ORION HEALTH

Introduction to Machine Learning in Healthcare

ORION HEALTH

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 pathway 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, 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 (bugs 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 (24Tbs)!

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.

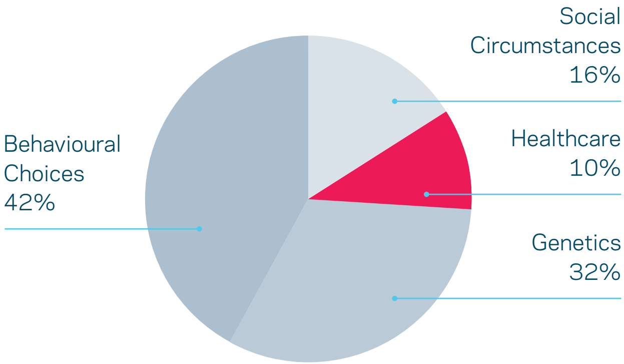
So how will a doctor process all this new information in the short time in which they see their patient? The truth is they can't. It will require high-powered computing, using insights from machine learning – a type of artificial intelligence that enables computers to find hidden insights without being programmed. Algorithms will interrogate vast data sets and surface 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.

At Orion Health we are at the forefront of developing both areas. We have invested in a world-leading, multi-million-dollar research initiative called the Precision Driven Health. It 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 our software manages over 100 million patient health records globally, making us one of the few health software companies in the world capable of carrying out machine learning analysis.

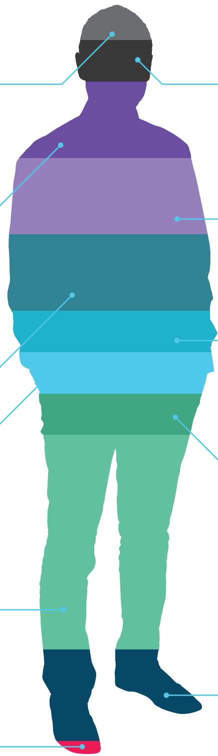
We believe this is an exciting time to be part of the global healthcare sector and so we have produced this brief introduction to machine learning. As a discipline health is transforming into a mathematical science, but at its heart it will always be about enabling the perfect care for everyone, anywhere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

Factors Influencing Health



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

Data useful 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 microscopic bugs living inside us

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

Epigenetic (Methylome)
The set of nucleric and methylation modifications in a human genome

Imaging
Medical images, such as x-rays, scans, ultrasound

A Short Introduction to Machine Learning

By Dr Kathryn Hempstalk / Senior Data Scientist, Precision Driven Health

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 examples of machine learning being used in the real world.

Machine learning is when a computer has been taught to recognise patterns by providing it with data and an algorithm to help understand that data.

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 it can reason about this new information based on what it has previously 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 analysing enormous data sets.

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

Classification

The purpose of the Classification model is to determine a label or category – it is either one thing or another. We train the model using a set of labelled data.

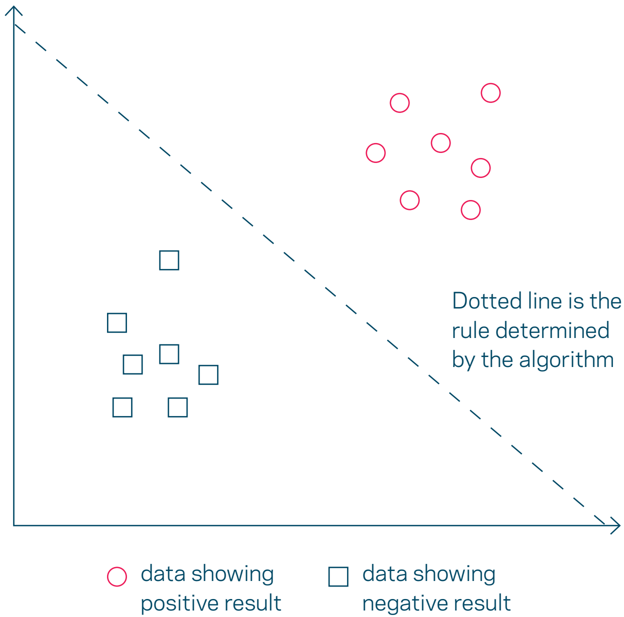
As an example, we want to predict if a person's mole is cancerous or not, so we create a model using a data set of mole scans from 1000 patients that a doctor has already examined to determine whether they show cancer or not. We also feed the model a whole bunch of other data such as a patient's age, gender, ethnicity, and place of residence.

