Group 7 Ethical 3

Nowadays medical data doubles every 3 years, so it became impossible for doctors to analyze all the data belonging to a patient. Therefore a new approach has to be found to assist a doctor in diagnosing and treating a patient. In modern society nearly everyone has access to the internet and is connected to each other. That is why many companies start to make use of this data with machine learning approaches like neural networks to develop systems that can access all the data and generate hypothesis out of it. These systems go away from the traditional approach of a rigid decision tree to cognitive systems. This chapter covers some approaches for medical software and the ethical questions coming with them.

Diagnosing through Data Processing

IBM Watson

IBM Watson is a cognitive system that can be applied to many areas. It mirrors the same cognitive abilities as human possess, with the difference that Watson performs these steps on a much higher scale and with higher speed:

- Observation: regard phenomena and evidence
- Interpretation: generate a hypothesis out of what was observed and what is known
- Evaluation: consider which hypothesis are right and which are wrong
- Decision: choosing the best option

One of the most famous achievements was the winning of the game show jeopardy, where Watson beat two human competitors.

Demonstration of a possible application for Watson health

To be able to use such techniques Watson no longer relies on structured data like tables but on unstructured data like articles or any other literature that is produced by humans. In order to interpret these kind of data watson can understand natural language by analyzing grammar and context of a given sentence. The collection of data that Watson relies on is called its corpus and is provided by humans who take care, that the data is reliable and true. The whole data set is preprocessed by Watson to build indices that grant faster access. After this phase Watson is trained by experts which provide a ground truth in the form of question and answer pairs. There is a process of continuous learning through constant interaction with Watson and the feeding of new data. Watson can give questions to answers by generating a hypothesis out of his knowledge and assigning each hypothesis a weighted average score.

Watson health is the medical application of Watson. It takes this huge amount of data and stores it in one centralized thinking hub where it can be combined with the cognitive power of Watson to give immediate feedback to doctors and patients which for example use the watson health app. If the doctor asks Watson, the program gathers all related information to the case, and presents them to the doctor. The doctor can then review the references, check the suggested treatment and can even see current clinical trials.

Watson Health already proved his value in cancer genome sequencing [1]. A study compared Watson for Genomics to a traditional genome sequencing technique to evaluate the the ability to derive treatment option from the genomic data of glioblastoma patient cells. While the traditional technique took 160 hours of manual human analysis, Watson was able to provide the whole genome sequencing data from the tumour cells within 10 minutes.

Other Examples
A project by the google company alphabet is called project baseline [2][3]. Their aim is to find out how health is defined. The project is conducted with the cooperation of the Stanford and the Duke University. Over a period of four years they are seeking around 10,000 people to represent different ages, backgrounds, and medical histories. The job of the participants is to visit a study site for a health test up to 4 times a year, wear the latest medical tracking devices and take part in surveys about their health status and lifestyle. With this huge amount of data gathered, they hope to find out by which criteria perfectly healthy human is defined.

The chinese equivalent to google baidu produced a software called Minwa similar to the IBM Watson model relying on natural language processing. It became famous after it beat google and microsoft in an image classification challenge. However the sucess was not long lasting, after they admitted on June 1, 2015 that it hat broken rules governing the challenge.

Health Tap launched an app called Dr. A.I. to which one can describe his symptoms. The app then asks follow-up questions based on your answers and finally compiles a list of the most likely causes of the symptoms ranked by the order or seriousness. Furthermore the app tells the consumer if he should head to a doctor or not. However the app is apparently not available on the european google play store [15].

