Covid Testing Optimisation Techniques

It's impossible to contain covid-19 without knowing who's infected, until a safe and effective vaccine is widely available, stopping transmission is the name of the game. While testing capacity has increased, it's nowhere near what's needed to screen patients without symptoms, who account for nearly half of the virus's transmission.

Our research points to a compelling opportunity for data science to effectively multiply today's testing capacity. If we combine machine learning with test pooling, large populations can be tested weekly or even daily, for as low as Rs220 to Rs370 per person. In other words, governments can safely reopen the economy and halt ongoing covid-19 transmission, all without building new labs and without new drugs or vaccines. The current alternatives, though, are not appealing. Infrequent testing (monthly seems to be the default in many proposals) or haphazard screening allow active cases to spread the virus for weeks before it's caught. And the price is still high at around Rs1000 per test.

Pooled testing, guided by machine-learning algorithms, can fundamentally change this calculus. In pooled testing, many people's samples are combined into one. If no virus is detected in the combined sample, that means no one in the pool is infected. The entire pool can be cleared with just one test. But there's a catch; if anyone in the pool is infected, the test will be positive and more testing will be required to figure out who has the virus. So, a key part of knowing how to pool is knowing the likelihood that certain people in the group will be positive, and separating them from the rest. How do we know that risk? That's where machine learning comes in. The risk of infection is evolving rapidly in the India, the relative odds in Maharashtra and Kerala have reversed in a matter of weeks.

Another advantage; pooled testing gets more efficient when disease prevalence is lower. If a population, say all students at a university, is tested daily, the risk of infection is dramatically lowered for everyone in the group, simply because testers remove positives from tomorrow's pool when they diagnose them today. That means tomorrow's pool can be even larger, which reduces the number of tests needed and thus the cost of testing the population. And with more frequent testing, people who are infected but don't have symptoms can stay home, further reducing spread and making pooled testing even more efficient. Daily testing can actively suppress the virus, whereas monthly testing really only allows us to see how badly things have gone.

The last pillar of prevention through testing requires accounting for the virus's spread between people and, therefore, for risk that is correlated. Using machine learning to model social networks has been a growing focus for researchers in computer science, economics, and other fields. Such algorithms, combined with data on jobs, classrooms, university dorms, and many other settings, allow machine-learning tools to estimate the potential that different people will interact. Knowing this likelihood can make group testing even more powerful. Machine learning can give us the precise individual-level estimates we need to make pooling work, by identifying those likely to test positive and keeping them out of large pools. Such groups of risky and non-risky participants can be made based on the following factors that can be fed into the machine learning algorithm:

1) Occupation (for example, healthcare workers and policemen are more prone)
2) Residential location
3) Workplace location
4) Patients with obvious symptoms
5) Travel history
6) Whether participant has been in close contact with positively tested person or not.
7) Immunity level from previous hospital records.
8) Blood Type (Once researchers have established evidence that people with type O blood have lower risk)

To tide over the constraints of time and resources, several countries, including US and India, have gone in for pooled testing.

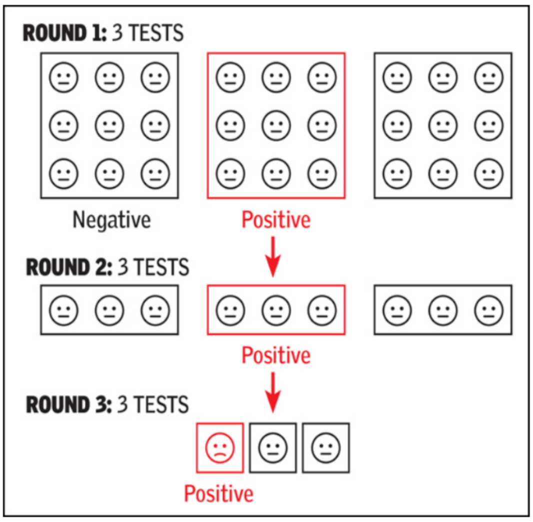
To understand pooled testing strategies (1-3), we'll make some assumptions:

