and its minimum value depends on the expected number of positive individuals and testing errors. The number of layers is arbitrary, but having $k = q + 1$ guarantees that all pairs of individuals will be tested together on at least one pool. The amount of redundancy in this strategy and its ability to detect and correct errors depends on selecting a proper number of layers and pools in these layers. Mallapaty [10] reports the usage of pooling strategies similar to STD in the context of SARS-CoV-2 on India and Israel.

2.2. Binary classifiers

A *classifier* is an algorithm that assigns a data point (an individual in a population) to a class among a discrete number of classes, according to the attributes' values of this data point. A classifier learns its behavior from historic data from which it derives patterns that maps attributes values to labels or classes.

Binary classifiers use supervised machine learning techniques to derive a classification model that predicts the existence or the absence of a given condition. The condition under interest for this research is being infected with the SARS-CoV-2 virus. Based on a set of individual profiles and the related results of PCR tests, the classifier tries to identify which attributes are important to separate infected individuals from healthy ones. Formally, let X be the set of attributes describing the population and let $x_{i,j}$ represent the value of attribute j for the individual i . Let $y_i \in \{0, 1\}$ be the classification value for individual i . The learning process produces a function f that maps X to Y and whose output is the probability of an individual belonging to a class. A threshold value for this probability indicates whether the individual is classified as infected or healthy.

There are many techniques to build binary classifiers, including heuristic rules, logistic regression, decision trees, random forest, multi-layer perceptron, and deep learning. As the SARS-CoV-2 scenario is new, there is close to no expertise available. There are few labeled datasets with a reasonable number of cases, but not big data. To test our approach, we selected three machine learning techniques: logistic regression [17], random forest [18], and multi-layer neural networks [19].

Logistic Regression: this is a statistical method for finding an equation that best fits the mapping between the set of attributes to a binary output (classification) [17]. It generates the b_s coefficients of a *logit* function. The *logit* function computes the odds of an individual on a given class using Equation 1, in which p is the probability of the condition showing in the individual. The logistic regression aims to maximize the likelihood of observing the sample values.

$$\begin{aligned} \text{logit}(p) &= b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n \\ \text{logit}(p) &= \ln\left(\frac{p}{1-p}\right) \end{aligned} \quad (1)$$

Random Forest: this is an ensemble method composed of a set of estimators that run in parallel to reach a classification [18]. Each estimator is a decision tree and the final result is reached by a majority among the results achieved by the independent trees. Each decision tree partitions the data into two subsets, according to the value of an attribute. On its turn, the random forest classifier divides its data into possibly overlapping sets that are distributed to train independent trees. There are two approaches for such data distribution: *bootstrapping* or *feature randomness*. Bootstrapping randomly selects individuals from the training data with replacement and distribute these individuals to the decision trees. Thus, each tree receives a subset of individuals with all their attributes. Feature randomness, on its turn, randomly selects attributes to be used by each decision tree in the random forest. Thus, each tree takes into consideration the whole population, though analyzing only part of the properties that describe its individuals.

Multi-layer perceptron: it is a neural network with an input layer, one or more hidden layers, and an output layer [19]. Nodes in the input layer are selected attributes describing the individuals in the dataset. The output layer of binary classifiers has a single node. Nodes in the hidden layers receive a set of stimuli (x_i) from the outputs of the nodes in the preceding layer, aggregate these stimuli considering their importance (w_i), and transform

the aggregated stimuli, by applying an activation function, into an output. If the output of the node is larger than a threshold value, it is propagated to the next layer. Otherwise, the propagation is refrained.

Training a multi-layer perceptron model is the task of finding values for the importance (w_i) of the inputs of the hidden layers that best fit the training set. The output of a multi-layer perceptron acting as a binary classifier is a probability. An individual is classified as having the condition if the probability produced by the perceptron when the individual's characteristics are presented in the input layer crosses a given threshold. The attributes selected to build the input layer may affect the quality of the trained model. Consequently, it is advisable to investigate the dataset before training. Configuring the architecture of the perceptron (number of hidden layers, nodes in each layer, and so on) is so far an artisan task.

2.3. Performance metrics

The performance of binary classifiers is evaluated according to the *specificity* and *sensitivity* metrics, which are derived from the count of true-positives (TP), false-positives (FP), true-negatives (TN), and false-negatives (FN). In opposition to false-positives (or false-negatives), true positives (or true negatives) represent cases on which the classifier correctly suggests that an individual is positive (or negative).

Sensitivity is calculated by Equation 2 and represents the classifier's ability to predict the presence of the condition, for instance, the ability to classify individuals as sick when they are sick. Specificity is calculated by Equation 3 and represents the classifier's ability to predict the absence of the condition, that is, the ability to classify individuals as not sick whenever they are healthy. Verdun et al. [14] calculated the expected specificity and sensitivity of five pooling strategies for SARS-CoV-2 testing.

The $F_{measure}$ is a harmonic combination of specificity and sensitivity, calculated by Equation 4. It represents a balance between the opposing metrics of specificity and sensitivity. Another metric derived from specificity and sensitivity is the area under the ROC curve (AUC), which represents the degree of separability of the classes known by the classifier.