The Cloud

Cloud computing has made data and the processing of it more ubiquitous, efficient and accessible. It offers access to data with geographical transparency and the protection against accidental deletion of data. Cloud-based systems are a key part of all of our personal lives, from Internet email services to wearables and app platforms. Whether in paper or in digital format, the privacy of health information in the US is protected by the Health Insurance Portability and Accountability Act. However, saving medical records and private health data in the cloud can be risky. For example, Eastern European hackers broke into Utah’s state health records database, gaining access to personal information on 780,000 patients including some 280,000 social security numbers [4]. But there are some misconceptions as well regarding cloud storage [5]:

- The provider owns patient data: many people worry that if their data is stored in the cloud they can no longer influence this. However there is a standard agreement between cloud storage vendors and healthcare providers where the former is obligated to return all data when the provider wishes to end the agreement.
- data stored on-site is more secure than cloud: keeping data close to home does not necessarily mean high-security levels. In fact, the chances of enterprise data centers suffering a malware/bot attack are four times greater compared with a cloud-hosting provider. Furthermore the data is stored with high level encryption techniques, and is therefore harder to understand than just a sheet of paper.

The biggest ethical challenge is to keep the right of privacy and confidentiality. So information of a patient should only be released to others if the patients agrees to it. If the patient is unable to decide for whatever reason the decision can be made by the legal representative. These informations are considered confidential and must be protected. One approach to do so is to allow only authorized individuals to have access to information. If the identity of the patient can not be determined from the information the privacy confinement holds no longer. So it is a good way to strip medical information of all personal identifiers before the data can reach the cloud.
Another danger is that the data stored at the cloud can become inaccurate due to improper use of options such as "cut and paste". So the system processing the cloud might receive inaccurate data and therefore predict a wrong diagnosis [6].

Cloud Computing is on the rise and has proven to assist doctors to offer the best treatment for a patient. Though it comes with certain risks, multiple strategies are available to reduce them and overcome barriers in the implementation of digital health records.

Ethical And Legal Issues

The current legal status in the US is that doctors are still accountable for their decision making. The main reason is that machine learning diagnose tools are regarded as an additional information source and are excluded from FDA oversight by the 21st Century Cures Act. However, the clinical decision support coalition has released some voluntary guidelines[16][17]:

- **Transparency**: CDS software should give clinicians the clinical evidence supportin their recommendation
- **Competent end users**: it should only by used by a trained expert in the field
- **Time**: the software should not be used to make split-second diagnoses
- **Access to outside information**: the doctor should consider other resources as well

One can not neglect that artificial intelligence is on the rise, be it with Tesla's self driving cars or Amazon's Alexa. Currently medical misdiagnosis is the third leading cause of death in the US and machine diagnosis shows potential to drastically reduce this rate. With more competent system the questions arises if the responsibility might shift one day. If an algorithm has a higher accuracy rate than the average doctor it seems wrong to place the blame on the physician. Furthermore doctors who are scared of malpractice tend to order more tests than necessary which would reduce the costs for insurances. Even right now studies show that this is not always the best approach to protect patients. A recent study showed that patients where doctors faced an increased risk of malpractice were 22 percent more likely to become e.g. septic [18].

One could hold the system itself accountable, however it is hard to actually punish a software or demand money from it, so this would leave affected patients unsatisfied. Another approach is to accuse the designers of the software, but since it’s often impossible to understand the decision making of an AI even for its designers, this does not seem right as well. The last option is to hold the organization behind the software responsible. This would result in a clear target, however it is unclear if companies then would stay in the market if they have such a high risk.

So in order to realize the benefits of AI in healthcare, the matter of responsibility still needs to be clearly defined. There will be no easy solution but not finding one will undermine patient trust, place doctors in difficult positions, and potentially hinder further investments [19].

Current legal situation in the US

Currently, the US agency FDA, which is responsible for approving medical therapies and drugs for the US market, has no regulation formalities for artificial intelligence assisted medical decision systems yet. Each case has been treated as individual cases yet, so that approving software systems were made with regulation guidelines like the regulations for mobile medical applications, which is the closest comparable document at the moment. In between, the 21st Century Cures Act excluded medical software from FDA's jurisdiction, but this state is unclear. There exists a draft guidance document with the name "Support Regulatory Decision-Making for Medical Devices", which gives an idea about how the FDA would analyse products falling into this category. For the future, the FDA is creating a new unit for digital health with experts in AI, software engineers and medical representatives, who are responsible for analysing and creating regulatory guidelines for medical support systems based on AI and machine learning.[20]

Current legal situation in the EU/Germany

The EU regulation dogma of medical devices defines that „any instrument intended by the manufacturer to be used for human beings for the purpose of, among other things, diagnosis, prevention, monitoring, treatment or alleviation of disease“ is considered by the regulation. It does not exclude software, so assistance systems with AI/machine learning, which have a medical purpose, can be regulated by this principles. Nevertheless, software has to be analysed case-by-case in terms of a regulation. [20]

But the assessment on software systems is in no ways comparable to the assessment on medical devices, as software does not directly affect a human’s body function.
The EU MDR was recently updated, with changes in classification of some medical products. It will become law 3 years after publication. This update was enforced by members of the European parliament, which wished a regulations in terms of robotics and artificial intelligence, especially in ethical standards and establish liability for accidents involving both the technologies.[21]

Arterys, an example for a FDA approved deep learning application

For the cause of helping doctors with diagnosing heart problems, a company named Arterys built a medical imaging platform, which is based on a self-learning artificial neural network. It helps doctors to understand how the heart is working by providing accurate measurements of blood ventricles. The FDA approved the software platform, as it passed tests which guaranteed that it can be at least as accurate as humans. But it overperforms medical representatives in terms of time: A doctor needs at least 30 minutes to one hour for a result, whereas the Aterys platform gave a diagnosis several seconds later. As the platform is a cloud based-deep learning application, the FDA approved it only after examining it according to data security standards, where Arterys uses strong data anonymization techniques in combination with secure transfer protocols and encryption. Overall, this is a very big step for further approvals of machine learning backed medical applications.[22]

Machine Learning

The goal of Machine Learning is to find a model \( y(x, w) \) that maximizes the probability \( p(z|x, w) \) for a given measured feature \( x \), and a target output \( z \). Machine Learning distinguishes between supervised and unsupervised learning. In supervised learning the input \( x \) and the target output \( z = f(x) \) are known. The function \( f(x) \) is unknown. Sometimes supervised learning is also called learning from examples. Examples for tasks which can be done by supervised learning are object recognition and handwritten digit recognition. Unlike supervised learning, in unsupervised learning only the observations \( x \), which are gathered by an unknown process, are known. So no ground truth is given. An example for an unsupervised learning task is source separation [7]. The datasets used for any kind of Machine Learning are mort often split into training and test set (most common with ratio 2:1). To prevent overfitting (having a perfect model for the training data, but getting real bad results for test data) the training set is often split again into training and validation set (again the common ration is 2:1) [8].

Decision Trees

A decision tree is a simple form of Machine Learning. Also every human already used decision trees but they did not recognized it actively. The image shows a very simple decision tree deciding which animal (dog, cat, bird, fish) it is on different features (fur, barking, flying) [8].

Neural Networks

A neural network (NN) is a complex form of Machine Learning. It uses non-linear functions to process the input data. A simple neural network with exactly one hidden layer is also called multi-layered perceptron [9].
Of course these trees can be much greater and more complex. The elements of a decision tree are [8]:

- **nodes**: these are so called feature test ("Does it have fur?", "Does it bark?", "Does it fly?")
- **branches**: outcomes of the feature test (mostly yes and no respectively true and false)
- **leaves**: region in input space and distribution of samples, the maximum of the samples is the decision (assuming the vector represents values for the different animals (dog, cat, bird, fish) then (1,0,0,0) would represent a dog, (0,1,0,0) a cat, (0,0,1,0) a bird and (0,0,0,1) a fish)

In the example, the leaves are represented in a human readable way.

**Random Forests**

Multiple decision trees can be trained on the same dataset if the split in training and test set is not always the same. This ensemble is the called a random forest. To get a decision from the forest, all trees within this forest decide on the given input and the result that is picked most often is the result of the forest [8].

**Support Vector Machines**

Support Vector Machines (SVM) transform the feature space into a higher dimension. In the new dimension exists a separating hyperplane that maximizes the distance between the classes [14].

**Deep Learning / Deep Neural Networks**

If a neural network has more than one hidden layer as shown above, it is called a deep neural network. Theoretically it can have infinite layers, but one has to be aware of overfitting and calculational limits (CPU /GPU, memory).