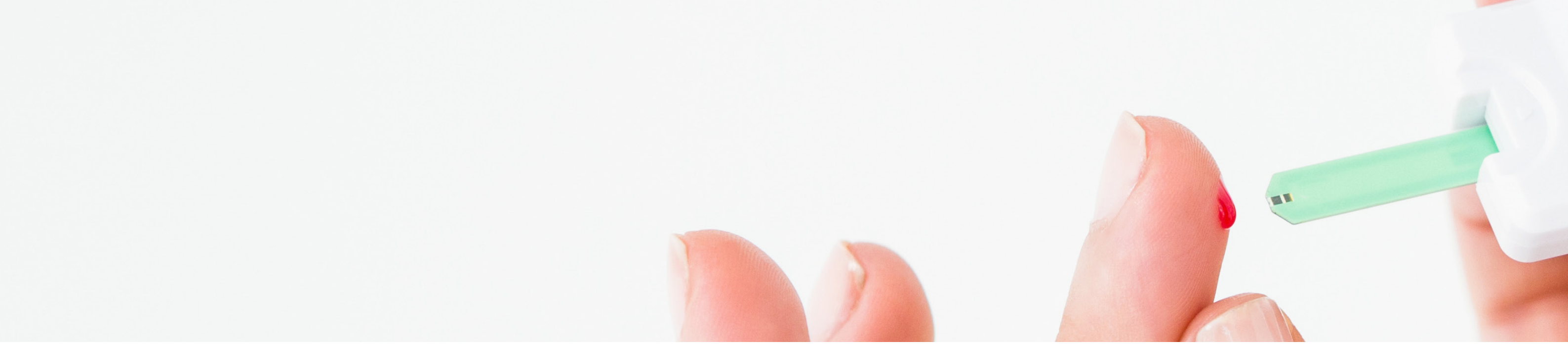
We then create a model which will enable us to present a new mole scan and decide if it depicts cancer or not.

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 it was given 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)





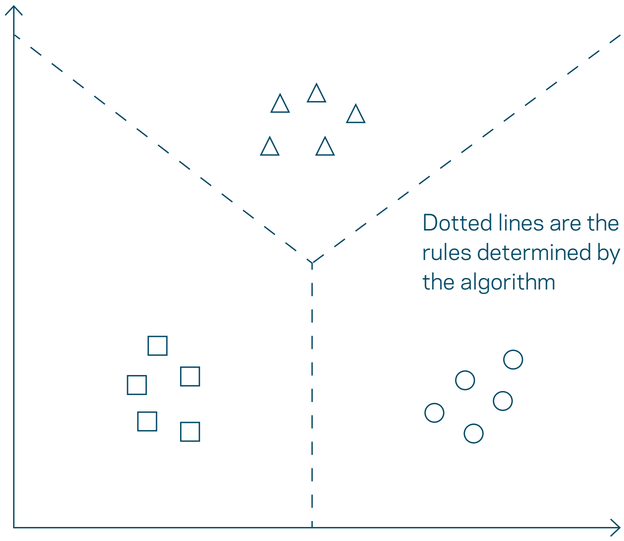
Clustering

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

An example might be that a GP creates a model based on a year's worth of patient data, to see if there is a pattern to the type of complaints she has seen in that year. The model might show that people living in a certain area have higher rates of cancer and this warrants further examination (is there an issue with 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. It has clustered the patient files into various groups, but it is up to a human to label each group.

