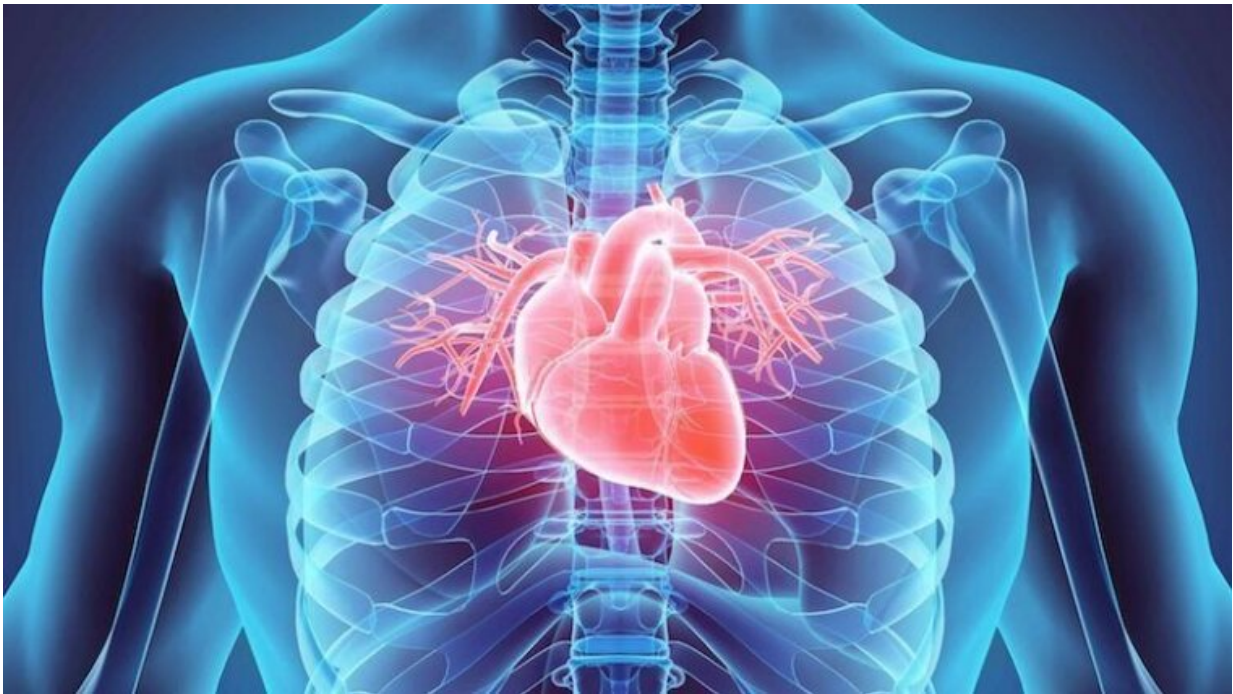# Using machine learning to estimate risk of cardiovascular death

September 13 2019, by Rachel Gordon

Using just the first 15 minutes of a patient's electrocardiogram (ECG) signal, an MIT system produces a score that places patients into different risk categories. Credit: Massachusetts Institute of Technology

Humans are inherently risk-averse: We spend our days calculating routes and routines, taking precautionary measures to avoid disease, danger, and despair.

Still, our measures for controlling the inner workings of our biology can be a little more unruly.

With that in mind, a team from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) came up with a new system for better predicting [health outcomes](#): a [machine learning model](#) that can estimate, from the electrical activity of their heart, a patient's risk of [cardiovascular death](#).

The system, called "RiskCardio," focuses on patients who have survived an [acute coronary syndrome](#) (ACS), which refers to a range of conditions where there's a reduction or blockage of blood to the heart. Using just the first 15 minutes of a patient's raw electrocardiogram (ECG) signal, the tool produces a score that places patients into different risk categories.

RiskCardio's high-risk patients—patients in the top quartile—were nearly seven times more likely to die of cardiovascular death when compared to the low-risk group in the bottom quartile. By comparison, patients identified as high risk by the most common existing risk metrics were only three times more likely to suffer an adverse event compared to their low-risk counterparts.

"We're looking at the data problem of how we can incorporate very long time series into risk scores, and the clinical problem of how we can help doctors identify patients at high risk after an acute coronary event," says Divya Shanmugam, lead author on a new paper about RiskCardio. "The intersection of machine learning and healthcare is replete with combinations like this—a compelling computer science problem with potential real-world impact."

Risky business

Previous machine learning models have attempted to get a handle on risk by either making use of external patient information like age or weight, or using knowledge and expertise specific to the system—more broadly known as domain-specific knowledge—to help their model select different features.

RiskCardio, however, uses just the patients' raw ECG signal, with no additional information.

Say a patient checks into the hospital following an ACS. After intake, a physician would first estimate the risk of cardiovascular death or heart attack using medical data and lengthy tests, and then choose a course of treatment.

RiskCardio aims to improve that first step of estimating risk. To do this, the system separates a patient's signal into sets of consecutive beats, with the idea that variability between adjacent beats is telling of downstream risk. The system was trained using data from a study of past patients.

To get the model up and running, the team first separated each patient's signal into a collection of adjacent heart beats. They then assigned a label—i.e., whether or not the patient died of cardiovascular death—to each set of adjacent heartbeats. The researchers trained the model to classify each pair of adjacent heartbeats to its patient outcome: Heartbeats from patients who died were labeled "risky," while heartbeats from patients who survived were labeled "normal."

Given a new patient, the team created a risk score by averaging the patient prediction from each set of adjacent heartbeats.

Within the first 15 minutes of a patient experiencing an ACS, there was enough information to estimate whether or not they would suffer from cardiovascular death within 30, 60, 90 or 365 days.

Still, calculating a risk score from just the ECG signal is no simple task. The signals are very long, and as the number of inputs to a model increase, it becomes harder to learn the relationship between those inputs.

The team tested the model by producing risk scores for a set of patients. Then, they measured how much more likely a patient would suffer from cardiovascular death as a high-risk patient when compared to a set of low-risk patients. They found that in roughly 1,250 post-ACS patients, 28 would die of cardiovascular death within a year. Using the proposed risk score, 19 of those 28 patients were classified as high-risk.

In the future, the team hopes to make the dataset more inclusive to account for different ages, ethnicities, and genders. They also plan to examine medical scenarios where there's a lot of poorly labeled or unlabeled data, and evaluate how their system processes and handles that information to account for more ambiguous cases.

"Machine learning is particularly good at identifying patterns, which is deeply relevant to assessing patient risk,'" says Shanmugam. "Risk scores are useful for communicating patient state, which is valuable in making efficient care decisions."

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology