Data Protection: Machine Learning in Health

Table of Contents

Introduction

Artificial intelligence in healthcare is an exciting field. There is the strong possibility that these technologies can greatly improve healthcare by automating complex tasks. As with any new technology, their implementation and use pose challenges that must be overcome. Our challenge is to develop artificial intelligence in a way that preserves patient's data protection rights.

This document makes practical recommendations on how to structure and lead a project in healthcare that uses machine learning, a subset of artificial intelligence to preserve data protection rights. It draws upon Big data, artificial intelligence, machine learning and data protection published by the Information Commissioner.

What is Machine Learning?

Machine learning (ML) is a technology that allows computer systems to perform a specific task without explicit instructions, relying on models and inference instead. ML algorithms build a mathematical model of sample data, known as "training data", to make predictions or decisions without being explicitly programmed to perform the task.

A good example of a task suited to ML is the recognition of spam email. Some spam email will be about selling drugs online, others will try and solicit username and password details, still more will ask for money. If we tried to write a computer program that dealt with each one individually it would be very long and fail to recognise new forms of spam.

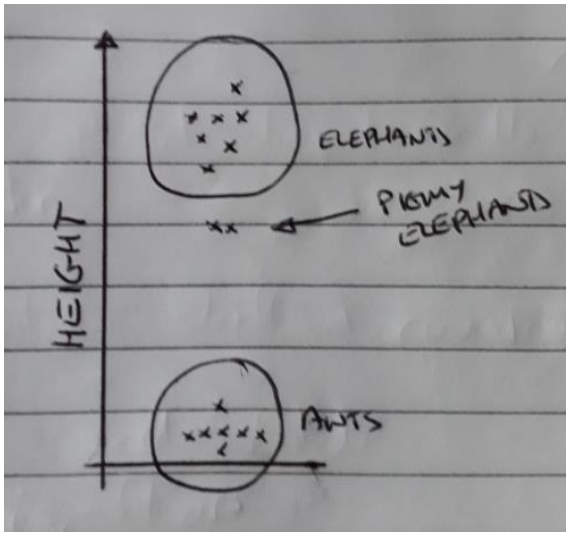
Rather than write a set of rules that define spam, the ML algorithm is shown a set of emails and told which is legitimate and which is spam. It then learns to recognise spam emails from the information provided and can make predictions when new emails are received.

Machine learning is not a solution to every problem. It is best used in the following situations:

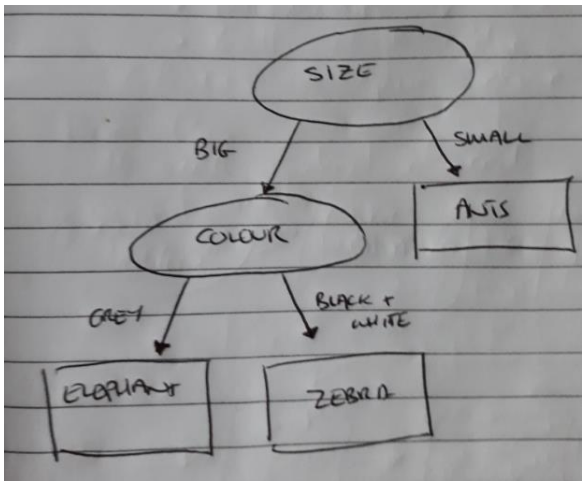
- **You cannot code the rules:** Many human tasks (such as recognizing whether an email is spam or not spam) cannot be adequately solved using a simple (deterministic), rule-based solution. Many factors could influence the answer. When rules depend on too many factors and many of these rules overlap or need to be tuned very finely, it soon becomes difficult for a human to accurately code the rules. You can use ML to effectively solve this problem.
- **You cannot scale:** You might be able to manually recognize a few hundred emails and decide whether they are spam or not. However, this task becomes tedious for millions of emails. ML solutions are effective at handling large-scale problems.

How does it Learn?