Clustering Model (Fig. 2)



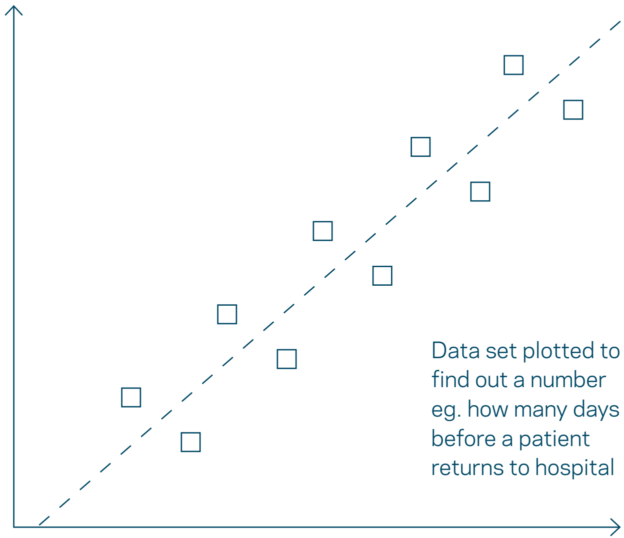
Regression

A Regression model is created when we want to find out a number – for example how many days before a patient discharged from hospital with a chronic condition such as diabetes will return.

We start with a data set that is already labelled with numbers (in this case, the number of days before each patient returns) and feed this and the patient's other data (such as age, gender etc) into the model. The model tries to learn a pattern that describes how long before a patient returns to hospital.

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

Regression Model (Fig. 3)



Deep learning

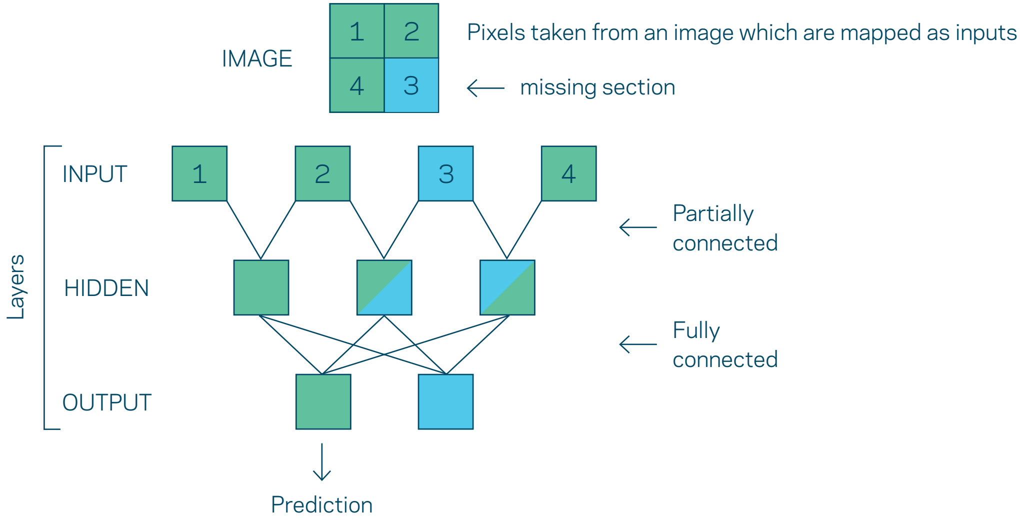
Deep learning is a buzzword we hear a lot but is often misused, in reality it is just a special case of one machine learning algorithm: artificial neural networks.

A neural network is an algorithm that was inspired by the ways a brain works, and 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 layer, 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 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 are a 'black box' algorithm. In other words, it is hard to describe the patterns they find to a human being.

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.