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

$$F_{measure} = 2 * \frac{Specificity * Sensitivity}{Specificity + Sensitivity} \quad (4)$$

3. Classifier-based Pooling

3.1. The intuition behind our approach

Consider a population of N individuals sending materials to be tested for SARS-CoV-2. Let this population have a known probability of testing positive (prevalence), represented as $p \in [0, 1]$. We propose using a binary classifier to partition this population into two groups: potentially-positive and potentially-negative individuals. Individuals in the first group are tested in isolation, while individuals in the second group are tested in pools according to the tournament strategy.

Dividing the population into two groups does not change the expected number of infected individuals. Let N_1 and N_2 be the number of potentially-negative and potentially-positive individuals, respectively. As illustrated in Equation 5, there will always be prevalence values p_1 and p_2 for the groups so that the expected number of people with the condition remains the same when dividing the population. Moreover, except for the trivial cases, several combinations of values for p_1 and p_2 comply with the equality.

$$(N_1 + N_2) \times p = N_1 \times p_1 + N_2 \times p_2 \quad (5)$$

Dividing the population randomly leads to a result in which $p_1 \approx p_2 \approx p$. However, a binary classifier can divide the population in a way that $p_1 \ll p \ll p_2$. Considering that N_1 and N_2 will be tested in pools of different sizes, say s_1 and s_2 , the total number of tests is $\lceil \frac{N_1}{s_1} \rceil + \lceil \frac{N_2}{s_2} \rceil$. If $s_1 \gg s_2$, the number of test kits required after dividing the population will be lower than N , and there will be an economy of lab test kits in comparison to testing individuals in isolation.

The novelty of our approach is that pool size and composition consider the prevalence and the individuals' probabilities to test positive under SARS-CoV-2 as determined by a classifier, instead of randomly assigning individuals to pools. We name this technique as *adaptive tournament pooling strategy*. As illustrated in Figure 2, the classifier uses additional information about the individuals to update their probability of being infected and scatters the population into a bi-modal distribution. Thus, instead of a uniform distribution stating the same probability of being infected for all individuals (the prevalence), one or more attributes are identified that may posit distinct probabilities for different individuals. We use the new distribution to better allocate individuals to pools. The expected number of test kits drops according to the classifier's ability to suggest an individual as free of SARS-CoV-2. Specificity is more important than accuracy and precision because the goal is to form as many SARS-CoV-2-free pools as possible.

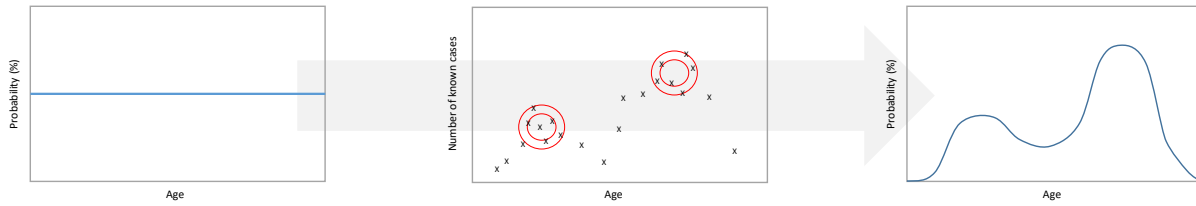


Figure 2. An example of using a SARS-CoV-2 classifier to reorganize the sample population leading to a bi-modal distribution.

A dataset is required to train the binary classifier. We do not postulate the structure or the contents of such a dataset, but the more it can separate individuals into potentially-positive or negative, the better. In general, clinical laboratory and hospitals record information from their patients, including demographic information and a list of reported symptoms, comorbidity, medications taken, medical treatments, and recent blood test results. The patient medical records also register information concerning the lab tests performed, including information such as the date and the result of a PCR lab test. It will be also part of their medical record the condition (SARS-CoV-2 positive or negative) lab test result. This dataset forms a rich base of information that can be used to build a classifier that, in its turn, can be used to distinguish new patients that arrive to be tested.

3.2. A reference strategy for the approach

Any pooling strategy discussed in Section 2.1 could be the reference pooling strategy for our approach. To determine the best strategy to be used along with the binary classifier for SARS-CoV-2, we performed a comparative simulation analysis. As such comparisons depend on a chosen pool size and a prevalence value for testing positive in a population, we selected a discrete number of each of these parameters. We tested pools of size 2, 4, 6, 8, and 10. Information from Brazilian laboratories suggests that PCR test kits might fail if materials from more than ten people were mixed in a pool. This provided an upper-bound for our analysis.

Regarding prevalence, we relied on data from Jun 26, 2020 to determine the prevalence of SARS-CoV-2 for the world population as approximately 0.1% (9,873,860 cases in 7.8 billion individuals) and Brazil as approximately 0.6% (1,274,974 cases in 212 million individuals). We also consider larger values to account for specific environments that might show a far higher prevalence. For instance, we have interacted with industrial partners whose prevalence values were close to 10%. Thus, we evaluated our proposal under the following prevalence values: 0.1%, 1%, 5%, and 10%.