Machine learning looks for patterns in the data. It uses different mathematical techniques to do this. For example, imagine we had a group of ants and elephants and needed to separate them. Using the height of the animals we find two clusters of data - one reflecting ants and the other elephants. The model would be able to correctly identify smaller elephants even though it was outside the existing data.



Another type of algorithm is known as a decision tree. Rather than look for clusters of similar data it tries to separate them out by asking different question. First it could consider size, then colour.



The selection of the correct algorithms and training them is undertaken by Data Scientists. Their job is to understand the data and the best algorithms for the problem. They need to understand the inner workings of the algorithms and should be able to explain them to a layman.

Using Machine Learning

Machine learning is divided into different stages:

1. **Frame Problem -** The first step is to frame the core machine learning problem in terms of what is observed and what answer you want the model to predict. For example, a sepsis ML algorithm would be framed using patient observations (blood pressure, temperature, respiratory rate, etc) and the likelihood of sepsis developing.

2. **Prepare Data** - The algorithms need to be 'trained' using cleaned and prepared data. Continuing our example this would require collecting data on patients and whether they developed sepsis. The data needs to be of a high enough quality to train the model and free from bias.

3. **Process Features** – Often, the raw data (input variables) and answer (target) are not represented in a way that can be used to train a highly predictive model. This stage amends the data to make it more useful. For example, it may be that the time of admission is not relevant to predicting sepsis, but the A&E wait time is. The former would be discarded, and the latter calculated prior to building the model.

4. **Build Model** – Feed the resulting features to the learning algorithm to build models

5. **Evaluate Model** - ML models will give a prediction score demonstrating their confidence in an outcome. This can be evaluated using data that was held out from model building. For example, Babylon demonstrated the efficacy of their GP at Hand AI tool using questions from the MRCGP medical exam although this is refuted by the College. The evaluation chosen must be suitable for the task.

6. **Use Model** - Use the model to generate predictions of the target answer for new data instances. For example, use the sepsis algorithm to predict likelihood with new patients.

There is a perception from the popular press that artificial intelligence is somewhat magical. In practice it is driven by humans providing the structure using their own judgement. The best models have well framed problems and high-quality data sets.

Implications for Healthcare

Machine learning is very interesting for healthcare as it is full of questions where there are no set rules. For example, identifying patients at risk of falls or predicting medication side effects. These can be deployed to support clinicians in their decisions or to replace them entirely.

Decision support for clinicians is not a new concept. In the year 2000, PRODIGY was used for prescribing decision support by GPs. The LACE risk algorithm provides a calculation of readmission risk. In both these cases the results are a guide for the clinician and should not be followed blindly.

The expectation for machine learning goes beyond traditional decision support with the hope of automating clinical processes. Deepmind is working with Moorfields Eye Hospital to prioritise urgent cases of eye disease with no human intervention.

As technology advances we need to consciously consider their use. The implications for healthcare are to ensure that the use of machine learning fits with the reasonable expectations of the patient and that they can exercise their data protection rights.

Patient Expectations & Ethics

The first question to ask is, "Should we do this?" Normally answering the question requires an understanding of the benefits for patients. For example, it is not useful to identify patients at risk of disease if this does not change the outcome. Machine learning projects are therefore best thought of as service improvement projects that use technology, not IT projects in isolation.

The next consideration is whether patients have a reasonable expectation of their data being used in this way. For medical records held by a Trust this is relatively straightforward. It becomes more complicated if data is used from outside the context of care, for example combining medical and social housing data. Reasonable expectations can be established by talking to patient groups, providing information and asking their views. It is something that can be developed through engagement.

Patients should understand how their medical records and other data is used. Healthcare organisations will normally approach this on a whole record basis rather than explaining about individual IT systems. The addition of ML will most likely mean the organisation guidance needs expanding to explain this new use, rather than specific guidance in that case. Patients are not normally experts in the management of medical records or complex IT systems and so it is important that any guidance produced can be easily understood.

Over the last 20 years the NHS has established that patients have rights over their data, reinforcing the legislative requirements of the Data Protection Act. Just under

1% of patients will exercise their data protection rights and wish to restrict or stop processing. ML – and the teams behind it - will need to respect this.