- Total number of people to be tested: 27

- Number of pools: 3
- People in each pool: 9

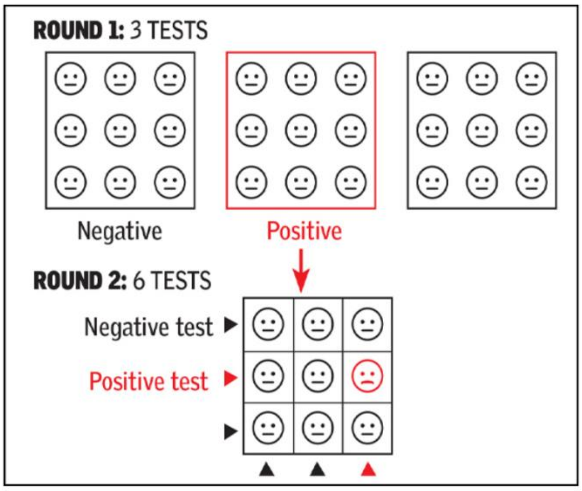
STRATEGY 1:

Now, say, of the 3 pools (let's call them A, B, C), two show no positivity, that is, these groups are clear of any infections and only one group of 9 people (say B) returns a positive result. That would mean that at least one of the 9 people in Pool B is positive. To zero in on which person that is all that the clinic has to do is to test the 9 samples individually. Thus, 27 people were tested using 3 + 9 = 12 kits.



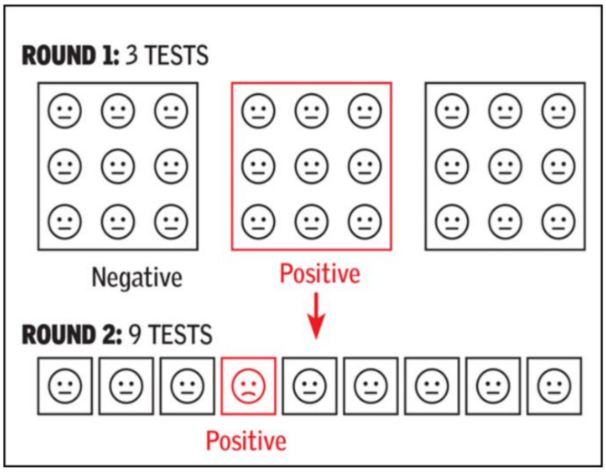
STRATEGY 2:

Again, assuming that only one of the three pools returns a positive test, it means that at least one of the 9 people in that particular group is positive. Now comes a variation on Strategy 1. Instead of testing the 9 samples in the positive pool individually, the clinic breaks them down into 3 smaller pools of 3 samples each and tests them again. When one of these smaller groups returns a positive result, the clinic can individually test just those 3 samples to zero in on the infected sample. In the second method, 27 people were tested using 3 + 3 + 3 = 9 kits



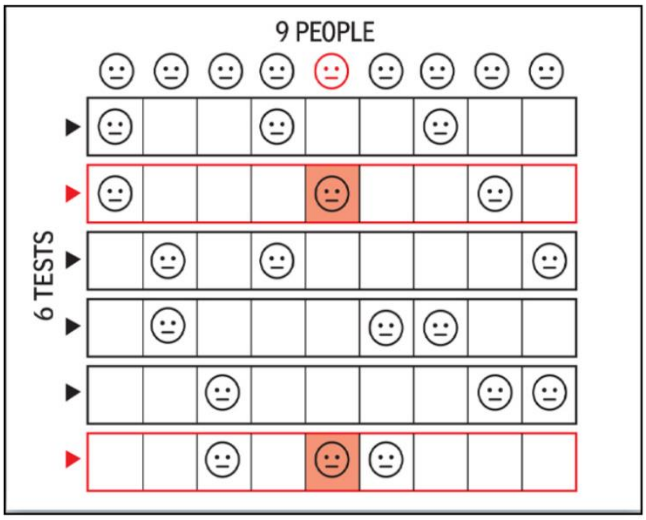
STRATEGY 3:

The above strategies leave room for improvement in terms of time taken to identify positive cases and the number of rounds of testing needed. To solve this, there are two other, slightly complicated, strategies. The first of these is like a grid system. Here's how it works: 27 samples are split into 3 pools of 9 each, just like the previous strategies. Once an infection is found, these 9 samples are split into horizontal (rows) and vertical (columns) grid. Now, the horizontal samples are each tested taking one row as one pool. So, that's three tests. The vertical samples are also tested with one column as one pool, which makes it another three tests. Only one sample can be common in the horizontal and vertical grids, thus leading the clinic to the infected sample. In the third method: 27 people tested using 3 + (3 + 3) = 9 kits.

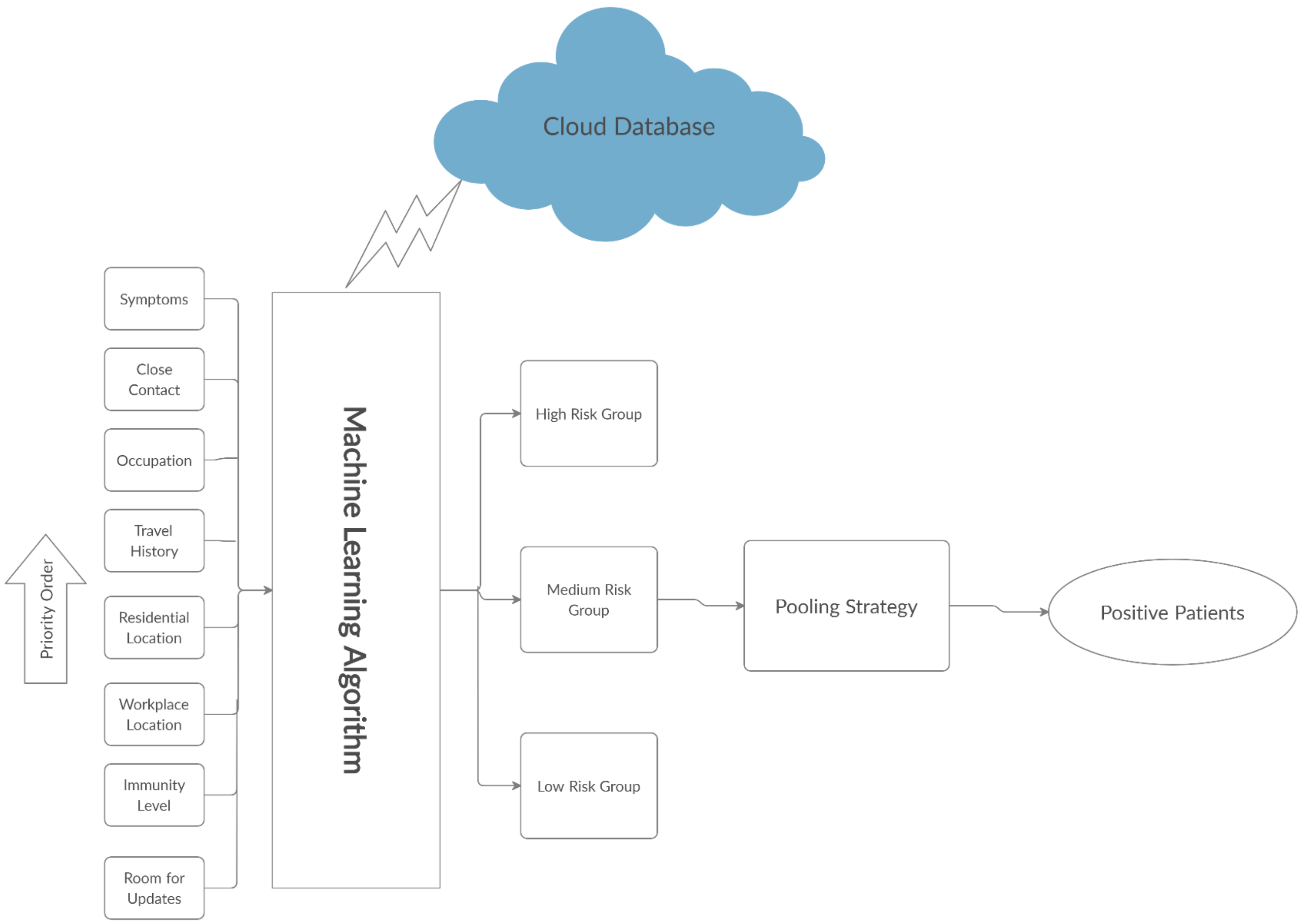


STRATEGY 4:

Time is of the essence and even two rounds of testing could be too much. In this strategy we do pooled testing in one round using "many overlapping groups". While this method increases the number of tests, it saves time. Samples are distributed in different pools (rows) as per mathematical rules. Every sample falls in two pools, and there is only one overlap between any two pools. So when you know which two pools have tested positive, you will also know which is the common sample that is positive. Every test should also include the same number of samples (3 in this case). In unpublished results from clinical trials in India, 5 positive samples were identified out of 320 using only 48 tests.



High Level Architecture Diagram



Our Solution

We use the AI Watson Assistant to ask the participant for the following information and save it on a database on IBM cloud. An additional column calculates the total points gathered by each participant which then decides what risk level group they get admitted to. The points for Travel, Residential