We also set the false-positive probability as 0% because there is virtually no likelihood that a probe designed for a SARS-CoV-2 sequence might produce results for material other than SARS-CoV-2 material. On the other hand, the false-negative probability is set as 5% to account for human error in collecting or manipulating the materials as well as laboratory errors.

We also assumed that individual samples can be used at least thrice to allow testing under the STD pooling strategy. Therefore, we have selected $q = 2$, $k = q + 1$ for complete redundancy analysis, and $n = poolsize \times 2$ for STD. Finally, for each simulation we performed a single test under each strategy, retesting the undetermined individuals in isolation afterward. It is important to notice that the single test consumes the materials of each individual thrice for STD, twice for the full grid, and twice for four individuals (randomly selected) in a border grid. The tournament strategy only uses the materials of any given individual once.

Table 1 shows the results for all pooling strategies using the selected pool sizes and values for prevalence. Each cell in the columns under “Tests saved” denotes the average (μ) and standard deviation (σ) percentile of the number of tests that were saved by using the pooling strategy (using a $\mu \pm \sigma$ notation). Boldface values represent the higher economy observed for a given prevalence. Each cell in the columns under “Errors” denotes the average and standard deviation of the number of errors observed after running the tests under the pooling strategy. These errors are due to the false-negative rate.

We run 100 scenarios and each executes 100 independent rounds of simulations. Each scenario samples a population of 1,000 individuals using a binomial distribution based on the selected prevalence to determine the number of cases of SARS-CoV-2. The simulation rounds under a scenario share the same population, but testing results under the same pooling strategy might differ due to distinct pool compositions and the false-negative rate. Then, undetermined individual results are retested in isolation. For instance, a pool of eight individuals testing positive requires eight new isolated tests to find which of the individuals are real positive cases. All tests performed in a round, either for a

pool or in isolation, are accounted against a higher economy. Economy values and errors are averaged over the simulation rounds and then over the scenarios for a given tuple of prevalence, group size, and pooling strategy.

The values in Table 1 show that the tournament strategy yields the largest economy in terms of the percentile of tests that were saved. On the other hand, it has a higher error rate if compared to the other strategies that involve redundancy. As expected, STD and full grid are not as efficient as a tournament but produce fewer errors, both due to redundancy. Distinct prevalence values lead to distinct pool sizes as the best under the simulation, but the tournament strategy is consistent in leading to the largest economy in every case.

3.3. On the quality of the classifier

The quality of the classifier, in terms of its specificity and sensitivity, can affect the performance of the pooling strategy. We have used simulation to compare a classic tournament strategy to the adaptive tournament pooling strategy driven by classifiers of different qualities. To allow a fair comparison, the classic tournament strategy also uses the results of the classifier as discussed below.

Given a classifier with $\%SN$ sensitivity and $\%SP$ specificity, we first build the population of N individuals according to the prevalence p . We randomly sample a binomial probability distribution to produce a number n of infected individuals under the condition. Then, we sample the number of false-negative individuals according to a binomial distribution of size n and probability $1 - \%SN$. Next, we sample the number of false-positive individuals according to a binomial distribution of size $N - n$ and probability $1 - \%SP$.

Then, we test the population using a given pooling strategy (adaptive or classic tournament) and count the number of test kits consumed. Next, we retest undetermined individuals in isolation, summing up the number of test kits required. We discount the number of tests performed from N and subsequently divide the result by N to calculate the economy in the number of test kits as a percentile of what would be required to test each individual in isolation. The higher the economy, the better is the pooling strategy.

The classic tournament divides the whole population in pools of a given size, building equal-sized pools for potentially-negative and potentially-positive individuals. The adaptive tournament builds pools of a given size for potentially-negative individuals and tests potentially positive individuals in isolation. Pool size is defined according to the prevalence and the results of the former simulation: ten individuals for 0.1% and 1%

prevalence, six individuals for 5% prevalence, and four individuals for 10% prevalence.

Table 2 shows the results of the simulation as the average economy over 100 scenarios, each running 100 independent simulation rounds. The first result of notice is the gain in economy obtained by using a classifier if compared to a naïve pooling strategy (Table 1), except for 0.1% prevalence. For the latter, a pooling strategy without a classifier presents better results because the frequency of positive cases in the population is smaller than the sensitivity and specificity proposed for the theoretical classifiers under investigation.

A perfect, hypothetical classifier (having 100% specificity and 100% sensitivity) would lead to an economy of 89.4%, 88.6%, 78.4%, and 65.3% for the classic tournament pooling strategy (respectively, under 0.1%, 1%, 5%, and 10% prevalence), showing considerable improvement over a tournament strategy that randomly distributes the population into pools. The economies produced for the adaptive tournament pooling strategy under the respective prevalence values are 89.9%, 89.0%, 79.2%, and 67.4%, respectively.

