



MACHINE LEARNING IN HEALTHCARE

Introduction to Precise Health, Vol. 1

THINKING SOFTWARE FOR LIFE



Precision Driven Health (PDH) is a seven-year multi-million-dollar research partnership aimed at improving health outcomes through data science. It will harness the unique combination of existing electronic healthcare data and world-class research capability to enable the development of data-driven healthcare solutions that can be applied globally.

For more information,
visit [**precisiondrivenhealth.com**](https://precisiondrivenhealth.com)



Orion Health is a health technology company that provides solutions which enable healthcare to over 110 million patients globally. Its open technology platform, Orion Health Amadeus, seamlessly integrates all forms of relevant data to enable population and personalized healthcare around the world. The company employs over 1,200 people globally and is committed to continual innovation, investing substantially in research and development to cement its position at the forefront of precision medicine.

For more information,
visit [**orionhealth.com**](https://orionhealth.com)

Contents

- > [A Tsunami of Information](#)
- > [A Short Introduction to Machine Learning](#)
- > [Machine Learning Models](#)
 - > [Classification](#)
 - > [Clustering](#)
 - > [Regression](#)
- > [Deep Learning](#)
- > [Applying Machine Learning to Healthcare](#)
- > [Machine Learning Timeline](#)
- > [Glossary](#)



A Tsunami of Information

Introduction by Ian McCrae, Orion Health Founder and CEO

15 minutes. That's how long your doctor has to see you, assess your complaint, diagnose a solution, and see you out the door—hopefully on the path back to wellness.

This isn't much time when you consider the wealth of information that he or she has to consider: your patient record, the medical research relevant to your complaint, the answers about your condition that you provide, and the basic examination ("say aaaaaaah") that is carried out.

So how will your doctor cope when faced with the tsunami of healthcare information that will occur when it is routine for your patient record to include data about your genome, your microbiome (microorganisms in your body) and your fitness regime?

Your electronic health record is fast becoming the most powerful tool in the medical toolkit. All the information will be stored in the cloud. It will have to be, because the size of the electronic file containing your complete patient record is estimated to be as much as six terabytes. That's a quarter of the whole of Wikipedia!



HOME



CONTENTS

A data file that large is required to enable the practice of precision medicine. This a new revolution in healthcare. It is the ability to target healthcare treatment specifically for an individual.

In addition to improving health outcomes, precision medicine will save vital health dollars because it is enabled by unique data insights that lead to more targeted treatments.

For example, when I had my genome mapped I found out that ibuprofen doesn't work for me. The only effect that drug has is on my wallet—that is, what it costs me to buy the pills.

So how will doctors process all this new information in the short time in which they see their patients? The truth is they can't. It will require high-powered computing that uses insights from machine learning— a type of artificial intelligence that enables computers to find hidden insights without being programmed. Algorithms will scan vast data sets and identify recommended treatment plans tailored to individuals.

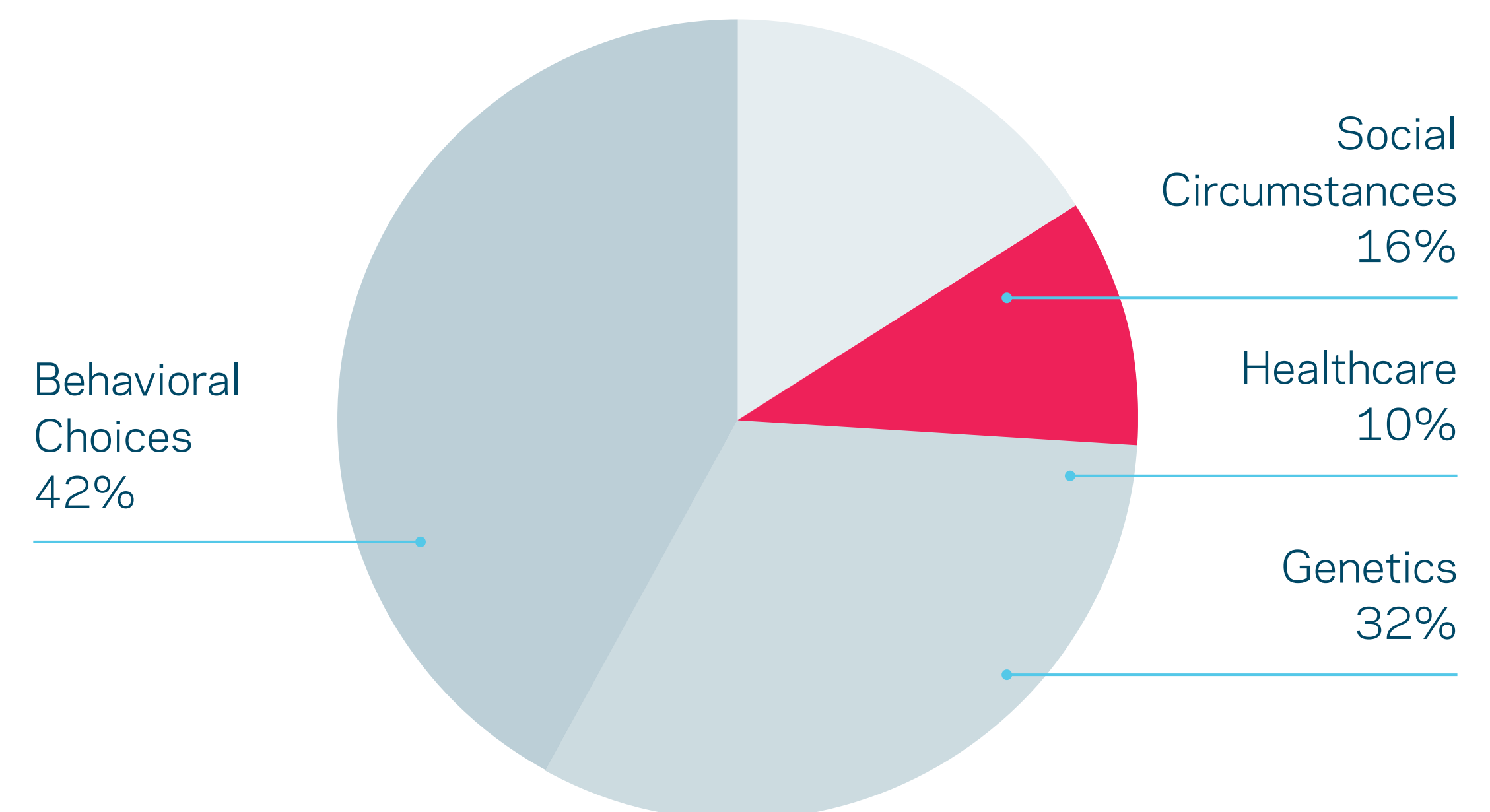
There are two things required for the successful application of machine learning in healthcare— intelligent algorithms and rich data sets.