Deep Learning Neural Network



Data is critical

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

Applying Machine Learning to Healthcare

By Dr Kevin Ross / Research Director, Precision Driven Health

While the healthcare sector is being transformed by the ability to record massive amounts of information about individual patients, the enormous volume of data being collected is impossible for human beings to analyse. Machine learning provides a way to automatically find patterns and reason about data, which enables healthcare professionals to move to personalised 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 they might be from different populations and environments.

Machine learning, on the other hand, can be refined using data that is available in that particular environment. For example, anonymised patient record information from a hospital and the area in which it serves.

An illustration of how healthcare providers can take advantage of machine learning is being able to predict hospital re-admission for chronically ill patients.

Identifying those patients most at risk of being re-admitted means they can have greater support after they have been discharged. 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 programmes.

As an example we undertook preliminary research on a publicly available dataset of 100,000 anonymised diabetic patient records from 130 US hospitals.

The analysis showed that machine learning approaches were 20% better at assessing the readmission risk of patients than the standard LACE risk scoring approach, currently used by US healthcare providers. LACE is calculated based on the patient's Length of stay, **A**cute admission through emergency, **C**o-Morbities (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 re-admission receive the greatest amount of post-hospital care. The more accurate the tool for identifying at-risk patients, the more targeted healthcare intervention can be in order to reduce re-admission 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 are:

- find combinations of drugs that should not be taken together,
- classify imagery, such as mole scans, to identify disease,
- assist with decisions about what condition a patient might have, or what treatment might work the best for this patient.

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

1805

Adrien-Marie Legendre describes the "method des moindres carres"
known in English as the least squares method – used widely in data fitting.

1946

John Von Neumann Electronic
Numerical Integrator & Computer (ENIAC), prompted Alan Turing to design his Turing Test.

1956

Arthur Samuel, The World's first self-learning program
Checkers learning program, alpha-beta pruning, IBM machine.

1965

James William Cooley and John Tukey
co-develop Fast Fourier Transform, an algorithm widely used in engineering, science and mathematics.

1980

Kunikho Fukushima proposes Necognitron
an artificial neural network that later inspires convolutional neural networks.

1993

R statistical language
First version released by Auckland University in NZ.

WEKA machine learning software
First version released by Waikato University in NZ.

1995

Tin Kam Ho
creates first algorithm for Random decision forests to enable better prediction performance.

2010

The launch of Kaggle
Website that serves as a platform for machine learning competitions.

2011

IBM Watson
defeats two human challengers at Jeopardy! It used a combination of machine learning, natural language processing and information retrieval techniques.

2012

Andrew Ng and Jeff Dean
create a neural network that learns to recognise cats by watching unlabelled images taken from frames of YouTube videos.

2015

Googles AlphaGo program
becomes the first Computer Go program to beat an unhandicapped professional human player – using a combination of machine learning and tree search techniques.

2016

Machine learning widespread
Evans Data's Big Data and Advanced Analytics Survey found that more than one-third of developers say they're using machine learning technologies in their big data and advanced analytics projects.

Glossary

Key ideas associated with Machine Learning

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 maximise its chance of success at some goal.

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

Big Data
Large and complex data sets that may be analysed computationally to reveal patterns, trends and associations – structured, unstructured and semi-structured data can be mined for learning.

Having a large dataset is crucial to use Machine Learning and Artificial Intelligence effectively.

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 – 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. Might involve traditional statistical methods and machine learning.

Deep Learning

Deep learning (term coined by Geoffrey Hinton in 2006) 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

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. Grants understanding of IT systems and technology structure within the business - allows informed decisions.

Precision Medicine

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

Predictive analytics

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

To learn more about the work of
Precision Driven Health

precisiondrivenhealth.com
@healthprecision

Precision Medicine starts here

Orion Health is embedding machine learning analysis into its clinical suite of solutions, which are delivered as part of its ground-breaking precision medicine platform Amadeus. The company provides technology solutions that capture, store and analyse patient health care records. Its software manages over 100 million patient records worldwide. Orion Health employs over 1200 people globally and is dedicated to continual innovation and investment to cement its position at the forefront of precision medicine.

orionhealth.com