An average difference of 1.3% is observed between the pooling strategies under comparison for any given specificity and sensitivity value under 0.1% and 1% prevalence. Such average difference grows to 2.7% for 5% and 4.9% for 10% prevalence. The smaller variance for smaller prevalence values is explained by the already large economy achieved when grouping a population with a very small number of infected individuals. In such a case, only a few pools will test positive and their individuals will be retested. Furthermore, it is expected that the classifier will suggest few individuals as potentially positive and the largest number of pools will include ten individuals. Differences in economy grow in favor of the adaptive tournament for larger prevalence, that is, for groups in which the incidence of the virus is larger than what has been observed in the world as a whole or any country.

The larger difference in the economy of test kits is observed for 10% prevalence and a bad classifier, having 70% sensibility and 70% sensitivity. This value does not vary significantly if the sensitivity of the classifier is increased, but is reduced to as few as 2.4% if specificity is increased. Increasing specificity implies a smaller number of false-positive individuals, reducing the number of individuals tested in isolation under the adaptive tournament pooling strategy and therefore bringing the number of tests closer to the classic tournament strategy.

Therefore, the adaptive tournament pooling strategy yields the best results in a simulated environment and for a theoretical classifier. We observe that the average economy increases along with the specificity of the clas-

Strategy	Pool Size	Tests saved				Errors			
		0.1%	1%	5%	10%	0.1%	1%	5%	10%
Border grid	2	-0.3 ± 0.3	-2.9 ± 0.9	-13.8 ± 1.9	-26.3 ± 2.4	0.1 ± 0.2	0.5 ± 0.7	2.6 ± 1.6	5.2 ± 2.3
	4	66.2 ± 0.5	62.0 ± 1.5	45.3 ± 2.9	28.2 ± 3.3	0.1 ± 0.3	0.8 ± 0.9	4.2 ± 2.0	8.2 ± 2.9
	6	79.4 ± 0.7	73.6 ± 2.1	51.6 ± 3.7	31.1 ± 4.0	0.1 ± 0.3	0.9 ± 0.9	4.4 ± 2.2	8.9 ± 3.2
	8	85.0 ± 0.9	77.5 ± 2.7	50.6 ± 4.6	28.0 ± 4.4	0.1 ± 0.3	0.9 ± 1.0	4.6 ± 2.3	9.1 ± 3.3
	10	88.0 ± 1.2	78.9 ± 3.3	47.6 ± 5.0	24.1 ± 4.7	0.1 ± 0.3	0.9 ± 1.0	4.6 ± 2.3	9.3 ± 3.5
Full grid	2	-0.3 ± 0.3	-2.9 ± 0.9	-13.8 ± 1.9	-26.2 ± 2.3	0.1 ± 0.2	0.5 ± 0.7	2.6 ± 1.6	5.2 ± 2.3
	4	49.6 ± 0.7	43.8 ± 2.0	21.3 ± 3.6	-0.1 ± 3.7	0.1 ± 0.2	0.5 ± 0.7	2.6 ± 1.6	5.2 ± 2.3
	6	66.4 ± 1.1	57.1 ± 3.3	25.1 ± 4.8	0.4 ± 4.1	0.1 ± 0.2	0.5 ± 0.7	2.7 ± 1.7	5.4 ± 2.4
	8	74.4 ± 1.6	61.8 ± 4.4	23.1 ± 5.4	-1.9 ± 3.9	0.1 ± 0.2	0.5 ± 0.7	2.7 ± 1.7	5.4 ± 2.5
	10	78.2 ± 1.8	63.4 ± 4.9	20.0 ± 5.6	-4.3 ± 3.5	0.1 ± 0.2	0.5 ± 0.7	2.6 ± 1.6	5.2 ± 2.3
STD	4, 2, 3	-50.4 ± 0.4	-53.8 ± 1.2	-67.9 ± 2.4	-83.3 ± 2.9	0.0 ± 0.2	0.5 ± 0.7	2.5 ± 1.6	5.0 ± 2.2
	8, 2, 3	24.3 ± 0.7	18.4 ± 2.0	-4.7 ± 3.8	-26.3 ± 4.0	0.0 ± 0.2	0.5 ± 0.7	2.5 ± 1.6	5.0 ± 2.2
	12, 2, 3	48.6 ± 1.0	39.7 ± 3.0	8.3 ± 4.9	-16.8 ± 4.6	0.1 ± 0.2	0.5 ± 0.7	2.5 ± 1.6	5.0 ± 2.2
	16, 2, 3	60.8 ± 1.4	49.4 ± 3.9	12.0 ± 5.7	-13.8 ± 4.7	0.0 ± 0.2	0.5 ± 0.7	2.5 ± 1.6	5.0 ± 2.2
	20, 2, 3	68.3 ± 1.8	53.3 ± 4.9	9.8 ± 6.4	-14.8 ± 4.6	0.1 ± 0.2	0.5 ± 0.7	2.5 ± 1.6	5.0 ± 2.2
Tournament	2	49.8 ± 0.2	48.1 ± 0.6	40.7 ± 1.3	32.0 ± 1.7	0.1 ± 0.3	1.0 ± 1.0	4.9 ± 2.2	9.7 ± 3.2
	4	74.6 ± 0.4	71.3 ± 1.2	57.4 ± 2.4	42.3 ± 3.0	0.1 ± 0.3	1.0 ± 1.0	4.9 ± 2.3	9.7 ± 3.3
	6	82.7 ± 0.6	77.7 ± 1.8	58.2 ± 3.4	38.8 ± 3.8	0.1 ± 0.3	1.0 ± 1.0	4.8 ± 2.3	9.7 ± 3.5
	8	86.7 ± 0.8	80.2 ± 2.3	55.5 ± 4.2	33.4 ± 4.5	0.1 ± 0.3	1.0 ± 1.0	4.9 ± 2.4	9.8 ± 3.6
	10	89.1 ± 1.0	80.9 ± 2.9	51.9 ± 4.9	28.2 ± 4.9	0.1 ± 0.3	1.0 ± 1.0	4.9 ± 2.4	9.7 ± 3.6