Healthcare already uses large sets of data in the interests of the patient as part of its commitment to public health and research. These structures and processes can be used for ML to meet the expectations of patients.

Using Medical Records

Machine learning works best with large amounts of comprehensive, high quality, electronic data. The quality of the output is dependent upon the quality of the input. Healthcare organisations will need to ensure their medical records meet this standard. Care should be taken to ensure that the input data set is not biased – i.e. that it reflects the gender, age, socio-economic, ethnicity and other characteristics of the patient population. The effort for this stage should not be under-estimated.

The Data Protection Act requires that the use of data is minimised to a specified purpose. This takes two forms. The first is the range of data provided to the algorithm. This requires a level of judgment. For example, if the algorithm was being developed to predict hospital re-admission then a judgement must be made as to whether patient age, postcode and whether they drove to the hospital was required. In many cases clinicians will have a clear view of the key data attributes through their experience. In others it may be that this is determined using trial and error.

The second consideration is the volume of data used. It is good practice that patients consent to their information being used in research and development. Judgement here links strongly to the patient's reasonable expectations as to how their data is used. Data may be anonymised, but care must be taken to ensure that it cannot be re-identified. The data Protection Act 2018 does not recognise pseudonymised data as being enough protection.

Profiling

GDPR makes specific mention of profiling:

> "Data subjects shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

This is because there is a danger that individual patients are disadvantaged through the ML model, either by being selected erroneously (a false positive) or excluded (a false negative). Since the results are produced by an algorithm it can be difficult to overcome this automatically. For example, Amazon developed a job recruitment tool that was racist due to bias in the training data.

Healthcare organisations must be alert to the efficacy of the machine learning model, measuring false positives/negatives and bias against current practice and the gold standard. This requires monitoring of both the inputs and the outputs through auditing on a regular basis.

Patient have the right to understand how their data is used and what to do if they object or the results are incorrect. In many cases the machine learning model is a tool to aid the clinician and so they can overrule it. Where this is not the case there need to be processes that circumvent the model and that patients are clear where it is being used.

Trust

Unfortunately, there is a lot of hype around artificial intelligence and some high profile cases of abusing data protection (e.g. Cambridge Analytica). Healthcare organisations therefore need to build trust in the use of the technology. This starts with the care relationship between the clinician and patient.

This trust needs to extend to the IT supply chain. ML is a cutting-edge technology with new entrants into the market. New firms may not be familiar with the cyber security and information governance standards used by the NHS. These are not optional.

Data Protection Guidance

The right place to capture and manage the data protection implications is a Data Protection Impact Assessment. This section seeks to set out how to manage the implications of machine learning in this context.

Lawfulness, Fairness & Transparency

The lawful basis for processing is best considered separately for the development and subsequent use of the ML model.

- **Development** - The protocols for using medical data for research and development are published by the Health Research Authority. The basis for development will often be consent.
- **Operation** - For NHS organisations the lawful basis for processing of data for treatment is defined in Article 6(1)(e) of the GDPR, i.e. "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller". Consent is not required.

Article 9 (2) (h) provides the legal basis for the processing of the patient's medical information by the Trust.

> *Article 9 (2) (h) – processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;*

The important point here is that machine learning is nothing new when it comes to the lawful basis for processing. The controls in place for R&D and patient care will often suffice.

The use of data needs to be transparent to patients, so they can exercise their data protection rights. Whilst patients have reasonable expectations their data will be used for the delivery of care, the use of artificial intelligence is currently novel and contentious.

Healthcare organisations should consider:

- Working with patients to determine their expectations of this new technology. This could be through a patient group or engaging directly. The results of a Citizen Jury run by the National Data Guardian (National Data Guardian – Patient Expectations of Data Sharing & Privacy) demonstrate how to do this well.

- Ensuring patients understand the use of data and their rights. Overall guidance is published on NHS.uk. Healthcare organisations should revisit their data protection publications, e.g. the statement on the website.
- Reviewing data protection processes to ensure that patient's rights can be fully enforced.