Orion Health is at the forefront of developing both areas. We have invested in the Precision Driven Health initiative, a world-leading, multi-million-dollar research program that

is investigating how the application of machine learning will enable new healthcare solutions that are more precisely tailored to a person's unique characteristics. And since our software manages over 110 million patient health records globally, we believe we're one of the few health software companies in the world capable of carrying out machine learning analysis.

We have produced this brief introduction to machine learning because this is an exciting time to be part of the global healthcare sector. As a discipline, health is transforming into a mathematical science, but at its heart, it will always be about enabling the perfect care for everyone, anywhere in the world.

Factors Influencing Health



McGinnis et. al. Human Affairs, Vol 22 (2)

Useful Data for the Practice of Precision Medicine

Social Data

Personal circumstances, such as living situation and income

Device Data

Information collected from apps that measure fitness and sleeping, electronic inhalers, etc.

Metabolome

Chemicals which are created, modified and broken down by bodily processes such as enzymatic reactions

Transcriptome

Messages created from DNA to form the template (mRNA) of proteins

Genome

Patient's complete set of genes "written" in DNA

Clinical Data

Patient's medical record

Exposome

Impact of the external environment, such as pollution and tobacco smoke, etc.

Microbiome

Collective name for 100 trillion microorganisms living inside us

Proteome

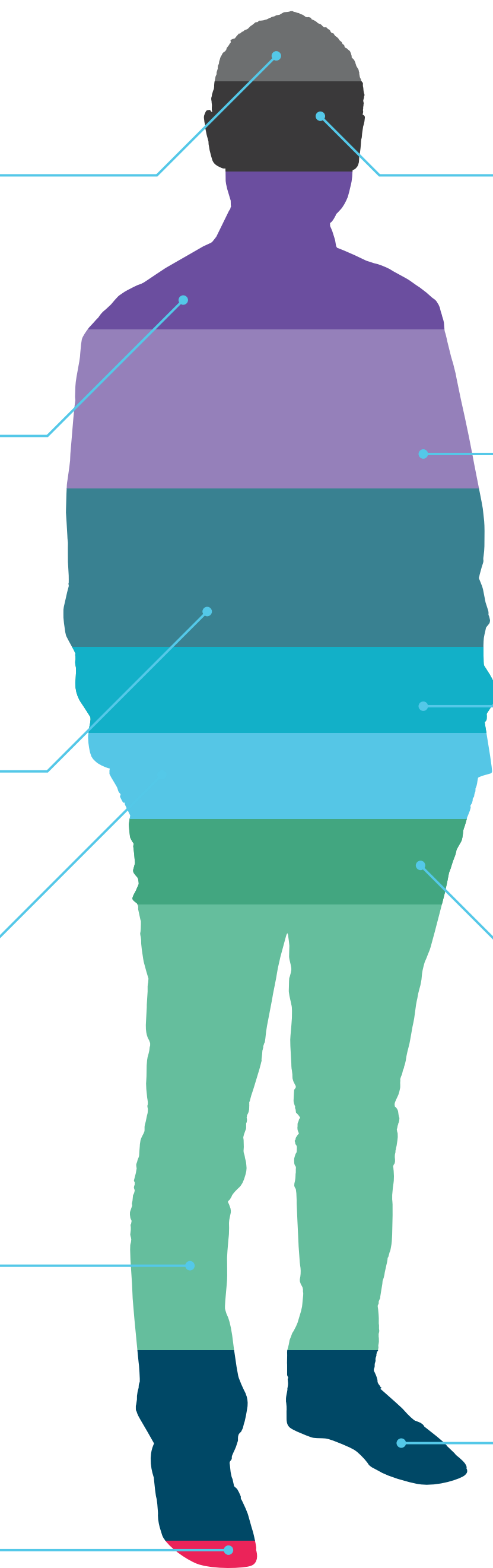
System of proteins, including enzymes, which are the building blocks of the body

Epigenetic (Methylome)

The set of nuclear and methylation modifications in a human genome

Imaging

Medical images, such as x-rays, scans, ultrasounds, etc.