and Workplace locations are calculated with the help of machine learning. The AI compares the user’s input with that of up to date covid hotspot locations from the cloud database to decide how risky the location is. The occupation and symptoms are also given points based on risk factor with the help of machine learning. The three risk level groups can be used by hospitals to apply one of the strategies described earlier and can help save time and cost per test kit.

AI Assistant:

Points	
Calculated By AI	List the symptoms you are experiencing: [V] more
	Have you been in close contact with anyone tested positive?
7	☐ Yes
0	☐ No
	What is your occupation?
0	☐ Work/study at home
0	☐ Unemployed
Calculated By AI	
	Travel History:
5	☐ Traveled anywhere internationally in the last 45 days.
Calculated By AI	Places visited in last 28 days: State, District, Locality [V] more

Points	
Calculated By AI	Residential Location (in lowercase): State: ☐ District: ☐ Locality: ☐
Calculated By AI	Study / Workplace Location (in lowercase): State: ☐ District: ☐ Locality: ☐
	Immunity Level: Have you ever had any of the following?
4	☐ Diabetes
3	☐ Hypertension
4	☐ Lung Disease
4	☐ Heart Disease
3	☐ Kidney Disorder
4	☐ Cancer

Database on IBM Cloud

Participant	Symptoms	Close Contact	Occupation	Travel History	Residential Location	Study/Workplace Location	Immunity Level	Total Points
Cyril Shaji	Nil	No	Work/Study at home	Nil	kerala, thrissur, kalathode	kerala, thrissur, kalathode	Nil	0
Kuriachen Sunil	Nil	No	Work/Study at home	Nil	kerala, kottayam, vallichira	kerala, kottayam, vallichira	Nil	0
Chandichen Sunil	Headache, fever, loss of sense of smell and taste	No	Work/Study at home	Nil	kerala, kottayam, vallichira	kerala, kottayam, vallichira	Nil	18
Meenakshi Sajeev	Nil	No	Work/Study at home	Nil	kerala, ernakulam, fort kochi	kerala, ernakulam, fort kochi	Nil	0

Prepared by Team T853

Cyril Shaji

Kuriachen Sunil

Meenakshi Sajeev

Chandichen Sunil