Table 1. Percentile of test kits saved and errors observed by using different pooling strategies with distinct pool sizes under four prevalence scenarios, as mean ± standard deviation.

sifier and slightly decreases with its sensitivity. So, the adaptive tournament pooling strategy makes better use of classifiers that generate few false-positives while being resilient to classifiers that generate false-negatives.

4. Evaluation on a public labeled dataset

4.1. Dataset description

To test the proposed approach, we used a public dataset available at the Kaggle platform. The dataset contains data from patients checked in at the Hospital Israelita Albert Einstein, in São Paulo, Brazil, in March 2020. The hospital had registered, by March 2020, 1223 confirmed cases of SARS-CoV-2. The dataset is a sample from this population and contains data from 5644 individuals with 111 attributes, including descriptive information about the patient, such as an identifier and admission date, but also attributes related to blood lab test results. The data were anonymized and the clinical data were standardized to have zero mean and unitary standard deviation. Fourteen percent of the patients recorded in the spreadsheet tested positive for SARS-CoV-2.

For our analysis, we took a sample of this dataset. We eliminated instances for which there were no PCR lab test results that represent the classifier’s target variable. We discarded attributes with more than 70% of empty values. We also eliminated instances with more than 30% of missing values. The remaining empty data points were substituted by the mean value for the attribute to build a completely filled dataset.

After the cleaning process, the resulting dataset contained 603 individuals with 38 attributes, including the test result for SARS-CoV-2, which is either positive or negative. This sample contained 83 positive individuals

(prevalence = 13.7%). Test results were considered correct for our purpose. We did a correlation study and sensibility test on the attributes and reached to a final set of eight attributes to be considered: *Patient age quantile*, *Platelets*, *Red blood Cells*, *Lymphocytes*, *Leukocytes*, *Eosinophils*, *Neutrophils*, and *Proteina C reativa mg/dL*.

4.2. Classifier configuration

After preparing the data we analyzed three different binary classifiers: Logistic Regression, Random Forest, and Multi-layered Perceptron Neural Networks (MLP NN). As discussed in Section 2, these are three classic supervised machine learning techniques adequate for building diagnostic systems from labeled datasets, as the one made available by the hospital. We implemented the classifiers in Python, using the `sklearn`¹ library.

We configured the neural network with a hidden layer having 100 nodes, a *tanh* activation function, a constant learning rate of 0.0005 (smaller than the default for better tuning), shuffling samples in each iteration, and a stochastic gradient for the solver. We mostly used the default values for the random forest model. We only altered the criterion to split classes to *entropy*, decreased the number of trees in the forest from 100 (default) to 50 to avoid over-fitting due to the small dataset, and set the number of attributes for triggering a split as a function of the square root of the number of attributes. For the logistic regression classifier, we changed the default solver to `liblinear` because the dataset used for training was small and balanced the class weight according to the inverse frequency of the class in the set.