HOME



CONTENTS

A Short Introduction to Machine Learning

By Dr. Kathryn Hempstalk, PhD Computer Science

Self-driving cars, Siri, and websites that recommend items based on the purchasing decisions of other people: what do these have in common? They are all real-world examples of machine learning.

Machine learning is when a computer uses an algorithm to understand the data it has been given and recognizes patterns.

We call the process of learning “training”, and the output that this process produces is called a “model.” A model can be provided new data and reason against it based on what the model has already learned.

Machine learning models determine a set of rules using vast amounts of computing power that a human brain would be incapable of processing. The more data a machine learning model is fed, the more complex the rules and the more accurate the predictions.

Whereas a statistical model is likely to have an inherent logic that can be understood by most people, the rules created by machine learning are often beyond human comprehension because our brains are incapable of digesting and analyzing enormous data sets.

In the following section, we will look at three types of models used in machine learning—classification, clustering, and regression—and review the concept of deep learning.

Classification

The purpose of the classification model is to determine a label or category—it is either one thing or the other. We train the model using a set of labeled data.

As an example, if we wanted to predict whether a person's mole is cancerous or not, we would create a model using a data set of mole scans from 1,000 patients that a doctor has already examined for cancer. We would also feed the model lots of other data, such as a patient's age, gender, ethnicity, and place of residence.

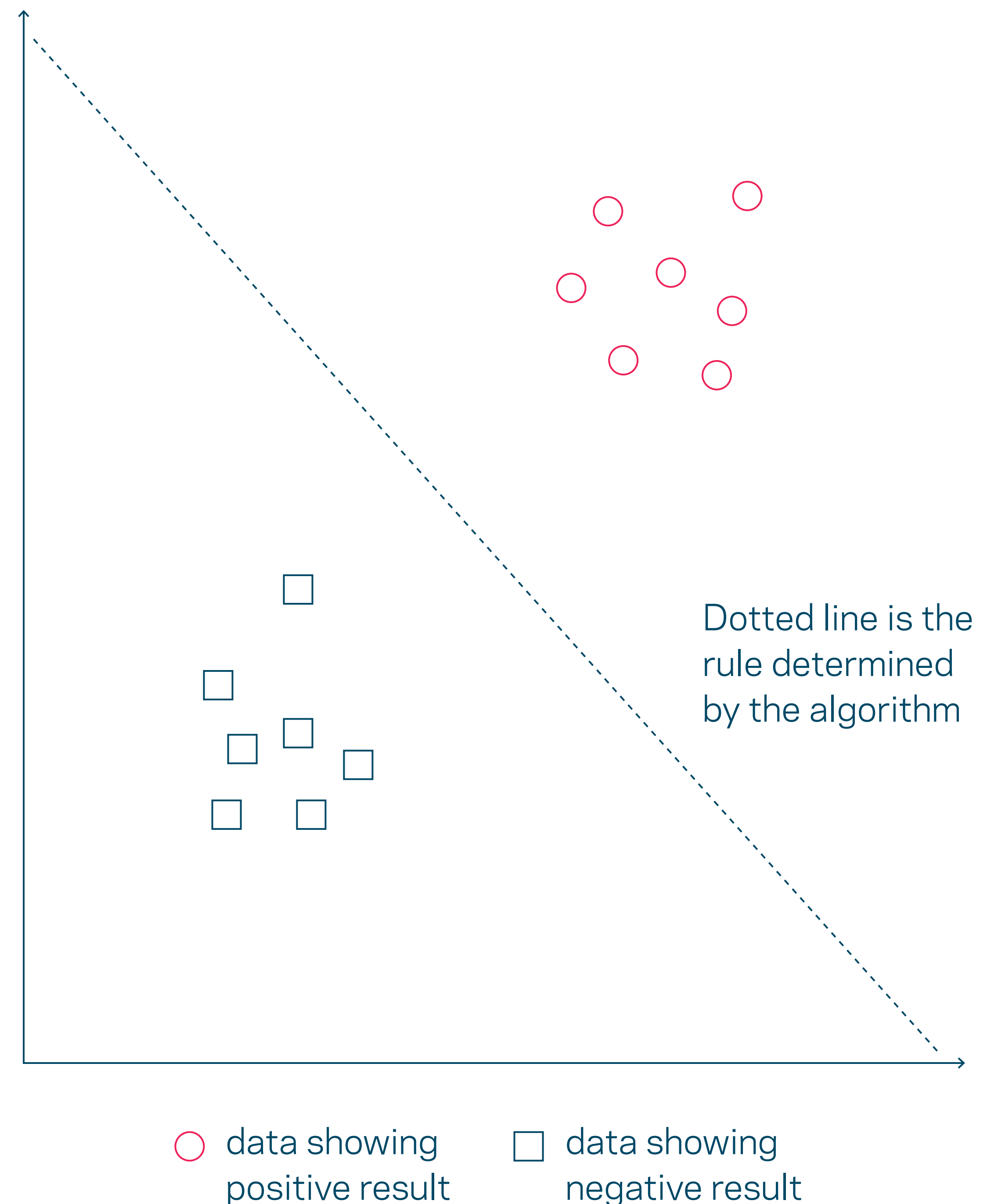
We would then create a model which will enable us to decide whether it depicts cancer.

