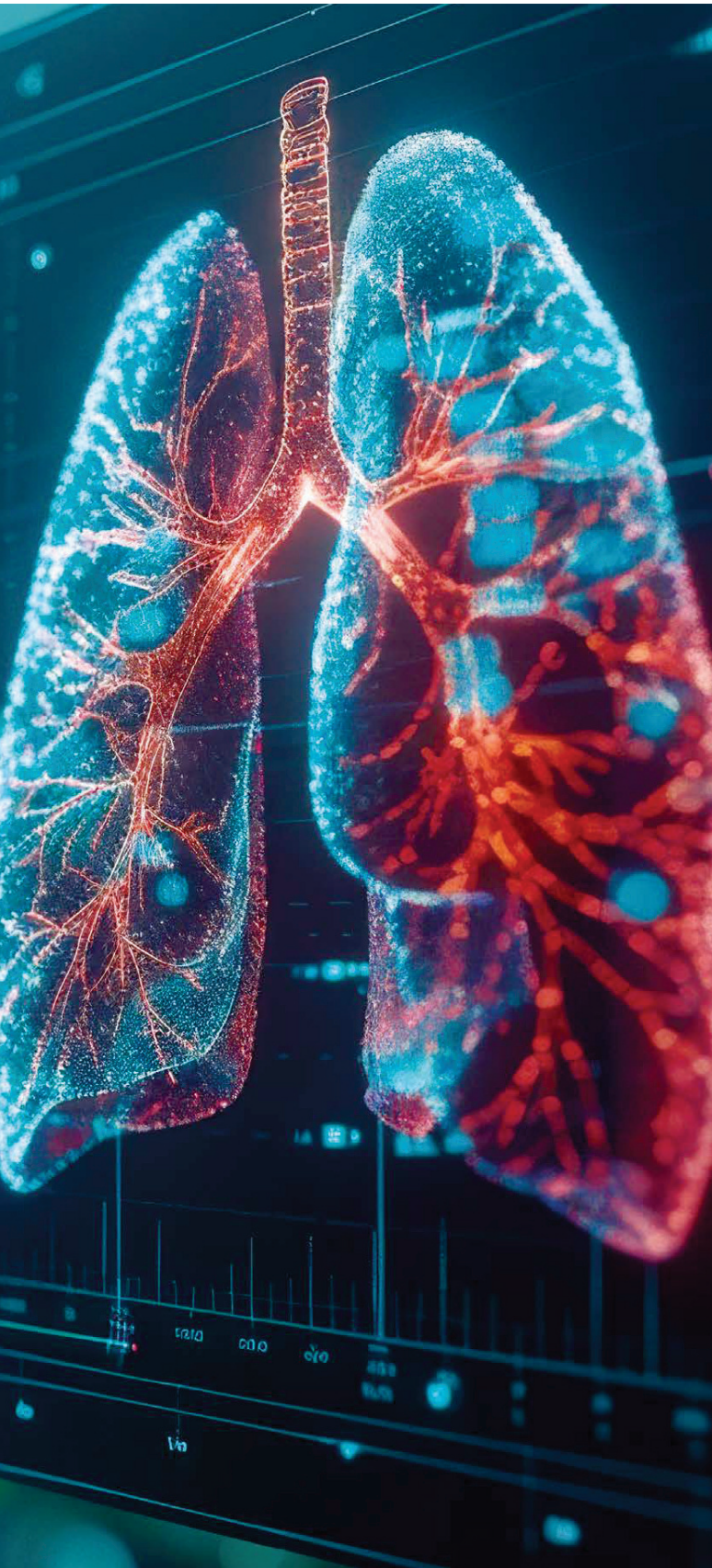


# Machines and Artificial Intelligence: A New Cure for Tuberculosis



A group of doctors and engineers uses data analysis models, machine learning and artificial intelligence to combat pulmonary tuberculosis in the most needy places in Colombia.

By David Mayorga

Photos by 123RF, Milagro Castro

DOI [https://doi.org/10.12804/dvcn\\_10336.425621\\_num7](https://doi.org/10.12804/dvcn_10336.425621_num7)

**T**he discomfort starts in the hypopharynx, the lower part of the throat. It is a tickling that goes up -sometimes causes dryness- and spreads through the walls of the area and becomes a cough: a slight contraction, followed by a guttural sound that only seeks to clear the tract between the lungs and mouth.