**Diagnosis**

**Stimulated Raman Histology with Deep Learning**
Stimulated Raman Scattering (SRS) microscopy enables rapid, label-free, high-resolution microscopic imaging. But, this is not applicable because free-space optics mounted on optical tables and state-of-the-art, solid-state, continuously water-cooled lasers are not suitable for use in a clinical environment (OR). Nevertheless, advances in fibre-laser technology allows the execution of SRS microscopy within the OR. Stimulated Raman histology (SRH) is the process of simulating haematoxylin and eosin (H&E) in SRS images. In the image you can see different steps and intermediate results of this simulation. The final result is the big part (g). It is a mosaic tiled image of several SRH fields-of-view (FOVs). The asterisks (*) in the bottom left indicates a microvascular proliferation. The dashed circle in the top left indicates a calcification. The dashed box in the top right shows the part which can be seen in the FOV (e). The scale bars in the bottom right of each part demonstrate 100µm [10].

The interpretation of these images is very labour- and time-intensive. That is why machine learning is used to interpret these images. For training the machine learning algorithm, 12,879 400µm x 400µm SRH FOVs from 101 patients are used. With the WND-CHRM framework, 2919 image attributes are calculated for machine learning. The neural network (NN) is fed with the normalized attribute data. For testing the accuracy a leave-one-out-approach (use all except one patient as training set and one patient as test set) is used. The NN makes predictions for a given FOV belonging to one of currently four categories [10]. These are [10,11]:

- non-lesional
- low-grade gliial
- high-grade gliial
- non-gliial (includes metastases, meningioma, lymphoma and medulloblastoma)

The NN is trained with CUDA (parallel computing platform) on a NVIDIA GeForce GTX 1080 GPU and the Theano deep learning framework [11].

In the US the number of hospitals performing brain surgery exceeds the number of certified neuropathologists [10]. With the new method the pathologists can do their work remotely [10,11] because the time for getting the result shrinks by a tenth[11]. This in turn leads to a decreased time in the OR and a decreased risk associated with the surgery. The algorithm has already an accuracy of 90%. Neuropathologists have an accuracy on the same data of 90% to 95%. This means that the algorithm is nearly as accurate as the human professionals. With more data which is acquired daily the accuracy will increase further. This might fix the variability among pathologists, but the latter are not replaced [11].

In the future, the next big step is a large-scale clinical trial. Also, the expansion to eight categories is possible if there are more samples. Even the research beyond brain tumors is conceivable [11].

**Radiogenomics**

Radiogenomics is the recognition of genomic properties of a tumor by their appearance in images. They provide access to invaluable genetic information to predict tumor progress and response to drugs and other treatments. It uses a GPU-accelerated deep learning technique. This results in an earlier and more accurate diagnosis and treatment of tumors. But, half of the patients get potentially misleading results. The NN is also trained with CUDA on a range of NVIDIA GPUs [12]. The developing team also won the second prize (50,000$) of the NVIDIA’s 2017 Global Impact Award [1,3].

**MGMT Prediction**

Glioblastoma multiforme (GBM) is the most common and deadliest type of brain tumor. It interferes with the DNA repair mechanism. It responds best to a mixture of radiation and chemotherapy [12,14]. Methylguanine methyltransferase (MGMT) is a key gene that encodes for a protein that repairs DNA [14]. The methylated version of MGMT suppresses DNA repair. The segmentation of MRI images is done with a random forest. Afterwards SVM and random forests are used to differentiate between MGMT methylated and unmethylated GBMs. The result shows that methylation is important to predict treatment response [14].

**Bibliography**
2. https://www.projectbaseline.com/
3. https://www.googlewatchblog.de/2017/04/project-baseline-google-schwester/
6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4394583/
7. slides of the lecture Machine Learning I by Patrick van der Smagt (an introduction to Machine Learning), fall semester 2016/2017
8. slides of the lecture Machine Learning I by Wiebke Köpp (Decision Trees), fall semester 2016/2017