The graph (Fig. 1) is an example of a classification model. The dotted line is the rule determined by the model that best separates the positive and negative results, based on the data provided during training.

If we were using our mole scan example, then in this graph the circles are the mole scans that show cancer is likely, and the squares are the mole scans that show that cancer is unlikely.

Each time we train the algorithm with additional data, it should improve the accuracy of the prediction.

Classification Model (Fig. 1)



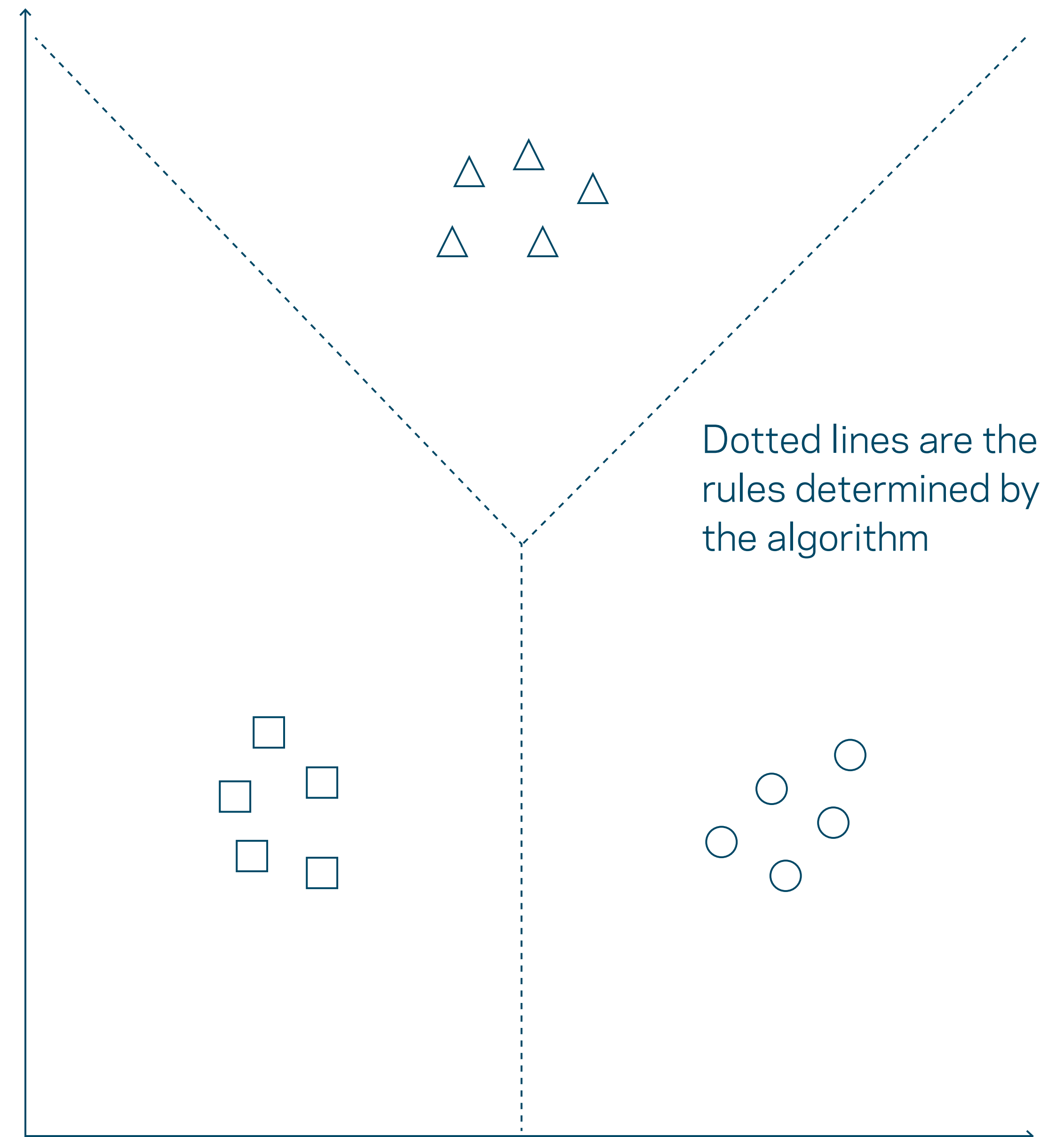
Clustering

We would create a clustering model if we had a whole bunch of data, didn't have a determined outcome, and just wanted to see if there were any distinctive patterns.

An example might be that of a primary care provider—in an effort to see if there is a pattern to the type of complaints she has seen in the past year—creating a model based on a year's worth of patient data. The model might show that people living in a certain area have higher rates of cancer, which would warrant further examination (e.g., is there an issue with the water supply? Is there pollution from a nearby factory?).

The graph (Fig. 2) shows the rules (dotted lines) determined by the algorithm using the data set. The patient files are clustered into various groups, but a human must label each group.