As the days go by, that cough becomes persistent. It appears in the early hours of the morning, before entering the shower, and persists throughout the day, when the feeling of thirst increases and chest pains appear. At night, the cough increases in intensity and even withstands the calming effect of hot drinks, herbal infusions and other home remedies with which you try to close the passage. And it is still there, as the person tries to fall asleep... Over the days or weeks, the cough intensifies to become a general discomfort that forces the person to go to the doctor.

Already in the office, sitting in front of the doctor, the questions begin: how old you are, where you live, with whom, for how long you have been feeling the discomfort ...

all are answered between small episodes of coughing. Then, with his stethoscope in hand, the doctor begins the usual check-up and as he listens, his suspicions become more solid. He goes to the closet and returns with a bag and two plastic jars: he orders the patient to cough with all his strength and spit inside one of them; he recommends repeating the action the next day. Both bottles will be sent to the lab for testing.

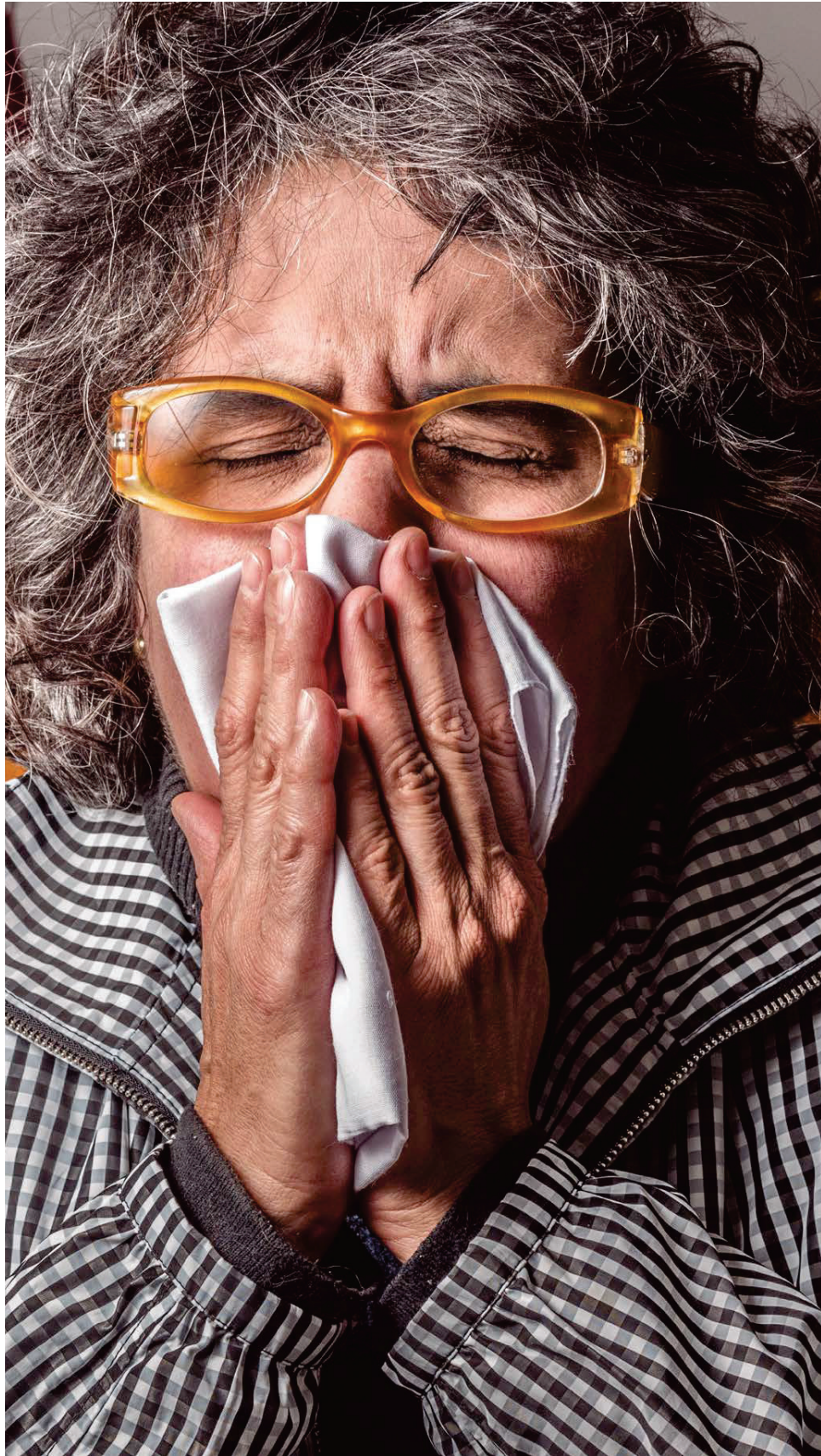
However, time passes and there is no news of the results. The cough becomes much more persistent and shocking. Now, even more, it hurts to cough. And around them new episodes of coughing arise within the family and among the closest friends. The despair and persistence of pain force the patient to return to the office. And then came the unexpected news: there were transport problems, the samples were contaminated and their analysis was not possible.