Data Minimisation

Personal data must be:

- Adequate – sufficient to properly fulfil your stated purpose;
- Relevant – has a rational link to that purpose; and
- Limited to what is necessary – you do not hold more than you need for that purpose.

There is a tension with this principle and the desire to develop an accurate model with the best data inputs possible. It is often the case that the more data used, the better the resultant model.

This may be addressed through the staged use of data:

- Using anonymised or synthetic records in the early stages of development and testing.
- Adding data in stages and removing data not useful to the development of the algorithm.
- Seeking consent for test data.

Best practice is for the approach to be documented and tested with patient groups to ensure it meets with their reasonable expectations.

Healthcare organisations should note that the ICO required the Royal Free sign an undertaking after using 1.6 million records for clinical safety testing of the Deep Mind streams tool. The ICO expecting that test data is either anonymised or a smaller subset of data used with the patient's consent.

Accuracy

ML models are only as accurate as the data they receive. There have been high profile cases where a model has been racist or sexist because the input data was biased.

Healthcare organisations therefore need to consider:

- Demonstrating the safety and effectiveness of the algorithm in clinical care with close monitoring of false positives and false negatives. This needs to be a regular task throughout the lifetime of the algorithm to ensure that it does not become ineffective over time.
- Ensure that the algorithm is trained with a balanced dataset to ensure that it does not become biased. There is a risk that the data may drive racism and sexism (see Nature).

- Allow patients to view (either online or through a subject access request) the personal data input into the algorithm so they can review and edit for any accuracy issues.
- The capability for patients to stop or correct accuracy for individual patients. This can be challenging for an algorithmic approach and so may have to involve removing the patient from the tool or developing an exception handling process.

Again, best practice is to have a documented approach to data accuracy.

Storage Limitation

Personal data should only be kept for as long as necessary. Timescales for clinical records are set at a national level in the Records Management Code of Practice and an organisational one through a policy document.

Storage limitations apply to a record type, e.g. mental health records should be retained for 20 years or 8 years after the patient has died. There is not a specific category for the retention of analytics data, although the following categories are relevant:

- Screening, including cervical screening, information where no cancer/illness detected is detected – 10 years.
- Research data sets – not more than 20 years.

Design of the ML system therefore needs to consider the deletion of records once the retention period has passed or making the data truly anonymous. Medical records already have long retention periods, so this should not be an issue for the model.

Integrity & Confidentiality (Security)

Healthcare organisation must ensure that there are appropriate security measures in place to protect personal data. The project will need to ensure that IT suppliers are able to conform to standards including ISO 27001, Cyber Essentials Plus and the DS&P Toolkit. There are also specific technical risks that need project focus:

- Access controls and audit log to datasets to ensure access is minimised and that only a small number of named staff have access to the whole dataset.
- Security of data in transit.

These are not new requirements but may be unfamiliar to new entrants in the marketplace.

Accountability

The accountability principle requires that healthcare organisations take responsibility for personal data and comply with the other principles. The key point here is that that there are appropriate measures and records in place to demonstrate compliance.

Machine learning project should explicitly consider data protection as part of its governance with this documented in the terms of reference. The governance could include:

- A data protection workstream.
- Patient representation on the steering group (e.g. representative from a patient group or charity). This should include monitoring of patient feedback including complaints.
- Approval at project gateways by the Trust's Data Protection Officer.
- Data protection risks as part of the risk log.
- Documented and approved approaches.

Conclusion

The machines are coming. Information governance teams should prepare themselves and healthcare project staff for artificial intelligence. There are already some key learning points from previous work:

- Machine learning is something that anyone can understand at a high level.
- The implications of the new technology can be managed through existing data protection controls.
- Successful projects will approach data protection by design from the start.
- Trade-offs need to be developed in consultation with patients. There is no yes or no answer to data protection.
- ML projects will therefore build in and engage with information governance professionals at the start. Work closely with your IG team to understand any legal implications and potential limitations.

The first step may be to get more familiar with ML and AI. There are great free online courses from sites such as Coursera and edX. Cloud providers offer free tiers of their service to practice and learn. For a busy IG professional that may be the place to start.