Clustering Model (Fig. 2)



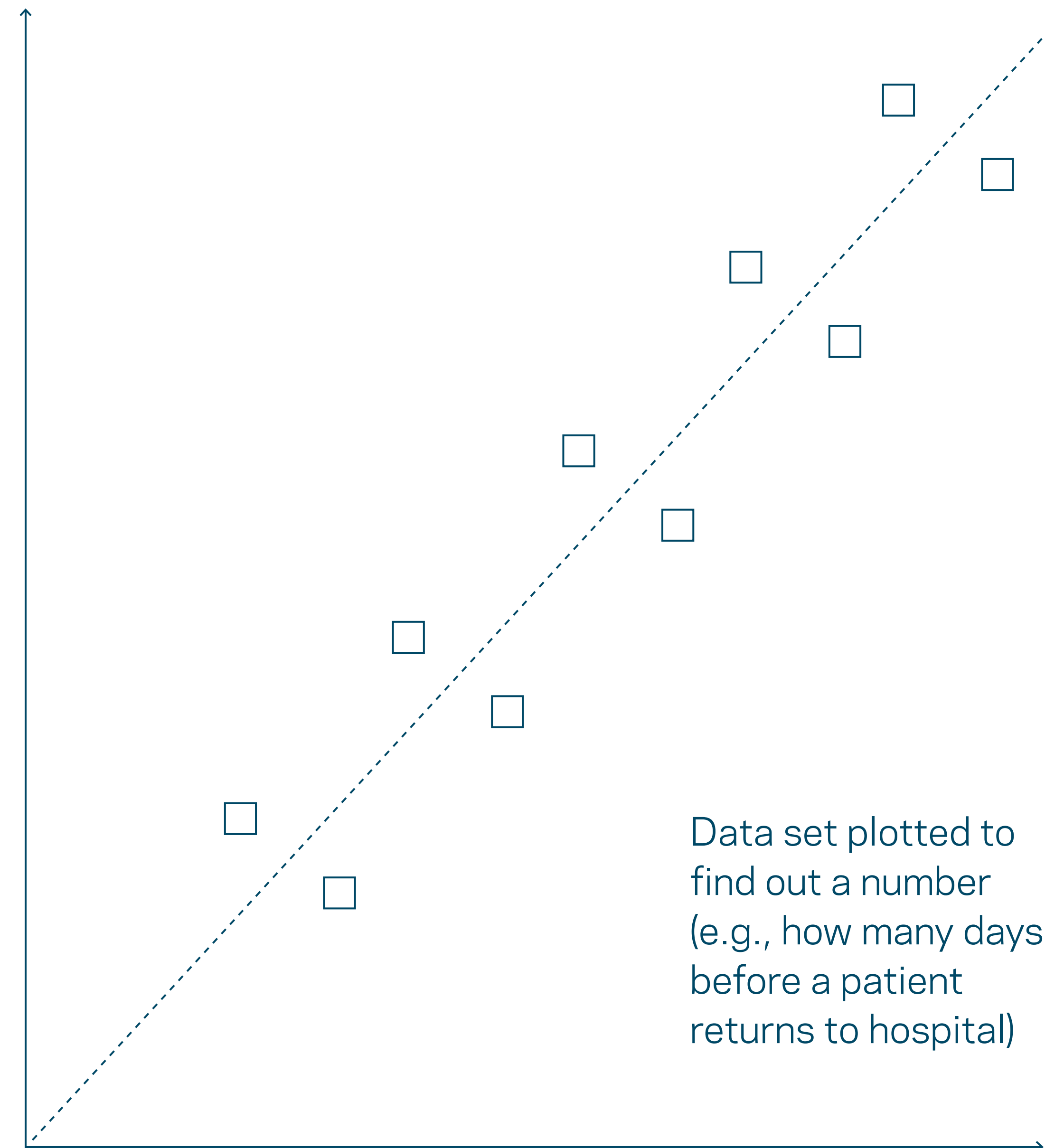
Regression

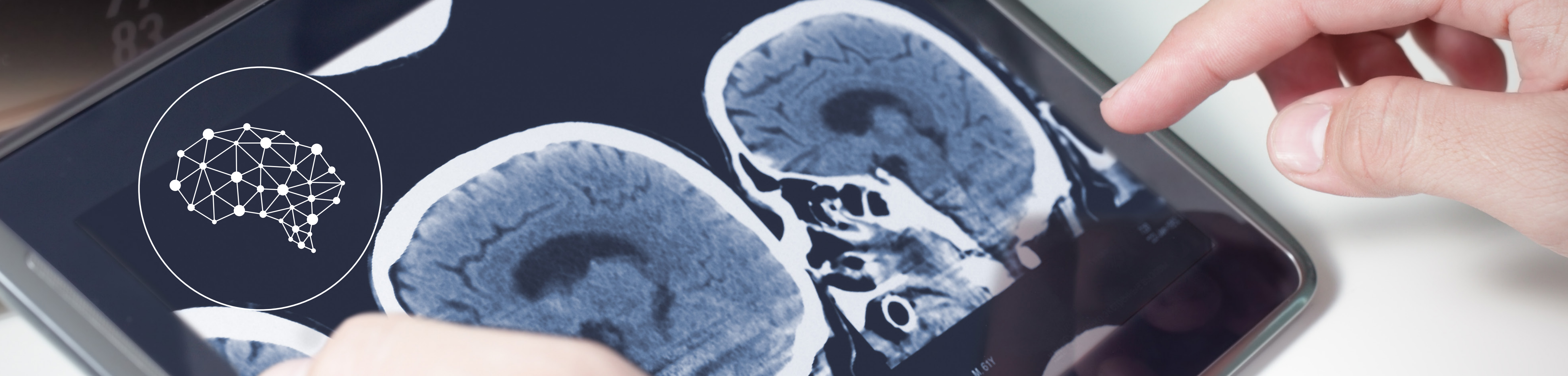
A regression model is created when we want to find out a number (e.g., how many days before a patient discharged from the hospital with a chronic condition such as diabetes will return).