¹<https://scikit-learn.org/stable/modules/classes.html>

Prev	Sens → Spec ↓	Adaptive Tournament						Classic Tournament					
		70%	75%	80%	85%	90%	95%	70%	75%	80%	85%	90%	95%
0.1%	70%	63.0	63.0	63.1	62.9	63.1	63.0	61.2	61.1	61.2	61.0	61.2	61.2
	75%	67.4	67.3	67.6	67.4	67.1	67.4	65.8	65.7	66.0	65.8	65.5	65.8
	80%	71.9	71.8	72.0	72.0	72.0	72.0	70.5	70.5	70.6	70.6	70.6	70.6
	85%	76.4	76.2	76.3	76.4	76.4	76.4	75.3	75.1	75.2	75.3	75.2	75.3
	90%	81.0	81.0	80.8	81.1	80.8	80.9	80.0	80.0	79.9	80.2	79.9	80.0
1%	70%	62.5	62.6	62.5	62.4	62.2	62.3	60.6	60.7	60.6	60.5	60.3	60.4
	75%	67.1	66.9	67.0	66.8	66.9	66.8	65.4	65.2	65.4	65.1	65.3	65.2
	80%	71.7	71.7	71.4	71.5	71.2	71.5	70.2	70.3	70.0	70.1	69.8	70.1
	85%	75.9	75.9	75.8	75.7	75.9	75.5	74.8	74.7	74.6	74.6	74.8	74.3
	90%	80.5	80.4	80.4	80.3	80.2	80.2	79.5	79.5	79.5	79.4	79.2	79.3
5%	70%	56.5	56.4	56.2	55.8	55.7	55.7	52.5	52.4	52.2	51.8	51.6	51.7
	75%	60.5	60.4	60.3	59.8	59.7	59.6	57.0	57.0	56.8	56.3	56.2	56.0
	80%	64.4	64.3	64.2	63.9	64.0	63.5	61.5	61.4	61.3	61.0	61.0	60.5
	85%	68.6	68.3	68.2	67.9	67.6	67.4	66.2	66.0	65.9	65.5	65.2	65.0
	90%	72.5	72.2	72.0	72.0	71.7	71.3	70.8	70.4	70.3	70.2	69.9	69.4
10%	70%	49.4	49.0	48.8	48.3	48.1	47.5	42.4	42.0	41.7	41.1	40.8	40.1
	75%	52.8	52.6	52.1	51.5	51.3	50.9	46.8	46.5	45.9	45.1	44.9	44.3
	80%	56.0	55.8	55.5	55.1	54.8	54.5	50.8	50.6	50.2	49.7	49.3	48.9
	85%	59.6	59.3	58.8	58.5	58.0	57.6	55.4	55.0	54.3	54.0	53.4	52.9
	90%	63.0	62.7	62.3	61.8	61.6	61.0	59.8	59.3	58.8	58.2	57.9	57.2
95%	70%	66.4	66.0	65.7	65.3	64.8	64.5	64.0	63.4	63.2	62.6	62.0	61.6

Table 2. Mean percentile of lab kits saved by using two pooling strategies driven by a classifier under four prevalence values and according to different classifier qualities.

4.3. Results and discussion

Table 3 presents the quality of the results produced by each classifier using the measures shown in Section 2. We used a 3-KFold method with 70% of the available data for training (422 individuals) and 30% of the data for testing (181 individuals).

Metric	Logistic	Random Forest	MLP NN
Sensitivity	0.44 ± 0.0	0.73 ± 0.05	0.62 ± 0.0
Specificity	0.72 ± 0.0	0.22 ± 0.0	0.64 ± 0.0
$F_{measure}$	0.55 ± 0.0	0.34 ± 0.01	0.63 ± 0.0
AUC	0.79 ± 0.1	0.61 ± 0.1	0.79 ± 0.1

Table 3. The performance of the classifiers in the dataset. Numbers in red represent low performance, while numbers in bold represent better performance. The MLP NN model presents the best overall results.

The three classifiers presented limited performance, running from an AUC of 0.61 for the random forest classifier to 0.79 for the Logistic Regression and the Multi-Layered Perceptron. Besides that, most sensitivity and specificity values are below the ranges we have selected for the simulations in the previous section. This limited performance is a consequence of using a small dataset comprising only 603 individuals. Consequently, a diagnosis coming from these classifiers should be taken with care. Nevertheless, the ordering effect is sensible when looking at the guidance they provide to the pooling process.

Table 4 supports our conclusion of the benefit a classifier can provide compared with the best results of the classic pooling strategy. Savings were calculated based

on the expected number of lab test kits that were saved according to the optimum group size shown in the third row. The Logistic Regression and the MLP NN classifiers provide significant savings in terms of needed lab test kits, going from 32.5% to a saving of 46.9%. These results confirm the insight gained through simulation, as the best classifier (MLP NN) is closely followed in performance by the classifier showing the largest specificity. Even the random forest classifier that did not perform well presented an increase in savings of 5.1%.

These findings support our claim that using a binary classifier to pre-order the individuals according to their probability to have the condition provides a positive effect on the number of required lab test kits. Even having poor ability to predict the incidence of the condition – presenting a large number of errors in the frontier between its classes – correctly predicting the “easy cases” makes the binary classifier a valuable pre-processor to optimize resource usage in testing. The code used in these experiments can be found in a GitHub repository².

We also performed a sensitivity test to verify the limitations of our approach. We varied the sensitivity and specificity of a theoretical classifier and calculated the expected number of lab test kits when performing the pre-test for allocating individuals into pools. As illustrated in Table 5, for prevalence values lower than 0.05, our approach is not worthwhile since it would require an almost perfect classifier. On the other hand, for prevalence values greater than 0.05, the pre-classification effectively saves PCR test kits. In the example in Table

²https://github.com/bicharra/Pooling_Methods

	Tournament	Adaptive Tournament Pooling Strategy					
	Group	Logistic Regression		Random Forest		MLP NN	
		Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
Group size	181	140	41	173	8	155	26
Prevalence	13.7%	5%	44%	11%	73%	6%	62%
Pool size	4	6	1	4	1	5	1
Pool economy	32.5%	58%	0%	39.4%	0%	54.8%	0%
Number of tests	$181 \times 67.5\% = 123$	$140 \times 42\% + 41 = 100$		$173 \times 60.6\% + 8 = 113$		$155 \times 45.2\% + 26 = 96$	
Lab test savings	32.5%		44.7%		37.6%		46.9%

Table 4. Results of the comparison between a classic tournament strategy and the adaptive tournament pooling strategy in a real database with three binary classifiers.