The evolution of these symptoms confirms the doctor's suspicions, but the high cost of molecular testing does not allow him to corroborate it immediately. Based on your experience, your doctor prescribes antibiotics, takes new samples, and orders a thorough care routine. Months and a half later the results are ready, and with them the doctor reconfirms his suspicions that the patient suffers from tuberculosis. However, there is something much worse: by that time, the close circle of the patient also suffers from uncomfortable and repetitive episodes of coughing.

### **All have been infected with the bacillus *Mycobacterium tuberculosis*.**

It seems surreal, but it is a scene that occurs in different populations of Colombia. According to data compiled by the Instituto Nacional de Salud (National Institute of Health) (INS, by its acronym in Spanish) in the [Sivigila](#) registration system, as of the first week of October 2023, exactly 17 754 cases of tuberculosis had been recorded across the territory. That sounds like a small number, given that Colombia's population exceeds 50 million –indeed, in 2022, the incidence rate of pulmonary tuberculosis, the most common type, was 29.39 per 100 000 population in 2022-. However, these numbers attract the attention of expert doctors such as pulmonologist Carlos Awad, who has been studying the disease for more than 35 years and recording its findings in detail.

“Tuberculosis is a preventable, treatable, curable disease, which is related to some special risk factors, which are likely to be nega-



tive facts. So, getting sick with TB would indicate that there is something wrong with the health system or the socio-economic side. This is why it is an important disease in the field of public health, such as leprosy, tetanus or measles. For the country, TB is a priority disease,” Dr. Awad explains.

Although retired, the specialist remains linked to the Santa Clara Hospital in Bogotá, after a life as coordinator of the Lung Laboratory of the institution, member of its ethics committee and even as deputy director of Research. Throughout his career he has witnessed the evolution of the figures of contagion of this disease, but, above all, he has treated his patients.

Sivigila's figures show that tuberculosis is present throughout the Colombian territory: from 10 cases in the archipelago of San Andrés, Providencia and Santa Catalina, to 3599 in the department of Antioquia. However, analysis of these numbers shows that their background is entirely sociodemographic, with 80.29% of confirmed cases located in strata 1 and 2. Going further allows us to see the face of this disease, because the most affected populations correspond to those living in overcrowded prisons (1120 registries), migrants (876, a figure that includes displaced persons due to the internal conflict) and indigenous people (720).

### The journey of artificial intelligence

A simple Excel sheet. Is the beginning, a new tool to treat tuberculosis in Colombia began to be built. This is a patient registry that pulmonologist Awad had been compiling while performing his duties at the Santa Clara Hospital (according to his consultation and from the data collected during the health brigades that the institution conducts in vulnerable populations). This was a valuable exercise, but it showed that there was a long way to go.

“The differences in data disposition are abysmal. For example, China, with its cities of millions of inhabitants, always studies very easily. Here we have to dig a little deeper into the data,” says Álvaro David Orjuela Cañón, an electronic engineer, PhD. of Electrical Energy (focused on computational intelligence in health) at the Federal University of Rio de Janeiro, and associate professor in the School of Medicine and Health Sciences at Universidad del Rosario.

In 2015, after completing his doctoral research with tuberculosis cases in several hospitals from Brazil ([the World Health Organization](#) –WHO– estimated 104 000 infections in the neighboring country during 2021), Orjuela landed in Colombia with the aim of researching and finding solutions for the treatment of the disease. And to do so, it brought not only the results of its trials in Brazil, but also an unexpected ally: artificial intelligence (AI).