We start with a data set that is already labeled with numbers (e.g, how many days before each patient returns to the hospital) and feed this and the patient's other data (e.g., age, gender, etc.) into the model. The model tries to learn a pattern that describes how long before a patient returns to the hospital.

The graph (Fig. 3) plots various patients against the number of days between hospitalizations, and the dotted line is the model that best predicts the outcome. This regression model predicts the time before rehospitalization for a new patient.

Regression Model (Fig. 3)





Deep Learning

“Deep learning” is an often-misused buzzphrase that refers to a special case of a machine learning algorithm: the artificial neural network.

A neural network is an algorithm inspired by the ways a brain works. It involves many nodes (or neurons) that are often connected together in layers to form a network. A neural network must have at least two layers—a layer of inputs and a layer of outputs.

There may be many “hidden” layers between the input layer and output layers, and these are used to extract more information by exploiting structure in the data. A network is considered “deep” if it has more than one hidden layer (see the “Deep Learning Neural Network” diagram [opposite], which illustrates the complexity of a neural network).

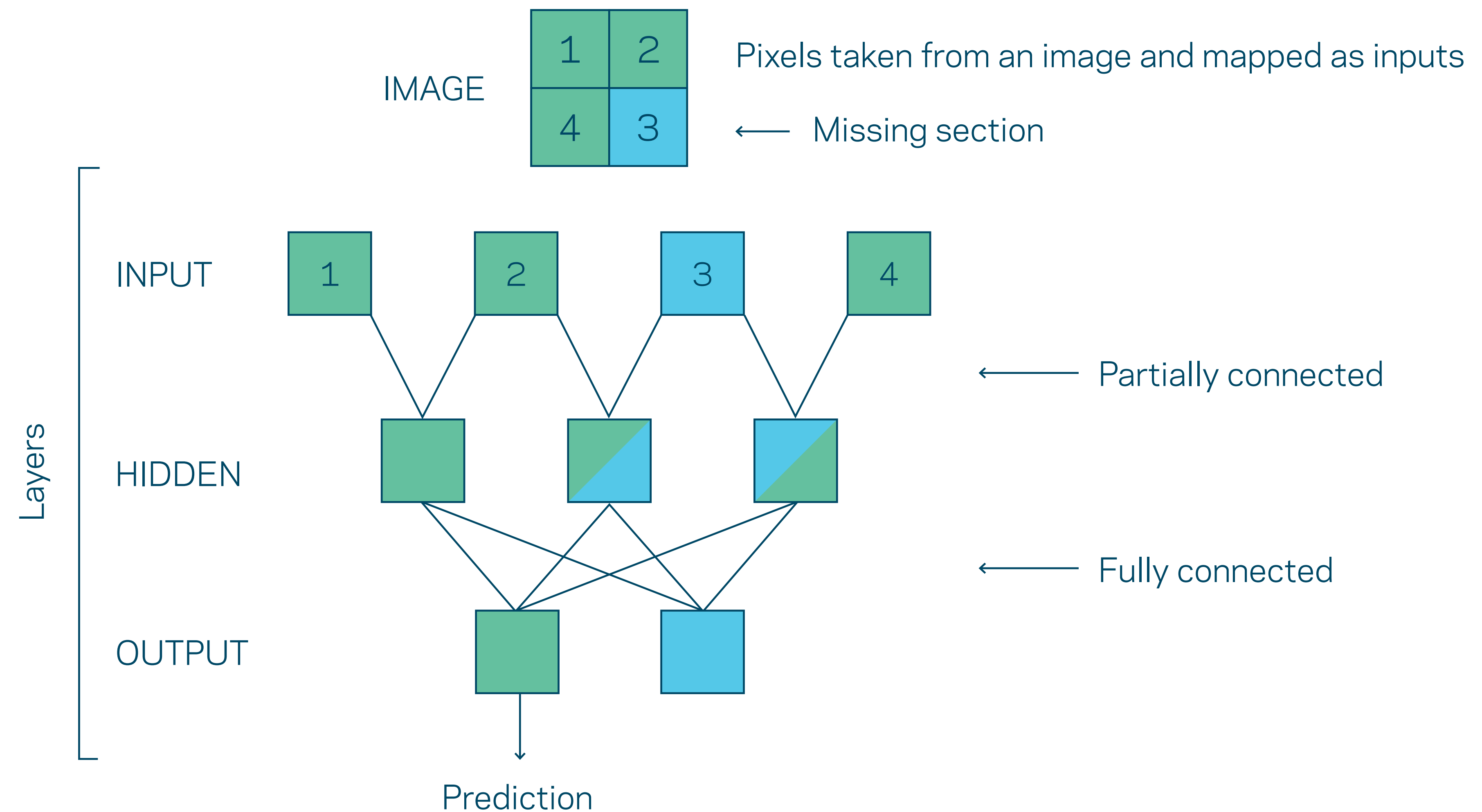
Neural networks are great at solving problems where the data is highly structured—like an image of a brain scan—but they are “black box” algorithms. In other words, it is hard to describe to a human being the patterns found by neural networks.

Despite being around for over 50 years, neural networks have only become popular and feasible in the last 10 years thanks to advances in both algorithm design and computing power.