5, with a population of 100 individuals and a pool size of 10 people, only a very potent classifier (sensitivity = 95% and specificity = 95%) reaches an expected number of lab test kits smaller than those saved by using the classic pooling strategy (16.3 *versus* 19.3). On the other hand, for prevalence values greater or equal to 0.05 the savings using our approach are significant even limited classifiers. The calculation process used to build Table is explained in Appendix A.

5. Conclusion

One of the main strategies for fighting the spread of the SARS-CoV-2 pandemics is continuous, massive testing. PCR tests offer a reliable method for identifying infected individuals. However, resources for applying these tests are scarce and expensive. This paper presented a novel approach to optimize testing for SARS-CoV-2 using pools. We studied different pooling strategies using simulation, and tournament pooling was selected as the most effective strategy.

Tournament pooling was combined with a binary classifier to build pools of potentially-negative individuals while testing potentially-positive ones in isolation. The role of the classifier is to distinguish individuals with a high probability of being infected from those with low probability (not to provide a diagnostic). We were interested in clustering individuals with the lowest probability of being infected, allowing larger pools of healthy individuals. This classifier-backed pooling strategy was named *adaptive tournament pooling strategy*.

We have implemented the proposed approach and evaluated its performance by using an actual dataset of individuals that checked in a Brazilian hospital and have been subjected both to blood and SARS-CoV-2 tests. Results showed that using a binary classifier decreases the number of required PCR kits to test a population, even if the classifier has limited prediction capability.

We believe that our approach is useful for any country with limited PCR test kit resources. It only needs data on the profile of individuals, as well as the results of their PCR tests, to train the classifier. We have shown

that our approach is worthwhile whenever the disease prevalence in the population that sends material to be tested is greater than 5%.

References

- [1] A. Kronbichler, D. Kresse, S. Yoon, K. H. Lee, M. Effenberger, and J. I. Shin, "Asymptomatic patients as a source of covid-19 infections: A systematic review and meta-analysis," *International journal of infectious diseases*, vol. 98, pp. 180–186, 2020.
- [2] E. Surkova, V. Nikolayevskyy, and F. Drobniowski, "False-positive covid-19 results: hidden problems and costs," *The Lancet Respiratory Medicine*, 2020.
- [3] M. N. Esbin, O. N. Whitney, S. Chong, A. Maurer, X. Darzacq, and R. Tjian, "Overcoming the bottleneck to widespread testing: A rapid review of nucleic acid testing approaches for covid-19 detection," *RNA*, pp. rna-076232, 2020.
- [4] K. van Oorschot, Y. Wang, and H. Akkermans, "Throbbing between two lives: Resource pooling in service supply chains," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
- [5] H. J. Scholl and S. AlAwadhi, "Pooling and leveraging scarce resources: The smart ecity gov alliance," in *2015 48th Hawaii International Conference on System Sciences*, pp. 2355–2365, IEEE, 2015.
- [6] D. J. Westreich, M. G. Hudgens, S. A. Fiscus, and C. D. Pilcher, "Optimizing screening for acute human immunodeficiency virus infection with pooled nucleic acid amplification tests," *Journal of Clinical Microbiology*, vol. 46, no. 5, pp. 1785–1792, 2008.
- [7] C. Gollier and O. Gossner, "Group testing against covid-19," *Covid Economics*, vol. 2, 2020.
- [8] N. Thierry-Mieg, "A new pooling strategy for high-throughput screening: the shifted transversal design," *BMC Bioinformatics*, 2006.
- [9] R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 14, pp. 436—440, December 1943.
- [10] S. Mallapaty, "The mathematical strategy that could transform coronavirus testing," *Nature*, vol. 583, pp. 504–505, July 2020.
- [11] C. A. Hogan, M. K. Sahoo, and B. A. Pinsky, "Sample pooling as a strategy to detect community transmission of sars-cov-2," *Journal of the American Medical Association*, vol. 323, no. 19, pp. 1967–1969, 2020.
- [12] B. Abdalhamid, C. Bilder, E. McCutchen, S. Hinrichs, S. Koepsell, and P. Iwen, "Assessment of specimen pooling to conserve sars cov-2 testing resources," *American*