His idea consisted of an instrument that would allow the doctor at the first consultation, and from the sociodemographic information provided by the patient (neighborhood of residence, conditions of his home, number of people with whom he lived and possibilities of access to the health system, among other data), as well as the description of his symptoms, determine with the help of statistical models and data processing the probability of whether or not he suffered from tuberculosis.



↑  
Álvaro David Orjuela Cañón, associate professor in the School of Medicine and Health Sciences at Universidad del Rosario, explains that it would be ideal to include this artificial intelligence-based analysis model in a mobile application that allows doctors to prioritize the limited resources at their disposal when deciding what treatment to apply to a patient with symptoms of tuberculosis.

It would be a preliminary step before the application of bacilloscopy or, in certain cases, other tests such as mycobacterial culture, a molecular test or, for extreme cases, a puncture that determines whether tuberculosis has affected the pleura of the lung. This would allow treatment to be started immediately, reduce the risk of infection, and, in extreme cases such as those in remote regions of Colombia with restricted health coverage, avoid patient cost overruns.

However, to build it first, everything needed data. And although he found a good harvest in the databases of the Sivigila, Orjuela soon ran into a difficulty: “The system compiles the cases already confirmed, but to train the AI models we needed to show them the discarded ones as well.”

When he tried to access the latter, he received many refusals from various health institutions in Bogotá. Until it aroused the interest of the pulmonologist Awad and the Hospital Santa Clara, reference in Colombia in the treatment of tuberculosis if you con-

sider that [It was opened in 1942](#) as a national tuberculosis sanatorium. The innovative AI-based analysis proposal began to be built, then, with the Excel file in which, for several years, Awad had been recording data about the patients he treated, a job in which Dr. Orjuela also collaborated.

The project itself would take a radical turn in 2018. Orjuela, who at the time worked at Universidad Antonio Nariño, received along with his colleague [Andrés Jutinico](#), the pulmonologist Awad and the [Integrated Sub-network of Health Services](#) Centro Oriente E.S.E. -to which the Santa Clara belongs- a grant from the Ministry of Science, Technology and Innovation to implement its idea,

called "[Generation of alternative models based on computational intelligence for screening and diagnosis of pulmonary tuberculosis.](#)" This support provided access to the institutional databases that were crucial for advancing to the next phase: training the AI tool through machine learning.

For this, it was necessary to implement up to five different models of information analysis (see box) necessary to debug the data and process them according to six variables within the context of Santa Clara, all inquiring about the socioeconomic and demographic reality of the patient: biological sex,

### Of Trees, Neurons, and Other Models of Artificial Intelligence (AI)

To construct artificial intelligence models, scientists have relied on various techniques. These include:

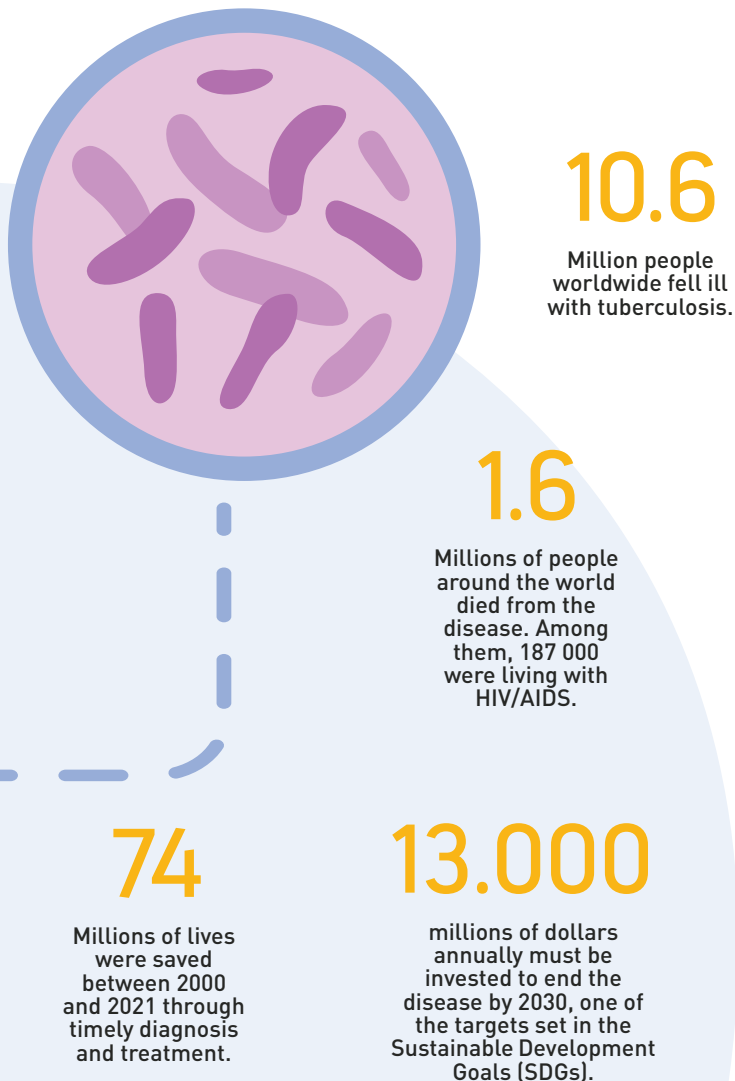
- **Decision Trees:** This method is used to solve problems in data classification. It is typically represented as a flowchart constructed with internal nodes that reflect the characteristics of a dataset; its branches constitute the decision rules, and each leaf (node) represents an outcome.
- **Random Forests:** Algorithms built from decision trees. This technique allows for ensemble learning by combining multiple classifiers to address a complex problem and improve the performance of the employed model.
- **Support Vector Machines:** Also used for classification tasks, these allow for the construction of the optimal line (or boundary) of decision to establish the opposition values between two categories of data and, through probability rules, predict their recurrence in the future.
- **Neural Networks:** Based on the model of brain organization and connectivity, this algorithm links decision nodes and enables the deciphering and identification of the processing path of a given data point.
- **Logistic Regression:** One of the most widely used algorithm. It is employed to predict the dominant variant of a group of independent variables, thereby forecasting an outcome.

## WHO figures on tuberculosis (2021)



age, population group to which it belongs, Bogota locality of residence, suffering or not of HIV-AIDS and whether they had received retroviral treatment. Subsequently, the results obtained in the company of medical personnel were analyzed.

“It was difficult to establish which data were useful and which were not, because there was a lack of information,” Orjuela explains. “We detected several inconsistencies, such as people who had started treatment long before they entered the hospital records. In those cases we turned directly to the doctors. They knew the details of the data acquisition, which served quite well.”



The research, which was expected to be concluded in 2020, also suffered from the consequences of the COVID-19 pandemic, as movement restrictions conditioned the phase of prospective studies that the research team had designed and forced the researchers to focus on retrospective calculations. However, in 2021 they concluded their implementation and a year later they shared their findings: first, in a forum with INS where they received interest from different scientific communities to expand the scope of their analysis, and second, in the article “[Machine learning in the loop for tuberculosis diagnosis support](#),” published in the specialized journal *Frontiers in Public Health*.

However, more work is still needed to make the research clearly practical. In Orjuela's opinion, it would be ideal to include this analysis model, along with the technology it employs, in a mobile application that allows doctors to prioritize the limited resources at their disposal when deciding what treatment to apply to a patient with symptoms of tuberculosis: “It is not a lie that the health infrastructure in the country is poor or insufficient. In some areas there is only one health post, not specialized medical staff or laboratories. That is why with this development we are looking to give a hand to that doctor who is in those areas so that he, depending on his expertise, knows what he can do with his patients.”

For Awad, however, this implementation is a milestone. His memory is very clear about all the efforts that have been made to make tuberculosis just a bad memory: from the first sanitariums of the early 20th century, copied from European models to combat the disease, through the coordinated global actions in the time of the explosion of HIV-AIDS (which weakens the defense system of patients and increases their vulnerability to the tuberculous bacillus), until the current moment, when the [National Plan for its eradication](#) contemplates coordinated protocols and actions to mitigate the resistance of the microorganism to medicines.

His years of experience allow him to elucidate that these efforts will overcome the reality of the Colombian health system and one of its main obstacles: limited resources. “This is an absolutely lucid investigation. Here what we are doing is trying to select a little better which patient to treat and with which procedure”, concludes the expert. “It is certainly an important step against TB, because AI is going to keep going and so is scientific research.” ■