Data is critical

Many major advances in machine learning actually occur from data sets, not algorithms. When IBM's Deep Blue defeated chess grandmaster Garry Kasparov, the algorithm was 14 years old, but the data set was only six years old. Google learned to translate Chinese to English using a 17-year-old algorithm and data they collected in 2015. On average, the time it takes for a major breakthrough to be made is 18 years for algorithms, but only three years for data sets.

Deep Learning Neural Network

[HOME](#)[CONTENTS](#)

Applying Machine Learning to Healthcare

By Dr. Kevin Ross, General Manager, Precision Driven Health

While the healthcare sector is being transformed by the ability to record massive amounts of information about patients, the enormous volume of data being collected is impossible for human beings to analyze. Machine learning provides a way to find patterns and reason about data, which enables healthcare professionals to move to personalized care known as precision medicine.

There are many possibilities for how machine learning can be used in healthcare, and all of them depend on having sufficient data and permission to use it.

Previously, alerts and recommendations for a medical practice have been developed based on external studies, and hard-coded into their software. However, that can limit the accuracy of that data because it might come from different populations and environments. Machine learning, on the other hand, can be refined using data that is available in that particular environment (e.g., anonymized patient information from a hospital and the area it serves).

An example of how healthcare providers can take advantage of machine learning is offering the ability to predict hospital readmission for chronically ill patients.

Identifying those patients most at risk of being readmitted makes it possible for providers to offer better post-discharge support. By lowering the rate of readmission, it not only improves the lives of those most at risk, it also helps save precious healthcare dollars, which can be used for wellness and prevention programs.

As an example, we undertook preliminary research on a publicly available data set of 100,000 anonymized diabetic patient records from 130 U.S. hospitals.

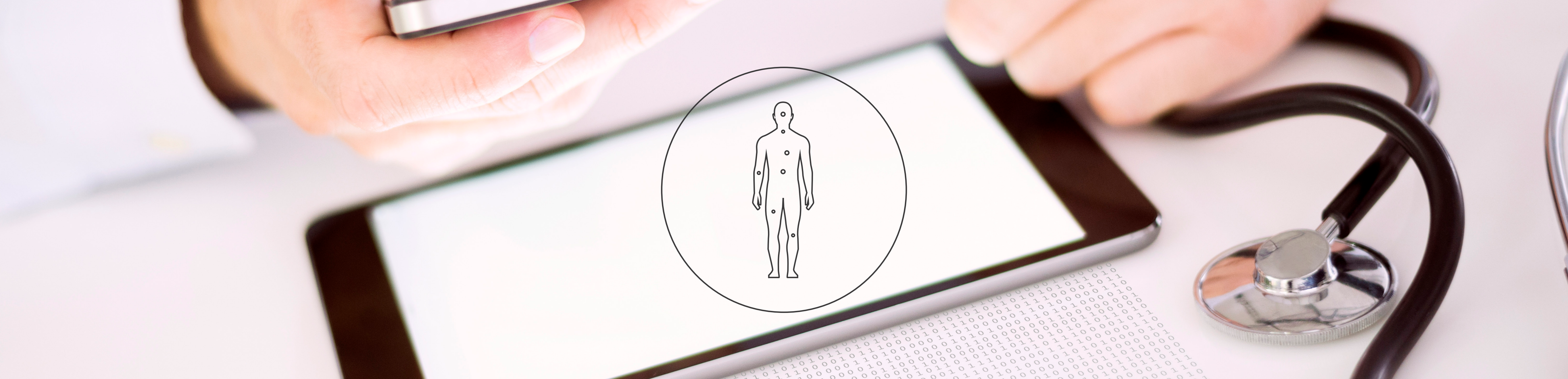
The analysis showed that machine learning approaches were 20 percent better at assessing the readmission risk of patients than the standard LACE risk-scoring approach, currently used by U.S. healthcare providers.



HOME



CONTENTS



LACE is calculated based on the patient's **L**ength of stay, **A**cute admission through emergency, **C**omorbidities (other illnesses), and **E**mergency department visits in the past six months. The machine learning models achieved a greater accuracy because they were able to explore patient and disease-specific factors—in addition to the factors already considered by LACE.

Those patients determined by tools such as LACE to be at the highest risk of readmission receive the greatest amount of post-acute care. The more accurate the tool for identifying at-risk patients, the more targeted healthcare intervention can be in order to reduce readmission rates and lower costs.