		Classifier configuration														
		Sensitivity	65%	65%	70%	70%	70%	70%	80%	80%	80%	80%	90%	90%	90%	90%
		Specificity	65%	70%	65%	70%	80%	90%	65%	70%	80%	90%	65%	70%	80%	90%
Prevalence	Classic	Adaptive tournament pooling strategy														
1%	19.6	45.2	40.7	44.8	40.3	31.4	22.5	43.9	39.4	30.5	21.6	43.0	38.5	29.6	20.7	16.3
5%	50.1	58.3	54.2	56.6	52.4	44.0	35.6	52.8	48.6	40.1	31.6	48.8	44.5	36.0	27.5	23.2
10%	75.1	71.5	67.9	68.9	65.2	57.6	50.0	62.8	58.9	51.1	43.3	55.7	51.7	43.7	35.7	31.6
15%	90.3	81.6	78.7	78.7	75.5	69.1	62.4	71.5	68.0	61.0	53.8	62.3	58.5	51.1	43.6	39.8
20%	99.3	89.2	86.9	86.3	83.8	78.4	72.8	78.9	75.9	69.7	63.3	68.4	65.0	58.1	51.1	47.6

Table 5. A comparison showing the expected number of required lab test kits calculated according to the classic pooling method and our approach varying sensitivity and specificity of the classifier. Numbers in red emphasize results calculated under specific classifier’s configuration for which the classic pooling outperforms our approach.

Journal of Clinical Pathology, vol. 153, pp. 715—718, April 2020.

- [13] F. Regen, N. Eren, I. Heuser, and J. Hellmann-Regen, “A simple approach to optimum pool size for pooled sars-cov-2 testing,” *International Journal of Infectious Diseases*, 2020.
- [14] C. M. Verdun, T. Fuchs, P. Harar, D. Elbrächter, D. S. Fischer, J. Berner, P. Grohs, F. J. Theis, and F. Kraemer, “Group testing for sars-cov-2 allows for up to 10-fold efficiency increase across realistic scenarios and testing strategies,” *medRxiv*, 2020.
- [15] L. Mutesa, P. Ndishimye, Y. Butera, J. Souopgui, A. Uwineza, R. Rutayisire, E. Musoni, N. Rujeni, T. Nyatanyi, E. Ntagwabira, M. Semakula, C. Musanabaganwa, D. Nyamwasa, M. Ndashimye, E. Ujeneza, I. E. Mwikarago, C. M. Muvunyi, J. B. Mazarati, S. Nsanzi- mana, N. Turok, and W. Ndifon, “A strategy for finding people infected with sars-cov-2: optimizing pooled testing at low prevalence,” 2020.
- [16] R. Kainkaryam and P. Woolf, “poolhits: A shifted transversal design based pooling strategy for high-throughput drug screening,” *BMC Bioinformatics*, 2008.
- [17] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, vol. 398. John Wiley & Sons, 2013.
- [18] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [19] V. Krková, “Kolmogorov’s theorem and multilayer neural networks,” *Neural networks*, vol. 5, no. 3, pp. 501–506, 1992.

Appendix

The estimation of the expected number of required SARS-CoV-2 test kits involves calculating the conditional probability of each of the four scenarios produced by the classifier: true positive, true negative, false positive, and false negative. We consider that if our classifier suggests that the individual has SARS-CoV-2, the individual is tested in isolation and consumes a test kit. Otherwise, the individual is submitted to the pooling process. We want to calculate the expected number of test kits required to test a population of N individuals, with prevalence p , and using a classifier whose performance is defined by its sensitivity (SN) and specificity (SP).

Consider a population of $N = 100$ individual swab materials sent to a lab for testing. Let the disease prevalence be $p = 10\%$ and a pool size of $PS = 10$. Assume a classifier with $SN = 80\%$ and $SP = 70\%$. The probability of the classifier returning a positive result (p^+) and a negative result (p^-) are given by:

$$p^+ = p \times SN + (1 - p) \times (1 - SP) = 0.1 \times 0.8 + 0.9 \times 0.3 = 0.35$$

$$p^- = 1 - p^+ = 0.65$$

Therefore, on average 35% of the individuals will be retested in isolation. The conditional probability of an individual having the condition given that the classifier’s result is negative (false negative) is given by p' calculated as:

$$p' = \frac{p \times (1 - SN)}{p \times (1 - SN) + (1 - p) \times SP} = \frac{0.1 \times 0.2}{0.1 \times 0.2 + 0.9 \times 0.7} = 0.0308$$

p_{pool}^- and p_{pool}^+ are, respectively, the probability of a pool of 10 people testing negative and positive.

$$p_{pool}^- = (1 - p')^{PS} = (1 - 0.0308)^{10} = 0.7315$$

$$p_{pool}^+ = 1 - p_{pool}^- = 0.2684$$

There will be two groups: G1 in which all individuals require a test kit because the classifier indicated them as infected, and G2 for which individuals will go through a classic pooling process. The expected number of required test kits, $E(total)$, will be the expected number of test kits for group G1, $E(G1)$, plus the expected number of test kits for group G2, $E(G2)$. In our case, 58.9 is smaller than the results of the classic tournament (75.1).

$$E(G1) = p^+ \times N = 0.35 \times 100 = 35$$

$$E(G2) = (p_{pool}^- + (PS + 1) \times p_{pool}^+) \times p^- \times \frac{N}{PS} = 23.9$$

$$E(total) = 35 + 23.9 = 58.9$$