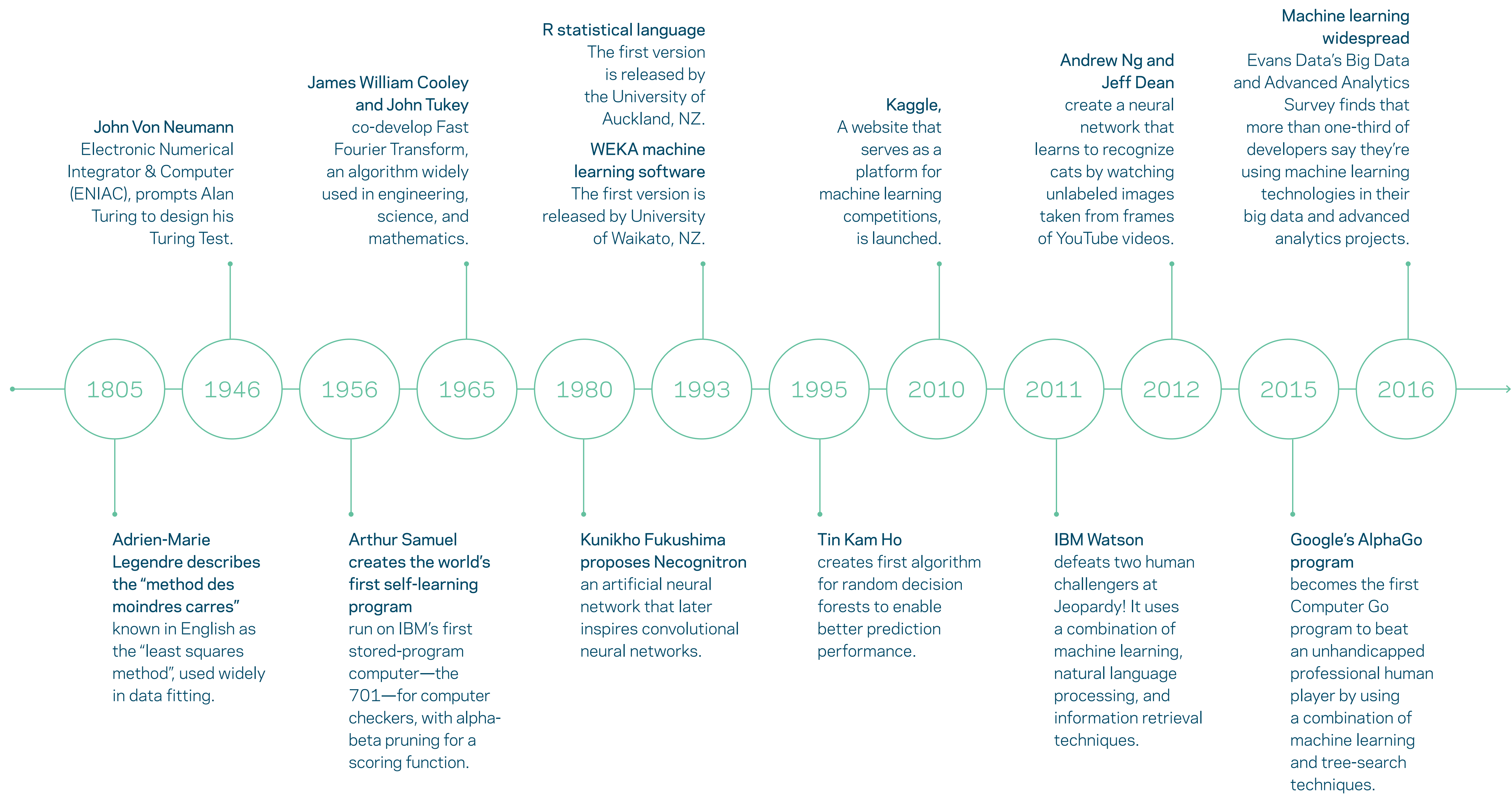
We calculated that the potential savings from the machine learning models were four times higher than those from LACE.

Predicting the risk of readmission following a hospital stay is just one example of how machine learning can be applied to solve some of most pressing issues in healthcare delivery. Other examples include:

- Finding combinations of drugs that should not be taken together
- Classifying imagery, such as mole scans, to identify disease
- Assisting with decisions about what condition a patient might have, or what treatment might work the best

This is an exciting field that seeks to assist healthcare providers—whether practicing in the hospital or in the community—in creating better health outcomes for their patients. We have only just begun to explore the possibilities.

Machine Learning Timeline



Glossary

Artificial Intelligence

Intelligence exhibited by machines. In computer science, the ideal “intelligent” machine is a flexible rational agent that perceives its environment and takes actions that maximize its chance of success at some goal.

Augmented Reality

A live or direct view of the physical world, supplemented by computer-generated sensory input (e.g., sound, video, graphics, or GPS data).

Big Data

Large and complex data sets that may be analyzed computationally to reveal patterns, trends, and associations. Structured, unstructured, and semi-structured data can be mined for learning. To use machine learning and artificial intelligence effectively, having a large data set is crucial.

Computational Techniques

Different approaches and methods to solving problems using systems/techniques that are mathematical or can be engineered and inputted into a computer.

Computer Learning

Another name for machine learning. It describes the process of how computers are “learning” through human input and training.

Data Mining

Can be considered a superset of many different methods to extract insights from data. It might involve traditional statistical methods and machine learning.

Deep Learning

A term coined by Geoffrey Hinton in 2006. It combines advances in computing power and special types of neural networks to learn complicated patterns in large amounts of data. It is a branch of machine learning.

Internet of Things

An increase in machine-to-machine communication and development of the Internet in which everyday objects have network connectivity. This will allow for constant instantaneous connection and data sending/receiving in real-time.

Machine Learning

Another name for computer learning. The goal of machine learning is to understand the structure of data so that accurate predictions can be made based on the properties of that data.

Operational Intelligence

A category of real-time, dynamic business analytics—delivers insight into day-to-day business operations. It offers an understanding of IT systems and technology structure within the business to allow informed decisions.

Precision Medicine

A medical model that proposes the customization of healthcare, with treatments, products, medical decisions, and practices being tailored to each patient.

Predictive Analytics

An area of data mining that deals with extracting information from data and using the information to predict trends and behavior patterns. It is used to make predictions about unknown future events.



HOME



CONTENTS

